Identification of optimal topography of the barotropic ocean model in the North Atlantic by variational data assimilation

Eugene KAZANTSEV

INRIA, projet MOISE, Laboratoire Jean Kuntzmann, BP 53,
38041 Grenoble Cedex 9, France

E-mail: kazan@imag.fr

Abstract

The use of the data assimilation technique to identify optimal topography is discussed in frames of time-dependent motion governed by nonlinear barotropic ocean model. Assimilation of artificially generated data allows to measure the influence of various error sources and to classify the impact of noise that is present in observational data and model parameters. The choice of length of the assimilation window in 4DVar is discussed. It is shown that using longer window lengths would provide more accurate ocean topography. The topography defined using this technique can be further used in other model runs that start from other initial conditions and are situated in other parts of the model's attractor.

2000 MSC: 86-08

1 Introduction

It is now well known that even the best model alone is not sufficient to make a good forecast. Any model depends on a number of parameters, it requires initial and boundary condition as well as other data. However, interpolating data from observational points to the model grid or smoothing of observed data is not an optimal way to incorporate these data in a model. Lorenz, in his pioneering work [25], has shown that a geophysical fluid is extremely sensitive to initial conditions. This fact requires to bring the model and its initial data together, in order to work with the couple "model-data" and identify the optimal initial data for the model taking into account simultaneously the information contained in the observational data and in the equations of the model.

Optimal control methods [24] and perturbations theory [29] applied to the data assimilation technique [20, 21] show ways to do it. They allow to retrieve an optimal initial point for a given model from heterogeneous observation fields. Since the early 1990s many mathematical and geophysical teams have been involved in the development of the data assimilation strategies. One can cite many papers devoted to this problem in the domain of development of different techniques for the data assimilation such as nudging [2, 5, 37, 39], Kalman filtering and ensemble methods [3, 6, 9, 10], and variational data assimilation (3DVAR and 4DVAR) [21, 23, 31, 36].

These methods have proved capable of combining information from the model and the heterogeneous set of available observations. They have led to a remarkable increase in forecast accuracy (see, e.g., [17]). The success of the data assimilation stipulates the development of

modern models together with methods allowing to integrate all available data in the model. Thus, in 1997, acknowledging the need for better ocean observations and ocean forecasts and with the scientific and technical opportunity that readily available satellite data had delivered, the Global Ocean Data Assimilation Experiment (GODAE) was initiated to lead the way in establishing global operational oceanography. Another example, the Mercator Ocean Group was founded in 2002 to set up an operational system for describing the state of the ocean, an integral part of our environment. Input for the Mercator system comes from ocean observations measured by satellites or in situ observations through measurements taken at sea. These measurements are assimilated by the analysis and forecasting model. The assimilation of observation data in a model is used to describe and forecast the state of the ocean for up to 14 days ahead of time.

However, the majority of the data assimilation methods are now intended to identify and reconstruct an optimal initial point for the model. The importance of precise knowledge of the starting point of the model, pointed out by Lorenz [25], lead to the fact that essentially the starting point is considered as the control parameter and the target of the data assimilation.

Of course, the model's flow is extremely sensitive to its initial point. But, it is reasonable to suppose that an ocean model is also sensitive to many other parameters, like bottom topography, boundary conditions, forcing fields, and friction coefficients. All these parameters and values are also extracted in some way from observational data, interpolated on the model's grid and can neither be considered as exact nor as optimal for the model. On the other hand, due to nonlinearity and intrinsic instability of model's trajectory, its sensitivity to all these external parameters may also be described by a fast growing function.

Numerous studies show strong dependence of the model's flow on the boundary data [1, 38], on the representation of the bottom topography [11, 15, 27], on the wind stress [8, 30], on the diffusivity coefficients [7], and on fundamental parametrizations like Boussinesq and hydrostatic hypotheses [26].

Although the bottom topography and the boundary configuration of the ocean are steady and can be measured with much better accuracy than the model's initial state, it is not obvious how to represent them on the model's grid because of the limited resolution. It has been known for 30 years that requiring the large scale ocean flow to be well represented, one has to smooth the topography to get only corresponding large-scale components of relief [16]. In this case, the influence of subgrid-scales has to be parameterized. But it is not clear how to apply the parametrization for a given model with a given resolution. It is shown in [35] that different smoothing of the topography pattern may significantly change the model's properties.

This paper is devoted to the use of the data assimilation procedure in order to identify an optimal bottom topography in a simple barotropic ocean model. The first attempt to use the topography as a control parameter was performed in [28] for a steady solution of a linear shallow-water model in a zonal channel. The control parameter in the data assimilation procedure is the parameter that is modified to bring the model within an estimated error of the observations. It was shown that the relationship between topography and the surface elevation does not have a unique inverse and hence all the details of the depth field cannot be recovered. In this paper we work also with a simple, but different model. First, the model's solution is not stationary. We control the topography in the time-dependent motion. And second, the model is nonlinear with chaotic intrinsically unstable behavior. Thus, our study is placed in the usual context of forecasting when the dependence of the model's flow on the controlled parameter is strong. In addition, barotropic ocean model is a well-studied toy model frequently used for a preliminary study.

Another important feature for data assimilation consists in the existence of a nonzero kernel of the sensitivity operator. That means there is no way to reconstruct the exact topography pattern by assimilation because of the presence of modes to which the flow is not sensitive at all. This fact must be taken into account in the data assimilation analysis.

The paper is organized as follows. The second section describes the model, its adjoint, and the data assimilation procedure. The third section is devoted to numerical experiments and discussion.

2   The Model

We consider the shallow-water model developed in [13, 34] with the rigid lid assumption

$$
\begin{aligned}
\frac{\partial u}{\partial t} - fv + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{1}{\rho_0}\frac{\partial p}{\partial x} &= \frac{\tau^{(x)}}{\rho_0 H_0} - \sigma u + \mu\Delta u, \\
\frac{\partial v}{\partial t} + fu + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{1}{\rho_0}\frac{\partial p}{\partial y} &= \frac{\tau^{(y)}}{\rho_0 H_0} - \sigma v + \mu\Delta v, \\
\frac{\partial(Hu)}{\partial x} + \frac{\partial(Hv)}{\partial y} &= 0,
\end{aligned}
\tag{2.1}
$$

where $u(x,y,t)$ and $v(x,y,t)$ are two velocity components, H is the water depth, ρ_0 is the mean density of water, H_0 is the characteristic depth of the basin, and g is the reduced gravity. The model is driven by the surface wind stress with components $\tau_x(x,y)$ and $\tau_y(x,y)$ and subjected to the bottom drag that is parametrized by linear terms σu, σv and to the horizontal eddy diffusion parametrized by n harmonic operator $\mu\Delta u$ and $\mu\Delta v$. The Coriolis parameter f is supposed to be linear in y coordinate: $f = f_0 + \beta y$. Although this linear approximation may result in significant errors on the North and on the South, we use it because our goal is to study the data assimilation method rather than physical adequacy of the model.

The third equation allows us to introduce the streamfunction ψ, such as

$$
Hu = -\frac{\partial\psi}{\partial y}, \quad Hv = \frac{\partial\psi}{\partial x}.
\tag{2.2}
$$

Denoting the vorticity by $\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ we get

$$
\omega = \frac{\partial}{\partial x}\frac{1}{H}\frac{\partial\psi}{\partial x} + \frac{\partial}{\partial y}\frac{1}{H}\frac{\partial\psi}{\partial y}.
\tag{2.3}
$$

This equation is used to solve for ψ given ω. It possesses a unique solution when $H(x,y)$ is always positive. Numerically, it can be solved by Cholesky decomposition method, for example.

Using this notation we calculate the curl of the first two equations of the system (2.1). We get

$$
\frac{\partial\omega}{\partial t} + (\omega + f)\,\mathrm{div}\,\vec{u} + u\frac{\partial(\omega + f)}{\partial x} + v\frac{\partial(\omega + f)}{\partial y} = \mu\Delta\omega - \sigma\omega + \frac{\mathcal{F}(x,y)}{\rho_0 H_0}
$$

or

$$
\frac{\partial\omega}{\partial t} + \mathcal{J}\left(\psi, \frac{\omega + f}{H}\right) = \mu\Delta\omega - \sigma\omega + \frac{\mathcal{F}(x,y)}{\rho_0 H_0},
\tag{2.4}
$$

where $\mathcal{J}(\psi,\omega) = \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial y} - \frac{\partial\psi}{\partial y}\frac{\partial\omega}{\partial x}$ is the Jacobian operator and $\mathcal{F}(x,y) = -\frac{\partial\tau_x}{\partial y} + \frac{\partial\tau_y}{\partial x}$.

The system (2.4) is considered in a bounded domain Ω and is subjected to the impermeability and slip boundary conditions:

$$\psi|_{\partial\Omega} = 0, \quad \omega|_{\partial\Omega} = 0. \tag{2.5}$$

The model is discretized in space by finite element method. Details of the discretization can be found in [19].

This system has been forwarded in time by the following scheme:

$$\frac{\omega^{n+1} - \omega^{n-1}}{2\tau} + \mathcal{J}\left(\psi^n, \frac{\omega^n + f_0 + \beta y}{H}\right)$$
$$= -\mu\Delta\frac{\omega^{n+1} + \omega^{n-1}}{2} - \sigma\frac{\omega^{n+1} + \omega^{n-1}}{2} + \frac{\mathcal{F}(x,y)}{\rho_0 H_0}. \tag{2.6}$$

The first step is performed using the two-stage process. On the first stage we calculate the value of $\omega^{1/2}$ at the time $\tau/2$. At the second stage we use this value to calculate $\delta\omega^1$ with the accuracy of second order at the time τ:

$$\frac{\omega^{1/2} - \omega^0}{\tau} + \mathcal{J}\left(\psi^0, \frac{\omega^0 + f_0 + \beta y}{H}\right) = -\mu\Delta\frac{\omega^{1/2} + \omega^0}{2} - \sigma\frac{\omega^{1/2} + \omega^0}{2} + \frac{\mathcal{F}(x,y)}{\rho_0 H_0},$$
$$\frac{\omega^1 - \omega^0}{\tau} + \mathcal{J}\left(\psi^{1/2}, \frac{\omega^{1/2} + f_0 + \beta y}{H}\right) = -\mu\Delta\frac{\omega^1 + \omega^0}{2} - \sigma\frac{\omega^1 + \omega^0}{2} + \frac{\mathcal{F}(x,y)}{\rho_0 H_0}. \tag{2.7}$$

This procedure helps us to avoid numerical oscillations at the beginning of the integration of the model.

2.1   Tangent linear and adjoint models

Let us suppose the couple $\psi(x,y,t), \omega(x,y,t)$ is a solution of the system (2.3), (2.4) with a given topography $H = H(x,y)$. If we perturb the topography by some small δH, we get another solution of the system $\{\psi + \delta\psi, \omega + \delta\omega\}$.

Our goal is to define the relationship between δH and $\delta\omega$ assuming both of them to be sufficiently small:

$$\|\delta H\| \ll \|H\| \ \text{ and } \ \|\delta\omega\| \ll \|\omega\|.$$

We start from the stationary equation (2.3). So far, the couple $\{\psi + \delta\psi, \omega + \delta\omega\}$ is a solution of the system with the perturbed topography, it must satisfy the equation (2.3):

$$\omega + \delta\omega = \frac{\partial}{\partial x}\frac{1}{H + \delta H}\frac{\partial\psi + \delta\psi}{\partial x} + \frac{\partial}{\partial y}\frac{1}{H + \delta H}\frac{\partial\psi + \delta\psi}{\partial y}. \tag{2.8}$$

Using the Taylor expansion and keeping only linear terms in δH, $\delta\psi$, and $\delta\omega$ we get the equation that allows us to compute $\delta\psi$ from $\delta\omega$, δH and the reference streamfunction ψ:

$$\nabla\frac{1}{H}\nabla\delta\psi = \delta\omega + \nabla\frac{\delta H}{H^2}\nabla\psi. \tag{2.9}$$

To get the equation for vorticity perturbation, we consider the evolution equation (2.4). As above, we write the equation for the perturbed topography using the Taylor expansion and neglecting higher-order terms.

Skipping the detailed development of the tangent model, we write it in a short matricial form

$$\frac{\partial \delta\omega}{\partial t} = A(\psi,\omega)\delta\omega + B(\psi,\omega)\delta H - D\delta\omega, \qquad (2.10)$$

where operators A, B, and D are defined as

$$A(\psi,\omega)\xi = -J\left(\psi,\frac{\xi}{H}\right) + J\left(\frac{\omega + f_0 + \beta y}{H}, \left(\nabla\frac{1}{H}\nabla\right)^{-1}\xi\right), \qquad (2.11)$$

$$B(\psi,\omega)\xi = J\left(\psi,\frac{\omega + f_0 + \beta y}{H^2}\xi\right) + J\left(\frac{\omega + f_0 + \beta y}{H}, \left(\nabla\frac{1}{H}\nabla\right)^{-1}\left(\nabla\frac{\xi}{H^2}\nabla\psi\right)\right), \qquad (2.12)$$

$$D\xi = -\mu\Delta\xi + \sigma\xi.$$

The system (2.10) starts from the zero initial state $\delta\omega(x,y,0) = 0$ because the purpose is confined to the study of the sensitivity of the solution to the topography, rather than to the initial state. We require the perturbed solution $\{\psi + \delta\psi, \omega + \delta\omega\}$ to have the same boundary condition as $\{\psi,\omega\}$ because we do not want to study the model's sensitivity to boundary conditions. Hence, perturbations $\delta\psi, \delta\omega$ in equations (2.9), (2.10) must satisfy $\delta\psi|_{\partial\Omega} = 0$, $\delta\omega|_{\partial\Omega} = 0$.

Using the same time stepping scheme as in the reference model, we get

$$\delta\omega^N = \delta\omega(T) = G(\psi,\omega,T)\delta H, \qquad (2.13)$$

where G is the product of tangent linear operators on each time step of the model.

To develop the adjoint model, we calculate the adjoint of the $G(\psi,\omega,T)$ matrix.

Several remarks can be made on the tangent model formulation. First of all, we can see the right-hand side of the tangent linear model (2.10) is composed of three terms A, B, and D ((2.11), (2.12)). The combination, $A - D$, is responsible for the evolution of a small perturbation by the model's dynamics, while the operator B determines the way the bottom topography perturbation is introduced into the model. The first term is similar for any data assimilation, while the second one is specific to the particular variable under identification. This term is absent when the goal is to identify the initial point because the uncertainty is introduced only once, at the beginning of the model integration. But, when the uncertainty is presented in the bottom topography or some other internal model parameter (like forcing term or boundary conditions), the perturbation is introduced at each time step of the model.

In order to see the structure of the operator $G(\psi,\omega,T)$, let us consider the case when (ψ,ω) is a stationary solution of (2.1). In this case operators A and B do not depend on time and the system (2.10) can be solved explicitly. The solution is

$$\delta\omega(T) = \underbrace{(e^{T(A-D)} - I)(A-D)^{-1}B}_{G(T)}\delta H. \qquad (2.14)$$

The only dependence on T is associated with the exponential $e^{T(A-D)}$. If we consider long-time limit $T \longrightarrow \infty$, we see singular values of G tend to singular values of a pure exponential operator $e^{T(A-D)}$. That means the growth rate of the norm of perturbation on long-time scales does not depend on the operator B. It is determined by the maximal Lyapunov exponent of the model, that is, it is the same rate as in the study of sensitivity to initial state.

Intrinsic model's instability dominates on these time scales and the source of the perturbation is no longer important. No matter how the perturbation is introduced in the model, in several days the growth becomes exponential in time with the rate determined by the model's dynamics.

On short-time scales, when the exponential is comparable with the identity matrix, the source of perturbation is also important because the matrix B cannot be neglected in the product (2.14). On these scales the model's sensitivity to topography differs from the sensitivity to initial conditions. It was shown in [19] that exponential operator starts to dominate on time scales $T > 3$ days in this model.

Another difference consists in the fact that the model state of the tangent linear and adjoint models includes one supplementary variable. In this paper we have to add the bottom topography as the third variable to the model state.

2.2   Minimization

The purpose of this work is to identify the optimal topography of the ocean model as the field $H^*(x, y)$ that realizes the minimum of the cost function $\mathcal{I}$:

$$
\begin{aligned}
\mathcal{I}\big(H^*(x,y)\big) &= \min_H \mathcal{I}\big(H(x,y)\big) \\
&= \min_H \int_0^T \|\omega(x,y,t,H) - \omega_{obs}(x,y,t)\|^2 \, dt,
\end{aligned}
\tag{2.15}
$$

where the norm is defined as

$$
\|\omega(x,y)\| = \left(\int_\Omega \omega(x,y)^2 \, dx \, dy \right)^{1/2}
\tag{2.16}
$$

Here, $\omega_{obs}(x,y,t)$ represents the model's variable reconstructed from observations. The minimization procedure is devoted to find the topography that ensures the closest model's solution to the observational variable.

To minimize the cost function $\mathcal{I}$ we need first to find its gradient. We start from writing the variation of the functional.

$$
\begin{aligned}
\delta\mathcal{I} &= \mathcal{I}\big[H(x,y) + \delta H(x,y)\big] - \mathcal{I}\big(H(x,y)\big) \\
&= \int_0^T \big\langle\!\big\langle \omega(t,H) - \omega_{obs}(t), \omega(t,H+\delta H) - \omega(t,H) \big\rangle\!\big\rangle \, dt \\
&= \int_0^T \big\langle\!\big\langle \omega(t,H) - \omega_{obs}(t), G(t)\delta H \big\rangle\!\big\rangle \, dt \\
&= \big\langle\!\big\langle \int_0^T G^*(t)\big(\omega(t,H) - \omega_{obs}(t)\big) dt, \delta H \big\rangle\!\big\rangle
\end{aligned}
\tag{2.17}
$$

where the scalar product $\langle\!\langle a, b \rangle\!\rangle$ is associated with the norm (2.16).

Thus, the gradient of the cost function can be obtained as the integral of the product of the adjoint operator G^* (2.13) and the difference between the model state and observations

$$
\nabla\mathcal{I} = \int_0^T G^*(t)\big(\omega(t,H) - \omega_{obs}(t)\big) dt.
\tag{2.18}
$$

The operator G^* represents the product of time steps of the adjoint model. In practice, we do not need to conserve the matrix $G^*(t)$ in memory. The multiplication of a vector by the adjoint operator is realized by the adjoint model integration starting from this vector.

The minimization procedure used here was developed by Jean Charles Gilbert and Claude Lemarechal, INRIA [12]. The procedure uses the limited-memory quasi-Newton method.

2.3 Experimental setup

The domain was chosen to represent the North Atlantic region. We assume that the domain is comprised in the rectangle between $78^0W \ldots 3^0W$ in longitude and $15^0N \ldots 65^0N$ in latitude. The boundary of the basin corresponds to the $1\,\mathrm{km}$ depth isobath of the ocean.

To obtain the forcing in this experiment we have used the data set "Monthly Global Ocean Wind Stress Components" prepared and maintained by the Data Support Section, Scientific Computing Division, National Center for Atmospheric Research. These data have been prepared by the routine described in [14]. From this data set we choose the mean January wind stress components τ_x and τ_y over the North Atlantic based on 1870-1976 surface observations. These data are presented on the $2^0 \times 2^0$ grid.

The forcing in this experiment is calculated from these data as

$$\mathcal{F}(x,y) = -\frac{\partial \tau_x}{\partial \varphi} + \frac{1}{\cos\varphi}\frac{\partial \tau_y}{\partial \lambda},$$

$$\varphi = 20^0 + y \times 50^0/L, \lambda = -40^0 + \frac{x \times 50^0/L}{\cos\varphi},$$

(2.19)

where $L = 5500\,\mathrm{km}$ is the characteristic length of the basin. The spatial configuration of wind stresses τ_x and τ_y is presented in Figure 1(a).

The bottom topography has been interpolated from the ETOPO5 5-minute gridded elevation data [33]. This topography is shown in Figure 1(b).
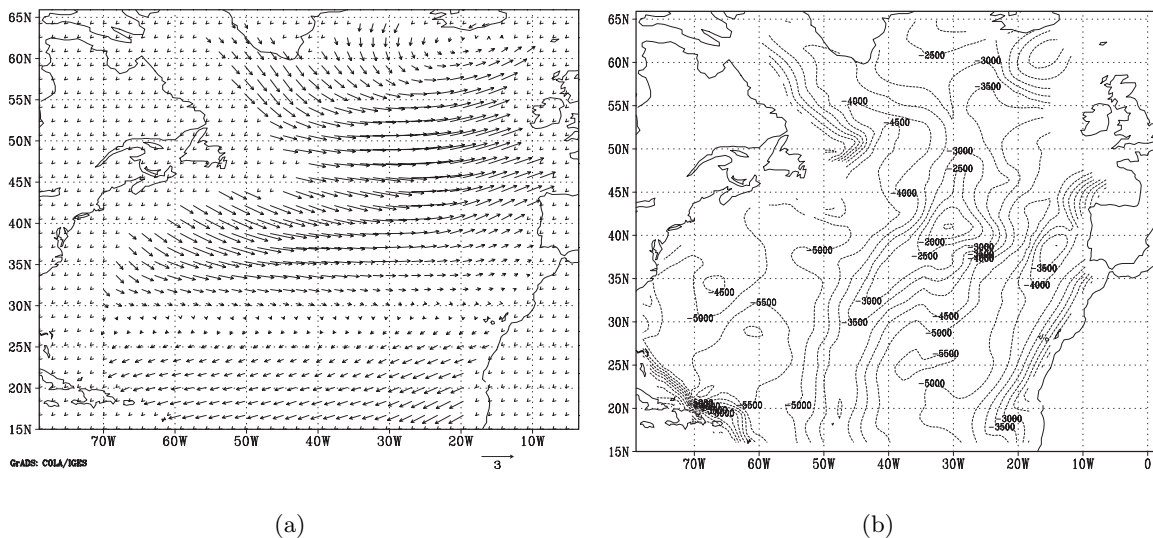

(a)   (b)

Figure 1. (a) Wind stress components τ_x, τ_y used as the forces of the model (the longest arrow corresponds to $3\frac{\mathrm{dyne}}{\mathrm{cm}^2}$) and (b) the model's topography (contours from $-500\,\mathrm{m}$ to $-6500\,\mathrm{m}$, interval $500\,\mathrm{m}$).

The coefficient of Ekman dissipation we choose as $\sigma = 5 \times 10^{-8}s^{-1}$. Lateral friction coefficient μ and forcing were chosen in order to avoid numerical instability which occurs due to the concentration of variability of the model at grid scale and to ensure chaotic behavior of the model's solution. This value has been taken to be $\mu = 300\frac{\text{m}^2}{s}$ that corresponds to the damping time scale $T_\mu = 6$ days for a wave of $100\,\text{km}$ length.

The model was discretized by the finite element method. So far, the model (2.4) under consideration is similar to a barotropic one and the solution produced by the barotropic model of the North Atlantic typically includes a western boundary layer with intense velocity gradients, the advantage of refining the triangulation along the western boundary of the domain is rather clear. A comparison of finite elements (FE) and finite difference (FD) models performed in [22] revealed that the difference arose between simulations by FE and FD techniques can be judged as insignificant when the number of FE nodes is about 6 times lower than the number of FD ones.

The package MODULEF [4] has been used to perform a triangulation of a domain in order to construct a system of $P2$ finite elements. This package produces quasi-regular triangulation based on the prescribed grid nodes on its boundary. We require the refining of the triangulation near the western boundary and especially in the middle of the domain where velocity gradients are extremely sharp. Obtained triangulation is composed of 195 triangles. The integration points set, being a union of vertices and mi-edges of triangles, counts 436 points. The resolution of this grid is about $40\,\text{km}$ near the American coast and about $300\,\text{km}$ near the European one.

3   Results

In this section we perform several numerical experiments in order to see whether the procedure is rapid and accurate especially when the observational data are noisy.

First of all, we address the behavior of the model's solution. Data assimilation is especially useful and necessary in situations when solution is chaotic and strongly depends on the control parameter.

As an evidence of irregular behavior of the discussed model's trajectory, in Figure 2 we can see the spectrum of the energy of the solution $E(t) = -\frac{1}{2\|\Omega\|}\int_\Omega \psi(x,y,t)\omega(x,y,t)\,dx\,dy$. This spectrum has been calculated from the energy double-daily time series of $\overline{T} = 50000$ days length by the discrete Fourier transform $E(t) = \sum_{k=0}^{50000}(a_k \sin(\frac{kt}{\overline{T}}) + b_k \cos(\frac{kt}{\overline{T}}))$. The amplitude of Fourier harmonics $E_k = \sqrt{a_k^2 + b_k^2}$ as a function of frequency $\nu_k = \frac{k}{\overline{T}}$ is presented in Figure 2. We see a uniform amplitude of low frequency Fourier modes that correspond to periods $1000 - 50\,000$ days and decreasing amplitudes corresponding to periods in a range from 1 to 1000 days. The rate of decrease is linear in logarithmic coordinates with the slope close to $-3/2$. That means the energy of the mode that corresponds to the wavenumber k depends on k as $E_k \sim k^{-3/2}$. This power law is close to the power law of the Kolmogorov's theory of turbulence $E_k \sim k^{-5/3}$. This fact shows the flow is turbulent, the solution's behavior is irregular, and there is no dominating periodic motions.

In order to test the efficiency of the proposed procedure, we generate artificial "observational" data using the same model with the reference topography that is presented in Figure 1(b). This fact ensures the existence of the absolute minimum of the cost function with vanishing value. Our purpose is to test the minimization procedure and its capacity to find the topography used to produce "observational" data, by assimilating these data in the model.
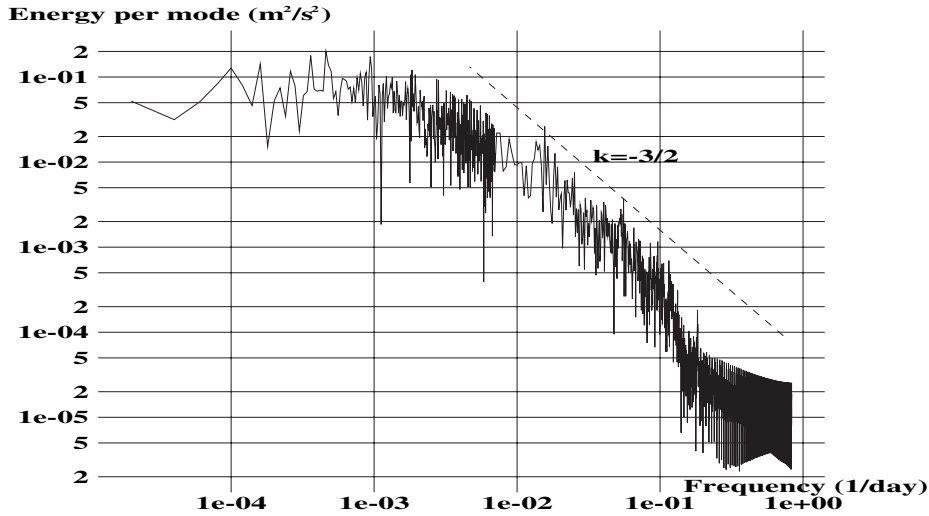
Figure 2. Energy spectrum of the model's solution.

The initial guess of the assimilation procedure is taken as a flat bottom at 4000 meters depth. Thus, we suppose we have no preliminary information about the bottom relief in order to perform the assimilation in the most difficult case. During minimization process, we keep track of two values that characterize the distance to the reference solution. The first one is the value of the cost function (2.15). It shows the decrease rate of the functional during the minimization. This value indicates how close is the assimilated trajectory to the reference one, but gives no information about the error in the reconstructed topography.

In fact, due to the presence of nonzero kernel of the Hessian of the cost function, its minimum is not unique. As it has been discussed in [19], the mode in the kernel can be easily seen from a simple analysis of the model (2.1). If we add to the topography H some perturbation which is proportional to H itself $\delta H = \alpha H$, the model remains the same. In this case, only the third equation of the system (2.1) is multiplied by $1 + \alpha$ and that does not disturb the equality to 0. Hence, the model exhibits no sensitivity to the perturbation $\delta H = \alpha H$ and this mode belongs to the kernel of the operator $G(T)$ (2.13).

So, the assimilation can only help us to find the reference topography multiplied by some arbitrary constant. We must either have some *a priori* information about the topography under reconstruction to be able to estimate this constant, or accept this nonunique result of assimilation. In this paper we choose to take into account this arbitrary constant in the post-processing. So, along with the cost function, we trace also the minimal values of the difference between the topography on the current iteration $H_n(x, y)$ and the reference topography $\bar{H}(x, y)$ in form

$$\eta_n = \min_{\alpha} \left\| H_n(x, y) - \alpha \bar{H}(x, y) \right\|, \tag{3.1}$$

where the norm $\|H\|$ is defined by (2.16).

In Figure 3 one can see an example of the evolution of the descent procedure. Starting at value $\mathcal{I} = 54$, the cost function decreases rapidly during first 10 iterations. During this time, the contribution of the most sensitive modes of the Hessian is minimized. In fact, after 10 iterations the model's trajectory is already close to the reference one. The value of the cost function has been divided by 200. At the same time, the topography is still far from the reference one, the value of η_n (3.1) has only been divided by 1.5. The reason of this is
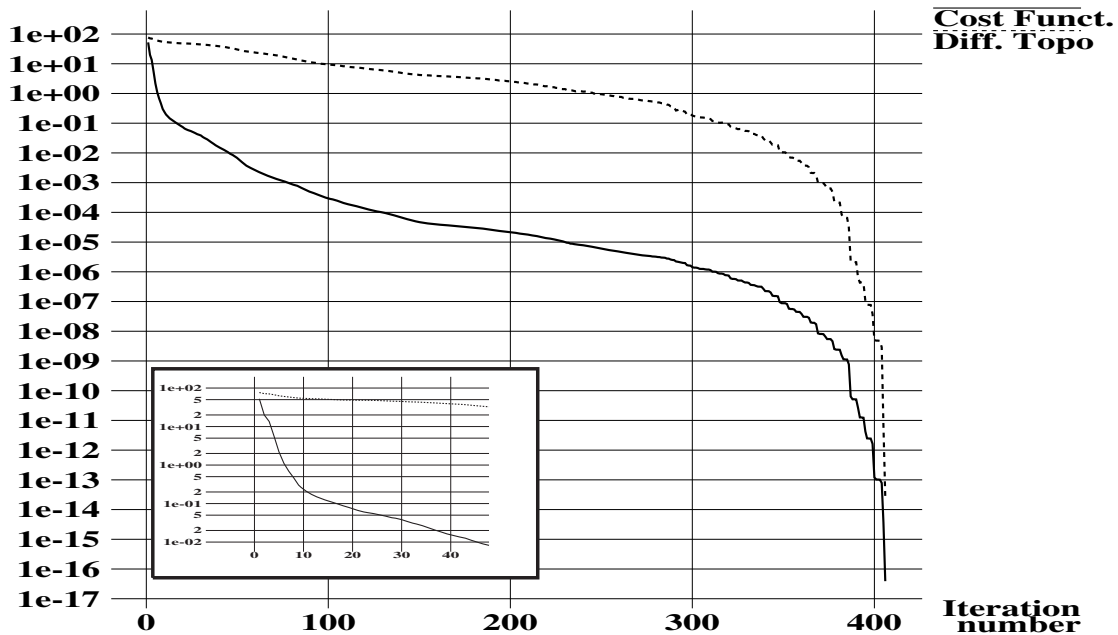
Figure 3. Cost function and η_n during minimization.

simple: all Hessian modes with low sensitivity have not been damped at this time. These modes contribute few in the cost function because of low sensitivity of the model. But their contribution in η_n is as important as the contribution of sensitive modes.

After that, the rate of decrease becomes slower. The procedure needs many iterations to damp modes with low sensitivity. The final value of the minimization vanishes, as supposed, because the cost function in this case has a clear minimum with zero value.

Thus, we may note only few iterations of the assimilation are sufficient to get a good trajectory which is close to the reference one. After 100 iterations the cost function, being divided by 10^5, is already negligible. However, the identified bottom topography may not be close to the topography of the reference experiment because of presence of numerous modes with low sensitivity that are not identified at this moment.

Assimilation error $H_n(x,y) - \bar{H}(x,y)$ at tenth ($n = 10$) and at hundredth ($n = 100$) iterations are shown in Figure 4. One can see that at tenth iteration the identified topography is much closer to the initial guess (flat bottom of 4000 m) than to the final pattern showed in Figure 1(b). Error in bathymetry in the region near American coast is about 1 km. That means the initial guess is improved by 500 meters. Mid-Atlantic Ridge is not identified at all as well as the Porcupine bank on the North-East of the basin. Errors in identification of the height of the Mid-Atlantic Ridge reach 2.5 km that is close to the difference between it's real height and initial guess of 4 km.

At the hundredth iteration, both Mid-Atlantic Ridge and the bottom topography to the West from the Ridge are identified. The error of identification in this region is of order of 300 meters. However, bottom relief is not well identified to the East and to the North from the Mid-Atlantic Ridge. Error in estimation is still big, reaching 1 km. As it is noted in [19], low sensitivity modes of the Hessian are concentrated in this region where flow is laminar. These modes are not dumped at 100th iteration and they provide the principal contribution in the error at this stage of assimilation.
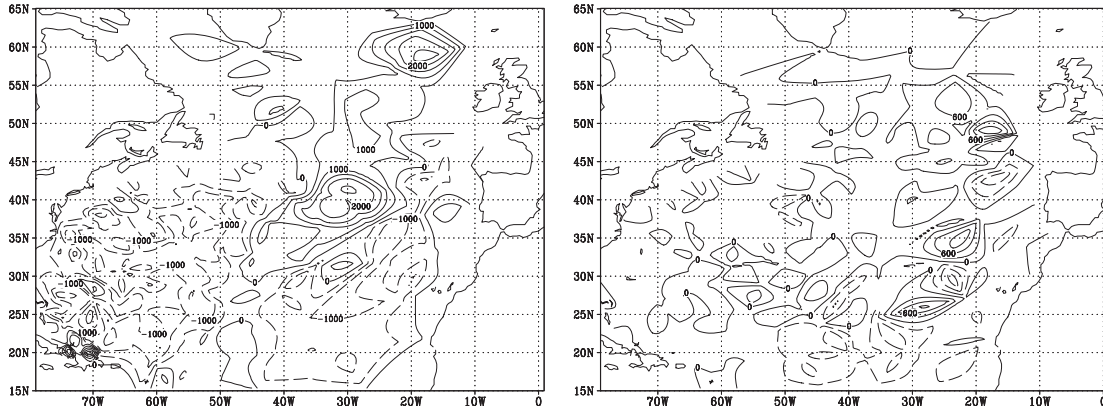
Figure 4. Difference between the reference and identified topography after 10 (left) and 100 (right) iterations. Contours from $-1500\,\mathrm{m}$ to $2500\,\mathrm{m}$ interval $500\,\mathrm{m}$ on the left and from $-900\,\mathrm{m}$ to $1200\,\mathrm{m}$ interval $300\,\mathrm{m}$ on the right.

3.1 Assimilation window: exact "observations"

The first question we address concerns the length of the assimilation window T, that is, the time span during which the assimilation is performed. There is no a priori information for the choice of T. In fact, we can choose T as one time step of the model and as one month as well. In order to see the influence of the window's width on the convergence of the assimilation process, we will distinguish two aspects of this influence: the window's width itself and the quantity of external information contained in the window. When we use the cost function (2.15) that is similar to classical 4DVAR cost functions, different windows contain different quantity of the observational information that is introduced into the model. Each window contains as many observational fields as time steps.

When we use the simplified cost function

$$\mathcal{I}_{\mathrm{simpl}}\big(H(x,y)\big) = \|\omega(T,H) - \omega_{obs}(T)\|^2 \tag{3.2}$$

that compares observational and model's data just at the end of time interval T, each window contains exactly the same quantity of information: only one observational field at the end of the window. Assimilating data by minimization of this cost function $\mathcal{I}_{\mathrm{simpl}}$ helps us to see the effect of the window's width itself, while using the original cost function (2.15) we will see the joint effect of the window and the information in the window.

When the assimilation window T increases, the computing time per iteration increases also because the iteration is composed by the integration of the direct model from 0 to T and backward integration of the adjoint model. Hence, looking for the optimal T, we should refer to the CPU time along with the number of iterations.

In order to see the dependence of the convergence rate on the window's width, we perform 8 experiments with different assimilation windows T. All other parameters are the same in this set. We start from the value $T = 0.1$ day that corresponds to one model's time step. For all subsequent experiments, we double T.

The "observations" are produced by the same model; hence the expected cost function's value must be vanishing at the end of assimilation. We use the criterium $\mathcal{I}(H(x,y)) = 10^{-16}$ to stop iterations. The number of iterations and the CPU time that are necessary to converge the minimization are shown in Table 1. The CPU time is expressed in minutes on Intel Pentium processor.

Table 1. Number of iterations and CPU time necessary to converge the minimization process with different T.

T(days)	$\mathcal{I}$		$\mathcal{I}_{\text{simpl}}$	
	Number of iterations	CPU Time	Number of iterations	CPU Time
0.1	2039	1.7	2039	1.7
0.2	1842	2.3	1970	2.3
0.4	1836	3.8	2095	4.1
0.8	1747	6.6	2060	7.3
1.6	1606	11.4	2243	15.1
3.2	1128	15.5	2257	29.5
6.4	891	24.2	2419	62.3
12.8	936	50.7	2398	123.5

Analyzing Table 1, we can see several tendencies. If we use the cost function $\mathcal{I}$ that uses observations over the assimilation window, the number of iterations becomes smaller when the assimilation window increases. Hence, each iteration becomes more efficient. It is not surprising, because it uses more external information. However, smaller number of iterations with longer T cannot compensate increasing of the CPU time per iteration. Each iteration requires almost double CPU time for double T, while the number of iterations is far from being divided by two. And, consequently, CPU time is lower for smaller assimilation windows.

In the case when the cost function $\mathcal{I}_{\text{simpl}}$ is used, the number of iterations seems to be constant for windows smaller than one day, but when T increases, the minimization requires slightly more iterations to converge. Indeed, no additional information is introduced into the model by enlarging the window. Therefore, there is no hope for the assimilation to be more efficient. In this case, hence, the convergence is undoubtedly more rapid when T is small.

As a consequence, when observations contain exactly the necessary information, the best assimilation window is one time step for both $\mathcal{I}_{\text{simpl}}$ and $\mathcal{I}$ (they coincide in this case).

3.2 Assimilation window: noisy "observations"

However, this experiment was carried out in frames of idealized situation. Artificial observations created by the same model contain the exact information about the model's topography. This is not the case if we use real observations containing measurement error. In addition to this, real observations include signatures of many different physical processes that are not taken into account in the model. In this case, the information about topography contained in observations can no longer be considered as exact.

In order to simulate the influence of possible errors in observational fields we add a white noise of several different amplitudes to assimilated data.

An example of this assimilation experiment is presented in Figure 5. As above, in order to produce the data to be assimilated, we run the model with the topography $\bar{H}(x,y)$ (see Figure 1(a)). After that, a white noise weighted by some small ε has been added to these data at each grid point and at each time step:

$$\omega'(x,y,t) = \omega(x,y,t) + \varepsilon \frac{\|\omega(x,y,t)\|}{\|r(x,y,t)\|} \times r(x,y,t), \tag{3.3}$$

where $r(x,y,t)$ is a random real number from the interval $-0.5 \ldots 0.5$.

An example of the noise effect is shown in Figure 5 on the left. The solid line represents the enstrophy of the reference experiment. Irregular dashed line with small dashes shows the trajectory with noise. So far the noise is irregular, it produces additional enstrophy to flow fields, providing the dashed line is almost always above the solid one. The smooth dashed line with long dashes represents the enstrophy of the trajectory of the model with identified topography. One can see this line is almost indistinguishable from the reference one.

The convergence of the cost function and the difference η_n defined by (3.1) during the assimilation process are shown in Figure 5 on the right. Contrary to Figure 3, neither the cost function nor the difference vanishes at the end of assimilation. They both tend to some final error. It is clear, this final error is due to the noise that cannot be assimilated by the model.
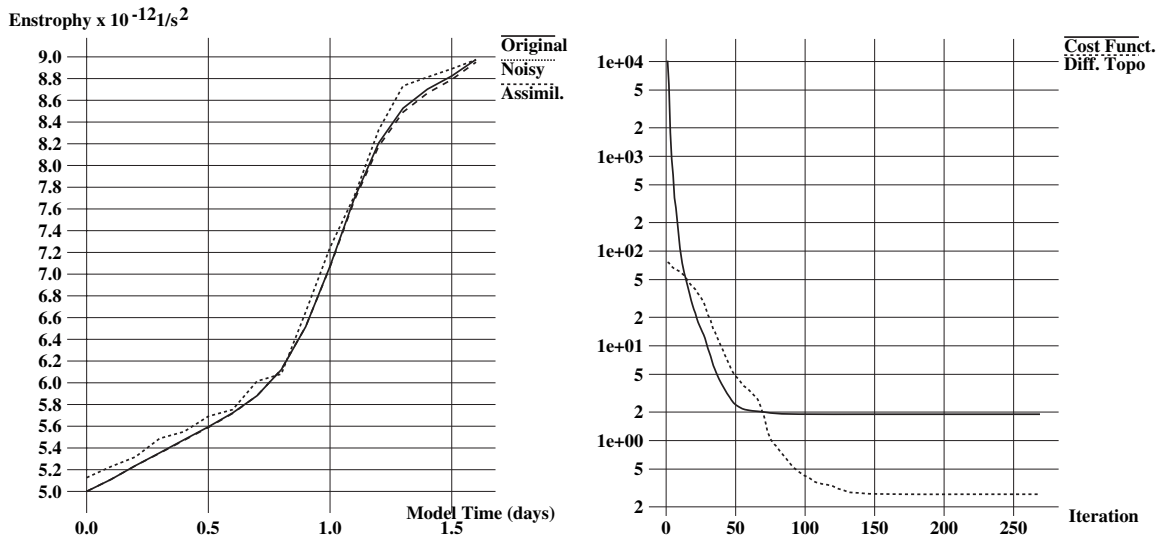

Figure 5. Noisy data. Left: Enstrophy of the reference (solid line), noisy (short dashes) trajectories, and the assimilation result (long dashes). Right: Convergence of the cost function (solid line) and of the difference η_n (dashed line) in assimilation of noisy data.

The question we ask, whether small values of T are still optimal in the case of assimilation of noisy data. In fact, in the case when artificial "observational" data were generated by the same model without noise, the result of identification of the topography was always the same. We were always able to reconstruct the exact topography of the model and reduce the cost function's value to zero. We were optimizing the computational cost rather than the result of assimilation. But when we assume the presence of noise in observational data, the optimality condition is not the same as in the previous case. Now it is better to obtain more accurate topography than lower computing time. It is not the convergence rate that we want to optimize, but the final error in the reconstructed topography.

In order to see whether there exists an optimal assimilation window in this case, we use noisy data $\omega'(x, y, t)$ as artificial observations in the assimilation procedure with the cost function (2.15) that uses observational data over the whole assimilation window. A set of experiments has been carried out with different windows. In the first experiment the window length was taken as the smallest possible one time step window. In each next experiment, the window length was two times longer. $T = 0.1, 0.2, \ldots 51.2$ days. Several examples of the convergence of η_n are shown in Figure 6 for $\varepsilon = 10^{-3}$. We see in this figure, when the noise is present, smaller assimilation windows provide worse results. Although the convergence
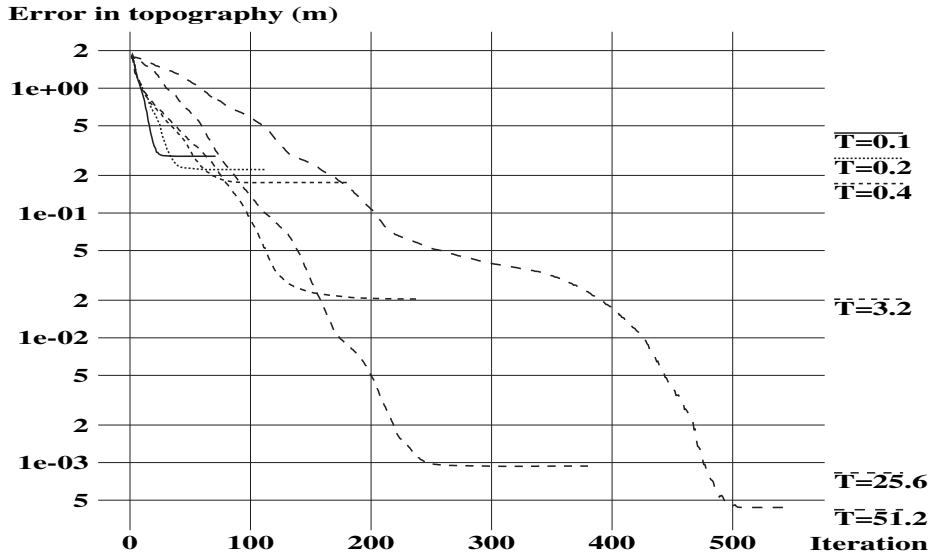
Error in topography (m)

Figure 6. Convergence of the difference η_n for different assimilation windows T in the case of noisy data.

rate remains better for small T, the final error of reconstruction is relatively big. Thus, if assimilation window is restricted to $T = 0.1$ days, the value η converges rapidly to the value 2.8×10^{-1} and after that remains stable. Multiplying T by 4 makes the convergence slower, but allows to reach $\eta = 1.7 \times 10^{-1}$. Using longer assimilation windows allows us to reduce final residual error in topography to $\eta = 2 \times 10^{-2}$ with T = 3.2 days; to $\eta = 9.4 \times 10^{-4}$ with $T = 25.6$ days and even to $\eta = 4.4 \times 10^{-4}$ with $T = 51.2$ days. However, the convergence becomes slower for longer windows. Not only the number of iterations of the descent procedure increases, but each iteration requires more computer time, because on each iteration the model and its adjoint are integrated for a longer time.

There exists, hence, no optimal value of T. Longer windows provide always better accuracy of the identified topography. This is clearly seen at Figure 7, where the final value of η is plotted for each window T for three different amplitudes of added noise $\varepsilon = 10^{-4}, 10^{-3}, 10^{-2}$ (these amplitudes are shown in the figure as small, medium, and big noises, respectively).

Increasing assimilation window, we increase the quantity of assimilated information and improve the accuracy of the resulting topography. This fact seems to be natural if we take into account that assimilated data are noisy. More noisy information introduced into the model helps to reduce the noise impact and results in better accuracy. The central limit theorem provides the expectation of the decrease rate of residual error η be proportional to $T^{-1/2}$. However, in Figure 7 we see two different decrease rates.

The first one can be seen for windows in range from 0.1 to 2 days. Indeed, the value of η decreases as $T^{-1/2}$ in this range. This fact shows the statistical law dominates for these windows. The noise contained in data influences directly assimilation results and we obtain the dependence on T predicted by the central limit theorem.

But when the window is longer than 2 days, we see more rapid decrease of residual error in topography. On this part the decrease rate is close to $T^{-3/2}$ and we can deduce the influence of noise in data is not as straightforward as before. In order to understand whether the change of the decrease rate has a physical meaning and occurs effectively at assimilation window $T = 1 - 2$ days, or this change has numerical origins and corresponds to several time steps
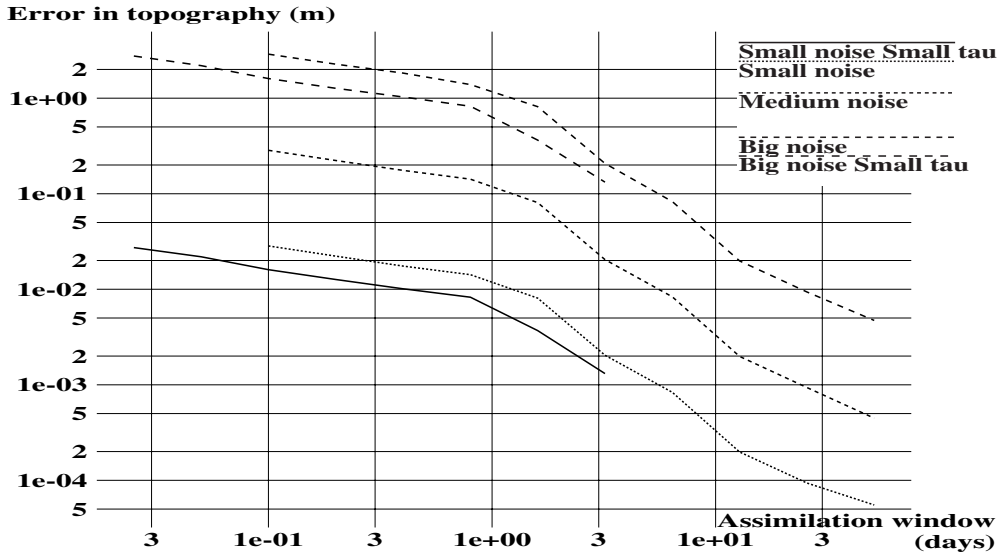
Figure 7. Final value of η_n for different assimilation windows T with small (solid lines), medium (short dashes), and big (long dashes) noises. Small ($\tau = 0.025$ days) and normal ($\tau = 0.1$ days) time steps are considered in experiments with small and big noises.

of the model rather than to physical time, we perform two similar experiments with 4 times smaller time step. Assimilation windows in these two experiments were chosen in range from 0.025 days (one small time step) to 1.6 days (64 small time steps). Corresponding lines can also be seen in Figure 7. Comparing them with lines corresponding to the 0.1 days time step, we see the same decrease rate: $T^{-1/2}$ for small windows and acceleration for bigger windows. That means the change in decrease rate is not related to numerical time stepping and must have physical origins. We see also both lines with smaller time step lie below corresponding lines with normal time step. It is natural: so far, the time step is smaller, more external information is introduced into the model each time unit. And more assimilated information results in better precision.

Moreover, comparing lines with small and normal time steps, we can see the residual error in topography depends on the quantity of assimilated data only. No matter how long was the assimilation window, it is the number of time steps (which is equal to the number of assimilated observational fields) that determines the residual error. Residual errors obtained with one, two, or three time steps are approximately equal for both the time step length 0.025 and 0.1 days.

Hence, the change in decrease rate has some physical reason. Indeed, it was shown in [19] the time scale 2-3 days is important in the sensitivity analysis of this model. On time scales lower than 3 days, the sensitivity is linear in time. That means an error in the model's solution induced by a topography perturbation grows linearly in time. But, on longer time scales, error growth becomes exponential. That means, during 2-3 days of model time the perturbation is introduced into the model and, after that, it follows the model's dynamics. During the phase of introduction, linear transition of perturbation from topography to model's variable dominates, resulting in linear dependence on time. But after 2-3 days, it is the model's dynamics that governs the error evolution. Being nonlinear and intrinsicly unstable, the dynamics ensures exponential error growth of a perturbation. On these time scales, topography perturbation evolves like any other perturbation from any other source.

Assimilating noisy data we see a similar difference between small and large scales. If the model's dynamics has not enough time to get adapted to the noise contained in assimilated data, then the noise effect is directly transmitted to identified topography. So, its influence on assimilation's results is similar to influence of a pure white noise. On large time scales the noise is transmitted by the dynamics and modified during transmission. As a result, we have more rapid decrease of the residual error.

3.3   Noise amplitude and assimilation error

In order to quantify the relationship between noise in model's parameters and error in the reconstructed topography, we perform a set of experiments assimilating noisy data. We examine the amplitude of the noise and also its origins. It is clear, bigger noise will provide bigger residual error in topography. But the source of the noise may also be important.

First, we add the noise in the model's initial conditions simulating the influence of interpolation or residual errors of reconstruction of initial point in the data assimilation process. Second, we perturb the forcing $\mathcal{F}(x, y)$ of the model (2.4) in order to simulate the difference between parametrization of physical processes in the model and real physical processes in observational data. And third, we add the noise to all grid-points at all time steps of the artificial observations to simulate the measurements errors.

Before data assimilation and topography identification, we will see the influence of different perturbation's sources on the model's variable. Each of above-mentioned parameters (initial conditions, forcing, vorticity) were perturbed by the noise (3.3) of a prescribed amplitude $\varepsilon = 10^{-3}$. In Figure 8 we trace evolution of the norm of the difference between the reference trajectory ω with no noise and noisy trajectory ω' during 150 days model run:

$$\xi(t) = \left\| \omega'(x, y, t) - \omega(x, y, t) \right\| = \sqrt{\int \left(\omega'(x, y, t) - \omega(x, y, t) \right)^2 dx\, dy}. \tag{3.4}$$

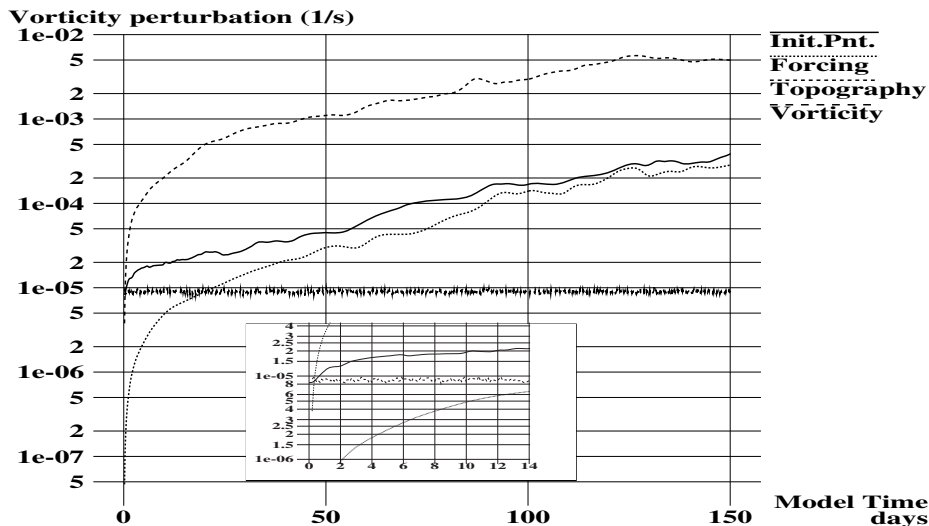Evolution of ξ during first 15 days is presented as a zoom in this figure.

Figure 8. Evolution of vorticity perturbation for different sources of noise: noise in initial point (solid line), noise in forcing (dotted line), noise added to vorticity (oscillating line), and noise in the bottom topography (dashed line).

The solid line in this figure represents the evolution of the noise added to initial conditions:

$$\omega_0'(x,y) = \omega_0(x,y) + \varepsilon \frac{\|\omega_0(x,y)\|}{\|r(x,y)\|} \times r(x,y),$$

where $r(x,y)$ is a random real number from the interval $-0.5\ldots0.5$.

Value $\varepsilon = 10^{-3}$ signifies the norm of initial point is perturbed by 0.1% of its value:

$$\frac{\|\omega_0'(x,y) - \omega_0(x,y)\|}{\|\omega_0(x,y)\|} = 10^{-3}.$$

This line starts at $\xi = 9 \times 10^{-6}\frac{1}{s}$ and grows rapidly during the first four days. This is the case of well-known super-exponential error growth (see, e.g., [32]) that reduces short-range predictability of chaotic systems because local Lyapunov exponents (that govern the error growth on short-time scales) are bigger than global exponents on long-time scales [18]. After the first four days, the value of ξ grows exponentially (linearly in logarithmic coordinates), with growth rate determined by long-time Lyapunov exponents. Beyond the 150 days interval, the line will reach saturation with values about $\xi = 10^{-2}\frac{1}{s}$ which represents the characteristic radius of the model's attractor.

The dotted line in Figure 8 represents the evolution of ξ in the case when the noise has been added to the forcing of the model. This line, obviously, starts from 0 but the rapid growth lasts more than 20 days. This happens because modification of the forcing of the model changes its attractor. The trajectory evolves first toward the new attractor and after that, the value of ξ grow also exponentially, in a similar way as the solid line.

An oscillating line represents the noise added to all grid-points at all time steps to the reference trajectory in order to simulate the measurements errors. This noise is not governed by the model's dynamics, consequently these lines are oscillating about constant values. This line oscillates about $9 \times 10^{-6}\frac{1}{s}$, the same value from which starts the solid line. This is natural because the noise that has been added to initial point is the same as the noise added to the vorticity.

And finally, in order to compare the influence of noise in all of these parameters with the influence of the same noise in the bottom topography of the model, we plot the fifth line. The dashed line in this figure represents the evolution of the norm ξ (3.4) when the noise of the same amplitude ε has been added to the topography. The amplitude $\varepsilon = 10^{-3}$ means the depth was perturbed by 4 meters in average. This line reveals stronger influence of the noise in topography field on the trajectory. The value ξ is equal to zero at the beginning of integration as well as in experiment in which the noise is added to the forcing of the model. Also similarly to the experiment with forcing, rapid increase of ξ lasts approximately 20 days. As well as the forcing modification, modification of topography also changes the model's attractor and the trajectory goes also first to the new attractor. However, the noise in topography generates much higher growth rate of ξ than the noise in forcing. That means, the attractor is more modified by a little change of topography than by an equal change of the forcing. This fact shows the importance of better identification of the topography of the model.

Error obtained in the reconstructed topography for different perturbations in model parameters and its trajectory is shown in Figure 9. The experiment was carried out in four sets. In each set one and only one parameter (initial point, forcing, vorticity or SSH of the trajectory itself) has been perturbed by random noise with ten different amplitudes. In the first set the "observational" data were generated by the model with perturbed initial conditions.
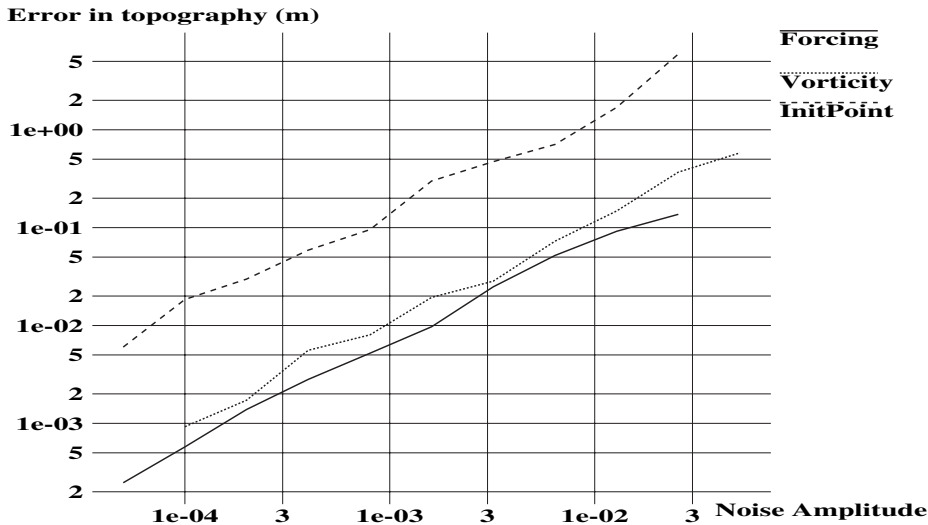
Figure 9. Residual error in topography η as a function of noise amplitude ε.

The perturbation amplitude was doubled in each of ten experiments beginning at the amplitude $\varepsilon = 5 \times 10^{-5}$. The amplitude of perturbation in the tenth experiment was equal to $\varepsilon = 2^9 \times 5 \times 10^{-5} = 2.56 \times 10^{-2}$.

The second set was performed with "observational" data obtained with non-perturbed initial point, but with noisy forcing. As before, the noise amplitude was doubled in each experiment beginning at the same value ε. And in the third set, neither initial point, nor forcing were perturbed, but the resulting trajectory was subjected to perturbation with amplitudes beginning at $\varepsilon = 10^{-4}$. In the fourth set, noisy sea surface height was used as observable variable with the same noise amplitudes as in the third set.

Assimilation window in all these experiments was chosen to be $T = 5$ days. The topography shown in Figure 1(a) used to create the reference "observational" data was the same in all experiments. The data assimilation of noisy data was performed up to stabilization of the minimization processed. We trace then the resulting norm of the difference between the reconstructed topography and the exact one, used to create the reference "observational" data.

The dashed upper line with long dashes shows the dependence of the final error in topography on the amplitude of the perturbation of initial conditions. Higher position of this line indicates stronger influence of errors in approximation of initial point of the model. This is not the case when the forcing is perturbed (lowest solid line). We can see that the model exhibits lower sensitivity. The amplitude of the noise in the forcing can be 100 times higher but the assimilation provides equal error obtained in the reconstructed topography.

The influence of noise in the trajectory of the model (dotted line) is a little higher than the influence of noise in the forcing but lower than the influence of noise in initial point.

The slope of all lines in Figure 9 is equal to one. That means the value of η (3.1) is linear function of ε.

3.4   Beyond the assimilation window

An important difference between identification of optimal initial point of the model and its optimal topography consists in the fact that topography is an internal model's parameter. It

must be identified once for all model runs, while initial conditions are external parameters and must be identified for each particular model run. The question we should ask in this case, whether the topography identified once by data assimilation is valid for other model runs that start from other initial conditions and are situated in other parts of the model's attractor.

In this experiment we perform different model runs using results of assimilation of noisy data obtained in previous section. All runs start from different arbitrary points on the attractor of the model with the reference topography. These initial points have been chosen as arbitrary points on the long trajectory of the model integrated with the reference topography.

As we have seen, assimilation of noisy observations results in an error of topography reconstruction. Consequently, it is hopeless to obtain the same solution with two different topographies. So, we will analyze the evolution of the relative difference between trajectories of models with the reference topography $\bar{H}$ and with the reconstructed topography H_{reconsr} containing errors due to assimilation of noisy data:

$$r_{\mathrm{reconsr}}^2(t) = \frac{\int_{\Omega} \left(\omega_{\bar{H}}(x,y,t) - \omega_{H_{\mathrm{reconsr}}}(x,y,t) \right)^2 dx\, dy}{\int_{\Omega} \left(\omega_{\bar{H}}(x,y,t) \right)^2 dx\, dy}. \tag{3.5}$$

This difference is compared with the effect of a simple perturbation of topography by random values at each grid-point:

$$H_{\mathrm{prtrb}}(x,y) = \bar{H}(x,y) + \frac{\eta}{\|r(x,y)\|} \times r(x,y).$$

The amplitude of perturbation in each experiment was exactly equal to the norm of the residual error in the topography after assimilation: $\|H_{\mathrm{prtrb}}(x,y) - \bar{H}(x,y)\| = \eta$. The model with perturbed topography has been integrated from the same initial point as the reference model for 180 days. During this time we calculate also the relative difference $r_{\mathrm{prtrb}}(t)$ that is similar to (3.5) but we use perturbed topography H_{prtrb} instead of reconstructed one H_{reconsr}.

Ten values of residual errors in topography η in range from $10\,\mathrm{cm}$ to $70\,\mathrm{m}$ have been tested together with ten equal perturbation amplitudes. Evolutions of $r_{\mathrm{prtrb}}(t)$ and $r_{\mathrm{reconsr}}(t)$ corresponding to the lowest and to the highest η are shown in Figure 10 on the left. Two solid lines represent the difference $r_{\mathrm{reconsr}}(t)$ for $\eta = 10\,\mathrm{cm}$ and $\eta = 70\,\mathrm{m}$. Two dashed lines show the difference $r_{\mathrm{prtrb}}(t)$ for the same amplitudes of random perturbation. One can see that for both η the difference obtained with the model with reconstructed topography is 5-10 times lower than that with randomly perturbed topography. Although amplitudes of the random perturbation and of the residual error are equal to each other, the trajectory of the model on the reconstructed topography corresponds better to the reference model. This can be explained by the fact that all sensitive modes in the topography have been damped during the data assimilation procedure. The residual error is essentially concentrated in modes that provide low impact on the trajectory. While the random perturbation of the topography adds noise to all modes and results in bigger model error.

We can see the same feature in Figure 10 on the right where differences $r_{\mathrm{reconsr}}(t)$ and $r_{\mathrm{prtrb}}(t)$ are shown at time $t = 60$ days for all ten amplitudes of the residual error and of the random perturbation. Model with the reconstructed topography shows 3-10 times lower difference with the reference model than the model with randomly perturbed topography. The dependence of the difference at $t = 60$ days on the amplitude of residual error is linear in logarithmic coordinates with the slope equal to 1. That means this dependence is simply linear. If we remind that the dependence of the residual error on the noise amplitude in assimilated data is also linear, we can deduce that the error in the model trajectory is proportional to the error in assimilated data.
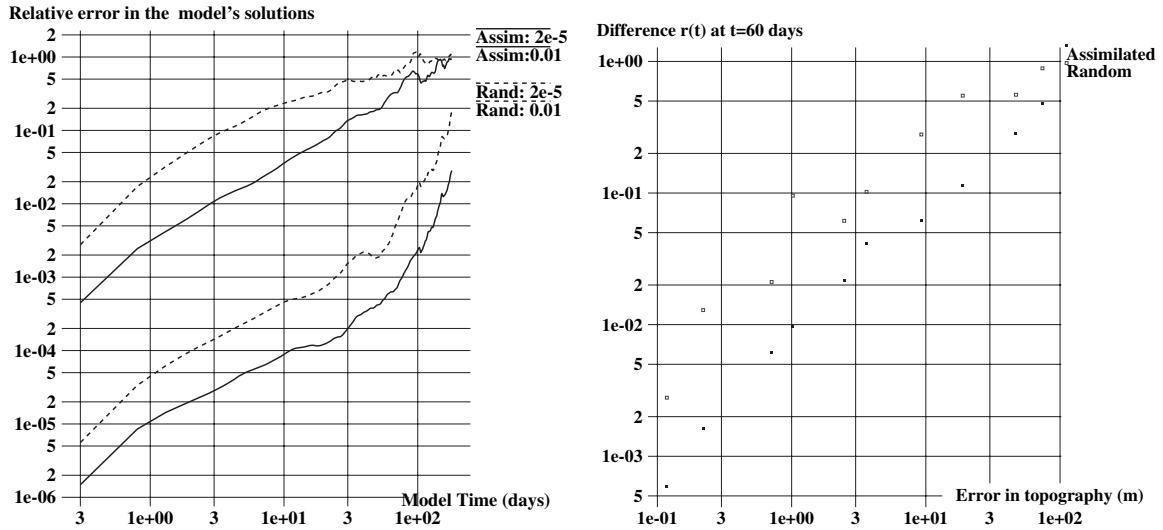
Figure 10. Evolution (left) and the value at 60th day (right) of the relative difference $r(t)$ between the reference model and models with inaccurate topographies: reconstructed topographies (solid lines) and randomly perturbed topographies (dashed lines).

4   Conclusion

We have studied the procedure of data assimilation for identification of the topography for a simple barotropic ocean model. Comparing this procedure with now well-developed data assimilation intended to identify optimal initial state of the model, we can say that there are both common points and differences as well.

Tangent and adjoint model are composed by two terms A and B ((2.11), (2.12)). The first one, A, governs the evolution of a small perturbation by the model's dynamics. This term is common for any data assimilation no matter what parameter we want to identify. The second one, B, determines the way how the uncertainty is introduced into the model. This term is specific to the particular variable under identification. This term is absent when the goal is to identify initial point because the uncertainty is introduced only once, at the beginning of the model integration.

The presence of nonzero kernel of the sensitivity operator constitutes another particularity of the data assimilation in this case. Exact topography cannot be reconstructed because there exists a mode the model is not sensitive to. Adding this mode to the topography of the model does not change its solution. Consequently, this mode cannot be identified from the model's trajectory. We must either have some a priori information about topography, or modify the model in order to suppress the kernel. In this paper the kernel can be suppressed by replacement of the average ocean depth H_0 by the real depth in each point $H(x,y)$ in the forcing term $\frac{\mathcal{F}(x,y)}{\rho_0 H_0}$ in (2.1). The presence of one-dimensional kernel has also been pointed out in [28] where the bottom topography was used as a control parameter for a steady state solution of a linear shallow-water model in a zonal channel. However, authors note the vector in kernel shows a nearly pure grid-scale noise that can be explained by technical aspects of the numerical scheme on a C grid. They point out this issue can be avoided by choosing the depth at the velocity points as control parameter.

When assimilated data are perfect and contain exact information, the minimization procedure always converges to the reference topography. In this case, it is preferable to use shorter assimilation windows because they require less CPU time due to shorter integration of direct and adjoint models on each iteration. The number of iterations necessary to assimilate the data depends on the quantity of these data only. When the simplified cost function that compares observational and model's data just at the end of assimilation window is used, almost the same number of iterations is necessary to converge in experiments with different assimilation windows because the quantity of external information introduced into the model is the same for any window in this case: just one field at the end of the window. And the same quantity of information leads to the same quantity of iterations.

When the 4D-VAR assimilation is used, new information is introduced at each model's time step. The quantity of external information is, hence, proportional to the length of the assimilation window. And in this case we see that longer windows require less iterations to converge. Each iteration becomes more efficient, but this increase of efficiency is not sufficient to compensate the increase of the CPU time necessary to perform each iteration. So, even in 4D-VAR, the total CPU time is bigger for longer windows.

When the assimilated data are noisy, more data results in a better accuracy of identified topography. It was shown that smaller time step and proportionally smaller assimilation window allows us to obtain the same precision in the reconstructed topography. However, the dependence of the residual error in topography on the quantity of assimilated data is not uniform for all windows. When assimilation window is shorter than 2 days, the residual error decreases as inverse of the square root of the window length. But when we assimilate data with windows longer than 2 days, the decrease rate of the residual becomes proportional to $T^{-3/2}$.

When the noise source is considered, we see the most dangerous noise lies in initial conditions. The same amplitude of noise in the forcing of the model and in its initial point may result in 30 times bigger error in topography in the second case. Consequently, one can think about data assimilation for the joint simultaneous identification of topography and initial point. On the other hand, the final result exhibits relatively low sensitivity to measurements errors and noise in the assimilated data.

We can state the topography identified once by data assimilation is valid for other model runs that start from other initial conditions and are situated in other parts of the model's attractor. The reconstructed topography can be used in the model for sufficiently long runs that exceed many times the length of the assimilation window. The final model's error due to inexact topography reconstructed assimilating noisy data depends linearly on the noise amplitude in these data.

Thus, we can state it is possible to use the assimilation of external data in order to reconstruct the bottom topography of a nonlinear barotropic model in the case of the time dependent motion. One open question at this time concerns possible principal difficulties related to baroclinicity and multi-layer models. In particular, optimization of topography may result in modification of the geometry of the basin at certain layers. Multi-layer model may also present theoretical particularity and invoke the question about differentiability of the model with respect to the topographic field.

Another point that has not been discussed here is the possible lack of external data. We have supposed all the necessary data are available, sea surface elevation and all velocity fields. In practice, however, only the surface height can be easily measured and assimilated.

And finally, it must by noted that the use of the bottom topography as control is only one example of a nontraditional control variable. A number of model parameters may require

such an optimization. One of them, an optimal choice of boundary conditions on rigid and open boundaries, seems to be the first necessity.

5   Acknowledgments

All the contour pictures have been prepared by the Grid Analysis and Display System (GrADS) developed in the Center for Ocean-Land-Atmosphere Interactions, Department of Meteorology, University of Maryland.

The author is grateful to the anonymous reviewer for valuable comments and constructive suggestions on the manuscript.

References

[1] A. Adcroft and D. Marshall. How slippery are piecewise-constant coastlines in numerical ocean models? *Tellus*, **50A** (1998), 95–108.

[2] D. Auroux and J. Blum. A nudging-based data assimilation method: The back and forth nudging (BFN) algorithm. *Nonlinear Process. Geophys.*, **15** (2008), 305–319.

[3] A. F. Bennett and W. P. Budgell. Ocean data assimilation and the Moan filter: Spatial regularity. *J. Phys. Oceanogr.*, **17** (1987), 1583–1601.

[4] M. Bernadou, P. L. George, A. Hassim, P. Joly, P. Laug, B. Muller, A. Perronnet, E. Saltel, D. Steer, G. Vanderborck, and M. Vidrascu. MODULEF: Une bibliothèque modulaire d'éléments finis. 2nd ed. Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, 1988.

[5] E. Blayo, J. Verron, and J.-M. Molines. Assimilation of TOPEX/POSEIDON altimeter data into a circulation model of the North Atlantic. *J. Geophys. Res.*, **99** (1994), 24691–24705.

[6] K. Brusdal, J. M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, and P. J. van Leeuwen. A demonstration of ensemble-based assimilation methods with a layered OCGM from the perspective of operational ocean forecasting systems. *J. Marine Syst.*, **40-41** (2003), 253–289.

[7] F. Bryan. Parameter sensitivity of primitive equation ocean general circulation models. *J. Phys. Oceanogr.*, **17** (1987), 970–985.

[8] F. Bryan, I. Wainer, and W. R. Holland. Sensitivity of the tropical Atlantic circulation to specification of wind stress climatology, *J. Geophys. Res.*, **100** (1995), 24729–24744.

[9] S. E. Cohn. An introduction to estimation theory. *J. Meteor. Soc. Japan. Ser. II*, **75** (1997), 257–288.

[10] J. Verron, D. T. Pham, and M. C. Roubaud. Singular evolutive extended Kalman filter with eof initialization for data assimilation in oceanography. *J. Marine Syst.*, **16** (1998), 323–340.

[11] M. Eby and G. Holloway. Sensitivity of a large-scale ocean model to a parametrisation of topographic stress. *J. Phys. Oceanogr.*, **24** (1994), 2577–2587.

[12] J. C. Gilbert and C. Lemarechal. Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Program.*, **45** (1989), 407–435.

[13] A. Gill. *Atmosphere-Ocean Dynamics*. Academic Press, London, 1982.

[14] S. Hellerman and M. Rosenstein. Normal monthly wind stress over the world ocean with error estimates. *J. Phys. Oceanogr.*, **13** (1983), 1093–1104.

[15] W. R. Holland. Baroclinic and topographic influences on the transport in western boundary currents. *Geophys. Astrophys. Fluid Dyn.*, **4** (1972), 187–210.

[16] A. M. Ilin, V. M. Kamenkovich, V. F. Kanayev, T. G. Zhugrina, and L. I. Lavrishcheva. An experiment in constructing a smoothed bottom relief of the world ocean. *Oceanologiya*, **14** (1974), 617–622.

[17] E. Kalnay, M. Kanamitsu, and W. E. Baker. Global numerical weather prediction at the national meteorological center. *Bull. Amer. Meteor. Soc.*, **71** (1990), 1410–1428.

[18] E. Kazantsev. Local Lyapunov exponents of the quasi-geostrophic ocean dynamics. *Appl. Math. Comput.*, **104** (1999), 217–257.

[19] E. Kazantsev. Sensitivity of a simple ocean model to perturbations of the bottom topography. Reseach report 6717, INRIA, 2008.

[20] F.-X. Le Dimet. A general formalism of variational analysis. Technical Report 73091, CIMMS report, Norman, OK, 1982.

[21] F.-X. Le Dimet and O. Talagrand. Variational algorithm for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A** (1986), 97–110.

[22] C. Le Provost, Ch. Bernier, and E. Blayo. A comparison of two numerical methods for integrating a quasi-geostrophic multilayer model of ocean circulations: Finite element and finite difference methods. *J. Comput. Phys.*, **110** (1994), 341–359.

[23] J. Lewis and J. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A** (1985), 309–327.

[24] J.-L. Lions. *Contrôle Optimal des Systèmes Gouvernés par des Équations aux Dérivées Partielles.* Dunod, Paris, 1968.

[25] E. N. Lorenz. Deterministic non periodic flow. *J. Atmospheric Sci.*, **20** (1963), 130–141.

[26] M. Losch, A. Adcroft, and J.-M. Campin. How sensitive are coarse general circulation models to fundamental approximations in the equations of motion? *J. Phys. Oceanogr.*, **34** (2004), 306–319.

[27] M. Losch and P. Heimbach. Adjoint sensitivity of an ocean general circulation model to bottom topography. *J. Phys. Oceanogr.*, **37** (2007), 377–393.

[28] M. Losch and C. Wunsch. Bottom topography as a control variable in an ocean model. *J. Atmos. Ocean. Technol.*, **20** (2003), 1685–1696.

[29] G. I. Marchuk. Formulation of theory of perturbations for complicated models. *Appl. Math. Optim.*, **2** (1975), 1–33.

[30] R. F. Milliff, W. G. Large, J. Morzel, G. Danabasoglu, and T. M. Chin. Ocean general circulation model sensitivity to forcing from scatterometer winds. *J. Geophys. Res.*, **104** (1998), 11337–11358.

[31] I. M. Navon, X. Zou, J. Derber, and J. Sela. Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Weather Rev.*, **120** (1992), 1433–1446.

[32] C. Nicolis, S. Vannitsem, and J.-F. Royer. Short-range predictability of the atmosphere: Mechanism for superexponential error growth. *Quart. J. Roy. Meteor. Soc.*, **121** (1995), 705–722.

[33] National Oceanic and Atmospheric Administration. Digital relief of the surface of the earth. Technical Report 88-MGG-02, NOAA, National Geophysical Data Center, Boulder, Colorado, 1988.

[34] J. Pedlosky. *Geophysical Fluid Dynamics.* 2nd ed. Springer-Verlag, 1987.

[35] T. Penduff, B. Barnier, M.-A. Kerbiriou, and J. Verron. How topographic smoothing contributes to differences between the eddy flows simulated by sigma and geopotential-coordinate models. *J. Phys. Oceanogr.*, **32** (2002), 122–137.

[36] J.-N. Thpaut and P. Courtier. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117** (1991), 1225–1254.

[37] J. Verron. Nudging satellite altimeter data into quasi-geostrophic ocean models. *J. Geophys. Res.*, **97** (1992), 7479–7491.

[38] J. Verron and E. Blayo. The no-slip condition and separation of western boundary currents. *J. Phys. Oceanogr.*, **26** (1996), 1938–1951.

[39] J. Verron and W. R. Holland. Impact de donnes d'altimtrie satellitaire sur les simulations numriques des circulations gnrales ocaniques aux latitudes moyennes. *Ann. Geophys.*, **7** (1989), 31–46.