Generalization of the dual variational data assimilation algorithm to a nonlinear layered quasi-geostrophic ocean model

Didier Auroux

HAL Id: inria-00189641

https://inria.hal.science/inria-00189641

Submitted on 21 Nov 2007

Generalization of the dual variational data assimilation algorithm to a nonlinear layered quasi-geostrophic ocean model

Didier Auroux

Institut de Mathématiques de Toulouse, UMR 5219, Université Paul Sabatier
Toulouse 3, 31062 Toulouse cedex 9, France

E-mail: `auroux@mip.ups-tlse.fr`

Abstract. In this paper, we present a generalization to nonlinear models of the four dimensional variational dual method, the 4D-PSAS algorithm. The idea of 4D-PSAS (Physical Space Analysis System) is to perform the minimization in the space of the observations, rather than in the model space as in the primal 4D-VAR scheme. Despite the formal equivalence between 4D-VAR and 4D-PSAS in a linear situation (both for model equations and observation operators), the dual method has several important advantages: in oceanographic cases, the observation space is smaller than the model space, which should improve the minimization process; for no additional cost, it provides an estimation of the model error; and finally, it does not have any singularities when the covariance error matrices tends to zero.

The idea of this paper is to extend this algorithm to a fully nonlinear situation, as it has been done in the previous years with other classical data assimilation schemes: the 4D-VAR and the Kalman filter. For this purpose, we consider a nonlinear multi-layer quasi-geostrophic ocean model, which mimics quite well the mid-latitude circulation. We recall the standard primal 4D-VAR scheme applied to this model, and then introduce an extended 4D-PSAS algorithm in the particular case of this nonlinear QG model.

We report then the results of extensive numerical experiments that have been carried out to compare this extended algorithm to the classical variational formulation, and to study its sensitivity to many parameters such as the nonlinearities, the number of available observations, the presence of an unknown term in the assimilation model, and to study the detection of the model error. As a matter of fact, it is found that this extended algorithm has kept the same advantages as in the linear case (model error detection, smaller sensitivity to various perturbations, more efficient minimization process). All these experiments suggest that it is an efficient assimilation scheme for oceanographic problems.

1. Introduction

It is now well established that the quality of ocean (and atmosphere) circulation forecasts is highly dependent on the quality of the initial conditions. Oceanic processes are governed by the general equations of fluid dynamics, and are nonlinear. Such nonlinearities are source of a huge sensitivity to the initial conditions, and then an ultimate theoretical limit to deterministic prediction. This limit is still far from being reached, and substantial gain can still be obtained in the quality of forecasts. This can be obtained through improvement of the observing system itself, but also through improvement of the geophysical models used to modelize the geophysical processes. For example, a major problem comes from the fact that sub-scales processes are associated with extremely large fluxes of properties, in particular energy. Seeking a numerical solution to the equations requires discretizing the equations, and therefore cutting off in the scales. It will be crucial to represent the fluxes associated to sub-grid processes by some additional terms in the equations [22, 36].

Over the past twenty years, observations of ocean circulation have become much more readily available, as a result of new satellite techniques and, at a lower level, large field programs. The use of altimeter measurements has provided extremely valuable information about the sea-surface height, and has allowed the oceanographic community to study more precisely both the general circulation of the ocean and the local dynamics of some particular regions (the Gulf Stream area, the Kuroshio extension, the Antarctic circumpolar current for example). Recently, the increasing amount of lagrangian float data (e.g. Argo is a set of nearly 3000 free-drifting profiling floats [2]) has greatly improved the collection of observations inside the ocean. Oceanographic data are currently extremely heterogeneous, both in nature, density and quality, but their number is still much smaller than the degree of freedom of the models. The growth of the available computing ressources indeed allows refinements of the grid size of general circulation models.

Environmental scientists are increasingly turning to inverse methods for combining in an optimal manner all the sources of information coming from theory, numerical models and data. The aim of data assimilation is precisely to combine the observations and models, in order to retrieve a coherent and precise state of the system from a set of discrete space-time data. Data assimilation covers all the mathematical and numerical techniques in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of time evolution and physical properties, and which allow us to blend as optimally as possible all the sources of information coming from theory, models, and other types of data.

There are two main categories of data assimilation techniques [43], variational methods based on the optimal control theory [28] and statistical methods based on the theory of optimal statistical estimation (see, for example, [8, 23] for an overview of inverse methods, both for oceanography and meteorology). The first class of methods (3D-VAR, 4D-VAR, 4D-PSAS, ...) was first introduced in meteorology ([27, 25, 42] for

the 3D- and 4D-VAR, [1, 4, 13] for the 4D-PSAS) and more recently for oceanic data (4D-VAR: [44, 41, 34, 40, 35, 32], 4D-PSAS: [30, 31, 8, 3]). The statistical methods (optimal interpolation, Kalman filter, particle methods, SEEK filter, . . .) were introduced in oceanography roughly fifteen years ago [18, 19].

All these techniques have first required linearized models and operators, and have been then generalized to fully nonlinear cases, except the 4D-PSAS algorithm. The Kalman filter [24, 17] has been extended to nonlinear cases through the ensemble Kalman filter (EnKF) and SEEK (singular evolutive extended Kalman) filter, and applied in oceanography to quasi-linear situations in tropical oceans [21, 15, 16, 12, 48] and to mid-latitude circulations [14, 37, 10]. In the same class of sequential algorithms, the particle methods have been recently studied in nonlinear situations [38]. The 4D-VAR algorithm has also been extended to nonlinear situations with its incremental scheme (see e.g. [39, 8]). But the 4D-PSAS algorithm has only been studied in situations in which the model equations and observation operators are linear or linearized around a reference trajectory [30, 31].

In practice, all data assimilation techniques encounter major difficulties due to computational reasons. The full Kalman filter would, in principle, require the manipulation of matrices with a dimension of typically 10^7 or 10^8 in an oceanic problem. The optimal control adjoint method often requires several hundred iterations of the minimization process to converge, thus implying an equivalent number of model runs. In this context, it is important to find new data assimilation algorithms allowing in particular a reduction of the problem dimension.

The 4D-PSAS (Physical Space Analysis System) scheme has been partly introduced and studied for this purpose and can be seen as a dual or weak formulation of the 4D-VAR algorithm. We refer to [1, 4, 13, 30, 31, 8] for a general description and study of this algorithm, and more specifically to [30, 31] for numerical studies on a primitive equation model. The idea is that the model can be seen as a weak constraint, the adjoint state being interpreted as the Lagrange multiplier of the model equations. Bennett [4, 8] introduced the so-called representers' method in linearized situations, consisting in solving iteratively and alternately some direct and adjoint equations. This approach has been applied to several ocean models, such as quasi-geostrophic and primitive equations ocean models, involving most of the time a local linearization of the state equations [5, 6, 7, 45]. In linearized situations, some weak-constraint approaches might not be convergent from a theoretical point of view, but a nonlinear situation is most of the time much less ill posed than its linearization, although this is usually difficult to prove theoretically.

From a different point of view, although it is mathematically equivalent in linear cases, it is possible to consider an optimization problem in the space of the observations, rather than in the model space as in the primal 4D-VAR scheme [1, 13]. The dual method has several important advantages in comparison to the primal method. First, in oceanographic cases, even if the number of available observations is constantly increasing, the observation space is still clearly smaller than the model space, which

speeds up the optimization process. Then, for no additional cost, it is possible to control and estimate the model error, whereas it is not possible in the 4D-VAR approach for a reasonable cost and without any reduction. Finally, the 4D-PSAS scheme requires the error covariance matrices, whereas the 4D-VAR needs their inverse and it is then not possible to theoretically have the different errors tend to zero. The 4D-PSAS has no singularities when the covariance error matrices tend to zero, and it allows us to always consider the model error, whether it is zero or not.

In this paper, we introduce an extension of the 4D-PSAS algorithm to a fully nonlinear situation. We have noticed that the presence of nonlinearities in the process could be taken into account in this scheme. For this purpose, we consider a nonlinear multi-layer quasi-geostrophic ocean model. Even if it is a simplified model, it is known to reproduce quite well the dynamics of the mid-latitude ocean circulation, at least in the statistical sense. Using this particular situation, we propose an extension of the dual algorithm that does not assume the model and operators to be linear, and in which we solve the fully nonlinear model equations. Moreover, in a linear situation, it is strictly equivalent to the 4D-PSAS algorithm.

We focus our interest on the reconstruction of a simple ocean system from altimetric surface observations. We first recall the variational adjoint method, using a strong constraint hypothesis (the ocean circulation model is assumed to be exact). The use of a cost function, measuring the mean-square difference between the observations and the corresponding model variables, allows us to carry out the assimilation process by an identification of the initial state of the ocean which minimizes the cost function [32]. It is then possible to consider the model as a weak constraint, the adjoint state being interpreted as the Lagrange multiplier of the model equations. The linear theory of duality, extended to a nonlinear case, allows us to consider the minimization problem in a dual way. The minimization of the dual cost function is then performed in the observation space, which is smaller than the state space. Moreover, there is no more need to assume that the ocean model is exact.

In this paper, this approach is shown to be able to consider nonlinearities in a more efficient way than the 4D-VAR. Moreover, some numerical experiments suggest that this extended algorithm is less sensitive to various perturbations of the process: presence of nonlinearities, number of available observations, presence of an unknown term in the assimilation model... It is numerically shown that the main advantages of the dual approach have also been preserved in the nonlinear extension (e.g. more efficient minimization process, detection of the model error).

The organization of this paper is as follows. In section 2, we introduce the equations of a nonlinear layered quasi-geostropic ocean model used for the theorical and numerical results. The primal method applied to our ocean model is given in section 3, and the extended dual scheme is introduced in section 4 for this particular nonlinear model. The results of extensive numerical simulations are given in section 5, and finally section 6 contains some concluding remarks.

2. Quasi-geostrophic ocean model

We consider here a layered quasi-geostrophic ocean model [22]. The behavior of most large scale geophysical flows is mostly due to the geostrophic equilibrium between the rotational effect and the pressure gradient. This model arises from the primitive equations (conservation laws of mass, momentum, temperature and salinity), assuming first that the rotational effect (Coriolis force) is much stronger than the inertial effects. The Rossby number, ratio between the characteristic time of the earth rotation and the inertial time, must then be small compared to 1. Secondly, the thermodynamic effects are completely neglected in this model. Quasi-geostrophy assumes that the horizontal dimension of the ocean is small compared to the size of Earth, with a ratio of the order of the Rossby number. We finally assume that the depth of the basin is small compared to its width. In the case of the Atlantic ocean, all these assumptions are not valid, notably the horizontal extension of the ocean. But it has been shown that the quasi-geostrophic approximation is fairly robust in practice, and that this approximate model reproduces quite well the ocean circulations at midlatitudes, such as the jet stream (e.g. Gulf stream in the case of the North Atlantic ocean) and ocean boundary currents [47, 9]. We finally assume that the forcing is due to the wind stress applied to the ocean surface, and that the dissipation of energy is essentially due to bottom and lateral friction.

The quasi-geostrophic model can be obtained by making a first order expansion of the Navier-Stokes equation with respect to the Rossby number. The zeroth-order development simply provides geostrophy but does not tell us anything about its time evolution, unlike the first order development which provides the time evolution of the geostrophic flow. The ocean is supposed to be stratified in n layers, each of them having a constant fluid density. The model system is then composed of n coupled equations resulting from the conservation law of the potential vorticity. The equations can be written as :

$$\frac{D_1\left(\theta_1(\Psi) + f\right)}{Dt} + A_4 \nabla^6 \Psi_1 = F_1 \qquad \text{in } \Omega \times]0, T[, \tag{1}$$

at the surface layer ($k = 1$);

$$\frac{D_k\left(\theta_k(\Psi) + f\right)}{Dt} + A_4 \nabla^6 \Psi_k = 0 \qquad \text{in } \Omega \times]0, T[, \tag{2}$$

at the intermediate layers ($k = 2, \ldots, n - 1$);

$$\frac{D_n\left(\theta_n(\Psi) + f\right)}{Dt} + A_1 \Delta \Psi_n + A_4 \nabla^6 \Psi_n = 0 \qquad \text{in } \Omega \times]0, T[, \tag{3}$$

at the bottom layer ($k = n$).

The notations are as follows :

- $\Omega \subset \mathbb{R}^2$ is the circulation basin and $]0, T[$ is the time interval,

- n is the number of layers,

- Ψ_k is the stream function at layer k, Ψ is the vector $(\Psi_1, \ldots, \Psi_n)^T$,

- θ_k is the sum of the dynamical and thermal vorticities at layer k :

$$\theta_k(\Psi) = \Delta\Psi_k - (W\Psi)_k,$$

with $-(W\Psi)_k = \dfrac{f_0^2 \rho}{H_k g}\left(\dfrac{\Psi_{k+1} - \Psi_k}{\rho_{k+1} - \rho_k} - \dfrac{\Psi_k - \Psi_{k-1}}{\rho_k - \rho_{k-1}}\right).$

- f is the Coriolis force (f_0 is the Coriolis force at the reference latitude of the ocean). In the β-plane approximation, the Coriolis force varies linearly with respect to the latitude.

- g represents the constant of gravity, ρ_k the fluid density at layer k (and ρ the average fluid density), and H_k the depth of layer k,

- $\dfrac{D_k}{Dt}$ is the Lagrangian particular derivative : $\dfrac{D_k}{Dt} = \dfrac{\partial}{\partial t} + J(\Psi_k, .),$

  where J is the Jacobian operator $J(f, g) = \dfrac{\partial f}{\partial x}\dfrac{\partial g}{\partial y} - \dfrac{\partial f}{\partial y}\dfrac{\partial g}{\partial x},$

- $\Delta\Psi_n$ represents the bottom friction dissipation, $\nabla^6\Psi_k$ represents the lateral friction (of biharmonic type) dissipation, and A_1 and A_4 are respectively the bottom and lateral friction dissipation coefficients,

- and F_1 is the forcing term, the wind stress applied to the ocean surface.

The initial conditions $\Psi_k(0)$ and some boundary conditions resulting from the mass conservation law [22, 3] complete the equations of the direct model.

3. Primal data assimilation method

3.1. Assumptions

We suppose that the data we want to assimilate come from satellite measurements of the sea-surface height, which is directly related to the upper layer stream function Ψ_1 by $h = \dfrac{f_0}{g}\Psi_1$. Thus, we assume that we have an observational stream function Ψ_1^{obs}. These observations are only available at times t_i, $i = 1\ldots N$, over the data assimilation period $[0, T]$, and are also discrete in space. We consider then that the vector $\Psi_1^{obs}(t_i)$ represents the observations of the ocean surface available at time t_i.

The control vector (which has to be determined) is the initial state of the stream functions at all layers $\Psi(0) = (\Psi_k(0))_{k=1\ldots n}$.

3.2. Cost function

We can define a cost function

$$
\begin{aligned}
\mathcal{J}\left(\Psi(0)\right) \;=\; & \frac{1}{2}\sum_{i=1}^{N}\langle R_i^{-1}\left(H_i\Psi_1(t_i)-\Psi_1^{obs}(t_i)\right), H_i\Psi_1(t_i)-\Psi_1^{obs}(t_i)\rangle \\
+\; & \frac{1}{2}\langle P_0^{-1}\left(\Psi(0)-\Psi^e(0)\right), \Psi(0)-\Psi^e(0)\rangle,
\end{aligned}
\tag{4}
$$

where P_0 and R_i are error covariance matrices for the background and observations respectively, H_i are (linear) observation operators connecting observations Ψ_1^{obs} and model solutions Ψ_1, and $\langle\,.\,,\,.\,\rangle$ is the canonical real scalar product. The background vector $\Psi^e(0)=\left(\Psi_k^e(0)\right)_{k=1\ldots n}$ is an a priori estimation of $\Psi(0)$, usually resulting from a data assimilation prevision on a previous period.

The first part of the cost function quantifies the difference between the observations and the state function, and the second part is a regularisation term (from the mathematical point of view) or a penalty term for deviations from the first-guess initial conditions (from a physical point of view). The inverse problem which consists in the minimization of $\mathcal{J}$ is then well-posed.

3.3. Gradient computation and adjoint model

In order to minimize the cost function, we need its gradient $\nabla\mathcal{J}$. Because of the large dimension of the model state vector (more than 10^6), it is not possible to compute directly the gradient by using finite difference methods. The gradient vector of the functional is then obtained by solving backwards in time the adjoint model [25]. The quasi-geostrophic adjoint equations are :

$$
\begin{aligned}
\frac{\partial\theta_1^T(\Lambda)}{\partial t}-\Delta J(\Psi_1,\Lambda_1) \;-\; & (W^T J(\Psi,\Lambda))_1 \\
& -\; J\left(\Lambda_1,\theta_1(\Psi)+f\right)+A_4\nabla^6\Lambda_1 = E_1
\end{aligned}
\tag{5}
$$

at the surface layer,

$$
\begin{aligned}
\frac{\partial\theta_k^T(\Lambda)}{\partial t}-\Delta J(\Psi_k,\Lambda_k) \;-\; & (W^T J(\Psi,\Lambda))_k \\
& -\; J\left(\Lambda_k,\theta_k(\Psi)+f\right)+A_4\nabla^6\Lambda_k = 0
\end{aligned}
\tag{6}
$$

at the intermediate layers, and

$$
\begin{aligned}
\frac{\partial\theta_n^T(\Lambda)}{\partial t}-\Delta J(\Psi_n,\Lambda_n) \;-\; & (W^T J(\Psi,\Lambda))_n \\
& -\; J\left(\Lambda_n,\theta_n(\Psi)+f\right)+A_1\Delta\Lambda_n+A_4\nabla^6\Lambda_k = 0
\end{aligned}
\tag{7}
$$

at the bottom layer, in $\Omega\times]0,T]$, where

- $\Lambda_1,\ldots,\Lambda_n$ is the adjoint vector,

- $\theta_k^T(\Lambda) = -\Delta\Lambda_k + (W^T\Lambda)_k$ is the *vorticity* corresponding to the adjoint state,

- and E_1 is the derivative of $\mathcal{J}$ with respect to Ψ_1 :

$$E_1(t) = \sum_{i=1}^{N} H_i^T R_i^{-1} (H_i\Psi_1(t) - \Psi_1^{obs}(t))\, \delta(t - t_i).$$

The coupling matrix W (see definition of the vorticity $\theta_k(\Psi)$ in section 2) can be diagonalized in $\mathbb{R}$, and we denote by P the corresponding similarity transformation matrix. The layer-mode transformation is required by the formulation of the boundary conditions, in addition to natural physical considerations [22]. We can then define the following modal adjoint vector $\chi = (\chi_1, \ldots, \chi_n)^T$:

$$\chi = P^T \Lambda,$$

and the space boundary conditions satisfied by the adjoint state Λ are then :

$$\begin{aligned}
\chi_1 &= 0 \qquad \text{in } \partial\Omega \times [0, T], \\
\int_\Omega \chi_k(t)d\sigma &= 0 \qquad \forall t \in [0, T], \quad \forall k \geq 2,
\end{aligned} \tag{8}$$

and

$$\Delta\Lambda_k(t) = 0 \qquad \text{in } \partial\Omega\times\,]0, T], \quad \forall k. \tag{9}$$

The gradient of the first part of $\mathcal{J}$ is obtained by solving equations (5-9) with a final condition of nullity of the adjoint state. The gradient of the second part of $\mathcal{J}$ is obtained directly by deriving it with respect to the control vector $\Psi(0)$, and we obtain :

$$\nabla\mathcal{J} = H(-\Delta + W)H^{-1} \begin{pmatrix} \Lambda_1(0) \\ \vdots \\ \Lambda_n(0) \end{pmatrix} + P_0^{-1}\left(\Psi(0) - \Psi^e(0)\right), \tag{10}$$

where H is the diagonal matrix with the layers' depths H_k on the diagonal.

The numerical minimization of $\mathcal{J}$ is then realized using a limited memory version of the BFGS quasi-Newton algorithm [11, 29, 3].

4. Dual method

4.1. Description of the method

The primal method has many disadvantages. First, the minimization process is often stopped before convergence to the minimum, because of the size of the state vector. Both from a theoretical and numerical point of view, as the dimension of the control vector can reach several millions, the minimization algorithm would need at least thousands of iterations to converge. It is then necessary (for computational cost reasons) to stop the algorithm after a fixed (usually a few tens) and small (particularly within an operational

point of view) number of iterations. Moreover, it is also impossible to take into account a model error : in the previous section, we have supposed that the model and the equations were perfect. This is obviously not the case (for example, not all parameters are well known). The only solution to incorporate the model error into the minimization process is to add corrective terms to the model, consider them as part of the control vector, and add a third term to the cost function. This is not computationally realistic because the size of the control vector would be multiplied by the number of time steps. Therefore, it is not possible to take into account in a straightforward way the model error in the primal variational approach.

A new approach to data assimilation problems has been proposed in the early 90's [4, 1, 13] in order to overcome these limitations. Rather than minimizing a cost function on the state space, the dual method works in the observation space (which is smaller than the state space). This method has been numerically studied in linear (or linearized) situations (see e.g. [30, 31] for its implementation in the oceanic primitive equation model MICOM), and our goal is to extend it to nonlinear cases. We present this extension in the next paragraphs.

4.2. Nonlinear generalized dual algorithm

Instead of solving first the direct equations and then the adjoint equations as in the primal variational approach, the dual method consists in solving first the adjoint equations in order to use the information contained in the observation vector, and then the direct equations in order to reconstruct a trajectory. The main issue with the dual method is the nonlinearity of the model, which makes it necessary to update the reference trajectory used to linearize the equations and the observation operators for each iteration in the adjoint state computation.

The dual algorithm for the quasi-geostrophic model can be described as follows [3]:

- Let m be an observation vector that can be directly related to Ψ_1 (assume that m is a vector containing an observation of a part of the ocean surface at different times t_i). Also assume that we have a background trajectory, in order to linearize the model and the observation operators.

- Solve the adjoint equations (resulting from a linearization of the direct model

around the background trajectory) :

$$\frac{\partial \theta_1^T(\Lambda)}{\partial t} - \Delta J(\Psi_1, \Lambda_1) \quad - \quad (W^T J(\Psi, \Lambda))_1 - J(\Lambda_1, \theta_1(\Psi) + f)$$
$$+ \quad A_4 \nabla^6 \Lambda_1 = \tilde{E}_1(m),$$

$$\frac{\partial \theta_k^T(\Lambda)}{\partial t} - \Delta J(\Psi_k, \Lambda_k) \quad - \quad (W^T J(\Psi, \Lambda))_k - J(\Lambda_k, \theta_k(\Psi) + f)$$
$$+ \quad A_4 \nabla^6 \Lambda_k = 0, \quad 1 < k < n, \tag{11}$$

$$\frac{\partial \theta_n^T(\Lambda)}{\partial t} - \Delta J(\Psi_n, \Lambda_n) \quad - \quad (W^T J(\Psi, \Lambda))_n - J(\Lambda_n, \theta_n(\Psi) + f)$$
$$+ \quad A_1 \Delta \Lambda_n + A_4 \nabla^6 \Lambda_k = 0,$$

with the final conditions $\Lambda_k(T) = 0$, $1 \le k \le n$, and where

$$\tilde{E}_1(m)(t) = \sum_{i=1}^{N} H_i^T R_i^{-1}(m(t) - \Psi_1^{obs}(t)) \, \delta(t - t_i).$$

- Solve the direct (nonlinear) equations

$$\frac{D_1 (\theta_1(\Psi) + f)}{Dt} + A_4 \nabla^6 \Psi_1 = F_1 + (Q\Lambda)_1,$$

$$\frac{D_k (\theta_k(\Psi) + f)}{Dt} + A_4 \nabla^6 \Psi_k = (Q\Lambda)_k, \qquad 1 < k < n \tag{12}$$

$$\frac{D_n (\theta_n(\Psi) + f)}{Dt} + A_1 \Delta \Psi_n + A_4 \nabla^6 \Psi_n = (Q\Lambda)_n,$$

with the initial conditions

$$\Psi_k(0) = \Psi_k^e(0) + (P_0 \Lambda(0))_k,$$

where Q and P_0 are error covariance matrices for the model and background respectively, and $\Psi_k^e(0)$ is an *a priori* estimation of $\Psi_k(0)$.

The resulting trajectory will be used as a model background for the next iteration.

- Define the operator $\mathcal{D}$ (depending on the previous linearization) on the observation space :

$$(\mathcal{D}m)(t) = \sum_{i=1}^{N} H_i \Psi_1(t_i) \, \delta(t - t_i), \tag{13}$$

where the observation operators H_i have been linearized.

We can then define the dual cost function as follows :

$$\mathcal{J}_\mathcal{D}(m) = \frac{1}{2} \langle \mathcal{D}m, m \rangle - \langle \Psi_1^{obs}, m \rangle, \tag{14}$$

where $\langle \, . \, , \, . \, \rangle$ is the canonical real scalar product in the observation space. $\mathcal{J}_\mathcal{D}$ measures the difference between $\mathcal{D}m$ and Ψ_1^{obs}, i.e. between the trace (in the observation space) of a solution of the direct model and the observation vector.
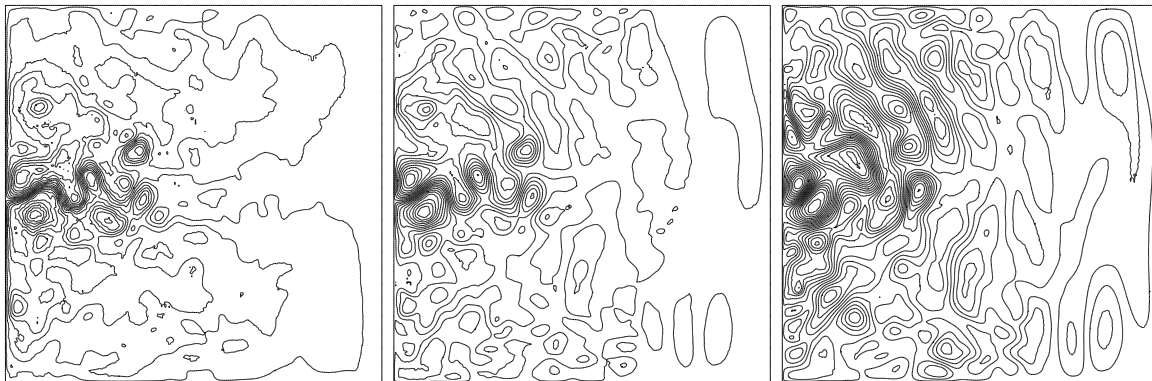
Figure 1. Initial state of the stream function for the three (from left to right: upper to bottom) layers of our quasi-geostrophic ocean model.

The minimization of $\mathcal{J}_{\mathcal{D}}$ is then performed in the same way as for the primal cost function, simply by using a quasi-Newton method such as a BFGS algorithm. Once the minimum has been found, it is easy to reconstruct the corresponding trajectory in the state space by solving (11-12).

We can observe that the minimization of the dual cost function takes place over a smaller space than the minimization of the primary one. Moreover, this method also takes into account the model error, which was numerically impossible in the classical approach.

5. Numerical results

5.1. Numerical parameters

5.1.1. Model parameters The numerical experiments are performed for a three-layered square, wind-driven, double-gyre ocean. The basin has horizontal dimensions of 4000 km × 4000 km and its depth is 5 km. The layers' depths are 300 meters for the surface layer, 700 meters for the intermediate layer, and 4000 meters for the bottom layer. The ocean is discretized by a Cartesian mesh of 201 × 201 × 3 points. The time step is 1.5 hour. The initial conditions are chosen equal to zero for a six-year ocean spin-up phase, the final state of which becomes the initial state for the data assimilation period. Figure 1 shows the initial state of the stream function for the three layers of our model. The double gyre circulation can be clearly seen on the upper layer stream function, and the six-year spin-up phase allows us to have nonlinear circulation dynamics.

Then the assimilation period starts (time $t = 0$) with this initial condition ($\Psi_k(0)$), and lasts 5 days (time $t = T$), i.e. 80 time steps. This period is quite small, although the typical length of assimilation periods is 3 to 7 days for mesoscale ocean current predictions. For example, the operational numerical prediction of Mercator Océan [33] uses a 7 days analysis window. The idea of data assimilation over longer periods is to split the assimilation period in small windows with a typical length of one week, using the analyzed state over a window as an initial guess for the next one.

Reduced gravity between layer 1 and 2	$0.0357 \; m.s^{-2}$
Reduced gravity between layer 2 and 3	$0.0162 \; m.s^{-2}$
Reference Coriolis parameter	$9.3 \times 10^{-5} \; s^{-1}$
Meridional gradient of the Coriolis parameter	$2 \times 10^{-11} \; m^{-1}.s^{-1}$
Lateral friction (biharmonic) coefficient A_4	$10^9 \; m^4.s^{-1}$
Bottom friction coefficient A_1	$10^{-7} \; s^{-1}$
Maximal amplitude of the wind stress forcing term F_1	$10^{-4} \; s^{-2}$
Rossby number	10^{-2}

Table 1. Quasi-geostrophic model parameters.

The other parameters of the model are given in table 1.

5.1.2. Minimization parameters The numerical method used to minimize the cost functions is a limited memory BFGS quasi-Newton method. The M1QN3 code by Gilbert and Lemaréchal [20] is used for our experiments. The maximum number of iterations is set to 40 because the huge size of the model doesn't allow us to obtain the exact minimum of the cost functions. The maximum number of simulations (cost function and gradient computations) is then 48. We have used the new approach introduced in [46], involving the direct BFGS formula for the diagonal preconditioner updates; the scaled version of the preconditioner is used for the minimization, but the original (unscaled) one is updated using the 5 newest pairs. This approach allows a good approximation of the inverse Hessian while improving further the minimization performance. In this context, it cannot provide an estimation of the error covariance, it only gives a measure of the local curvature of the cost function.

5.1.3. Numerical approach The experimental approach consists in performing *twin experiments* with simulated data. First, a reference experiment is run, from which the corresponding data are extracted. This reference trajectory will be further called the exact solution. Pseudo-observations are supposed to be obtained on every fifth gridpoint of the model, with a time sampling of 7.5 hours (every 5 time steps). Simulated surface data are then noised with a blank Gaussian distribution, and provided as observations for the cost function. The first guess of the assimilation experiments is chosen to be the reference state of the ocean one year before the assimilation period. The results of the identification process are then compared to the reference experiment. The covariance matrices P_0 and R_i are set proportional to the identity matrix. The background error covariance P_0 is usually the most difficult error covariance to estimate and has a crucial impact on the results [4, 8, 23]. It is usually chosen in agreement with the physical scales of the model, in order to represent the spatial structure of the errors. We should also mention that $\sqrt{P_0^{-1}}$ is often used for preconditioning the initial condition, which

is equivalent to choose $P_0 = Id$ but the difficulty is now to choose the other error covariance matrices [13]. In our study, for simplicity reasons, we choose P_0 equal to a multiple of the identity matrix, both in the primal and dual algorithms.

In all the following numerical experiments, the weights of R_i are set proportional to the level of observation noise. As the typical standard deviation of the observation noise is 10% in our experiments, we have set $R_i = 0.01\ I$ for all observation covariance matrices, where I is the identity matrix. The weight of P_0 is set to 1. We justify this choice by the fact that the error on the initial guess can be of the order of 100% of the true initial condition. Of course it is smaller than that, but the idea is to give more influence to the observations than it would be if the initial guess was very good.

Using these weights, considering that we have 17 observation times and each observation consists spatially in nearly one twenty-fifth of the upper layer, one should remark that there is a relative equilibrium between the two parts of the cost function defined in (4) in this particular situation. These weights are fixed all along the experiments, even when the number of observations is increased or decreased.

5.2. Convergence of the dual method and comparison with the primal method

*5.2.1. Convergence of the dual method* We first focus our interest on the numerical convergence of the dual algorithm we have proposed. In this part, we have used the background state of the model to initialize the process of minimizing the dual cost function. We have also used the exact (without noise) observations.

Figure 2 shows the evolution of the dual cost function (a) and its gradient (b) versus the iteration number in the minimization process. We can see that we have divided by 100 the cost function in nearly 20 iterations. The reduction is then slower. The gradient of the dual cost function is divided by nearly 10^3 in the same time, and is then nearly stable. The *exact* convergence of the minimization process is not shown, but the reduction of the cost function and its gradient is sufficient to consider the use of this method.

5.2.2. Comparison with the primal method in the linear and nonlinear cases In this section, we reduce the model dimension in order to study the loss of equivalence between the primal and dual methods in the presence of nonlinearities. We use a smaller discretization grid ($41 \times 41 \times 3$), which corresponds to a horizontal space step of 100 km. We consider a 15 year spin-up phase, after which the assimilation period starts and lasts 5 days (80 time steps). The model reduction allows us to perform a longer spin-up phase, providing a slightly larger jet stream and more ocean currents. As we are interested in the loss of equivalence between the primal and dual methods in presence of nonlinearities, the idea is to use the nonlinear model and its linearized version on this kind of turbulent states. The background is set to the real state of the system 10 days before the beginning of the assimilation period. Simulated surface data are obtained on every fifth gridpoint of the model, with a time sampling of 7.5 hours
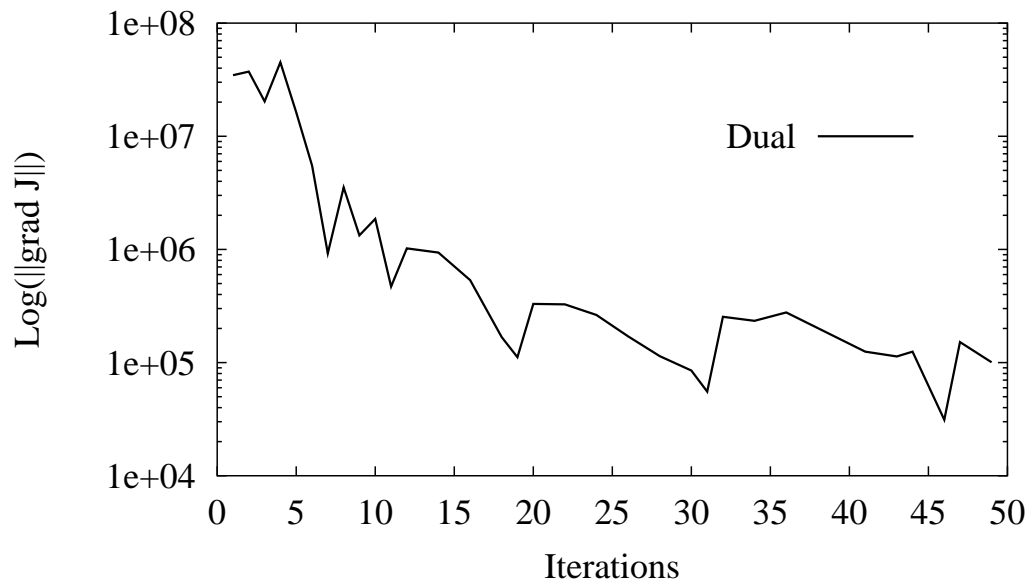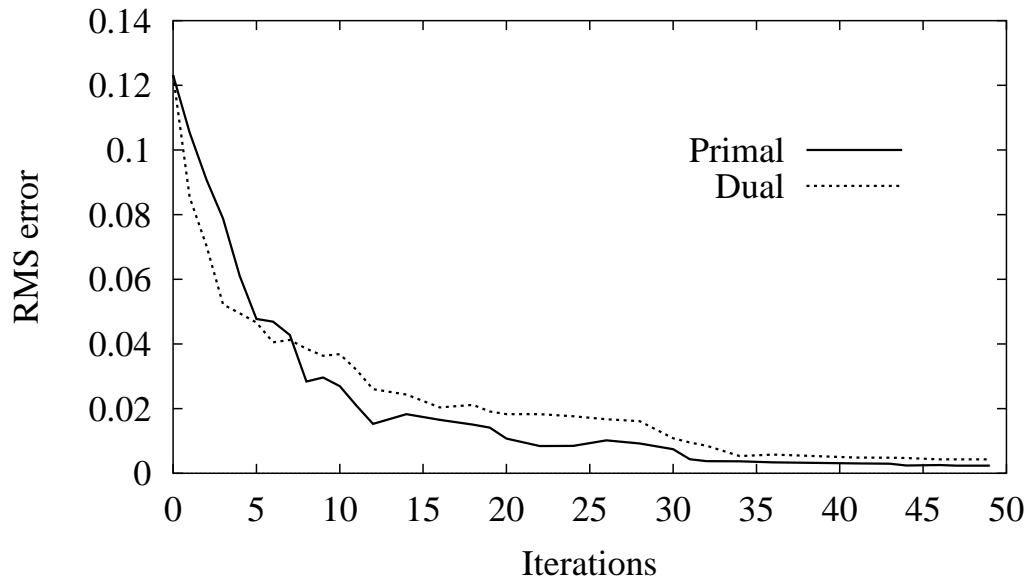
(a)
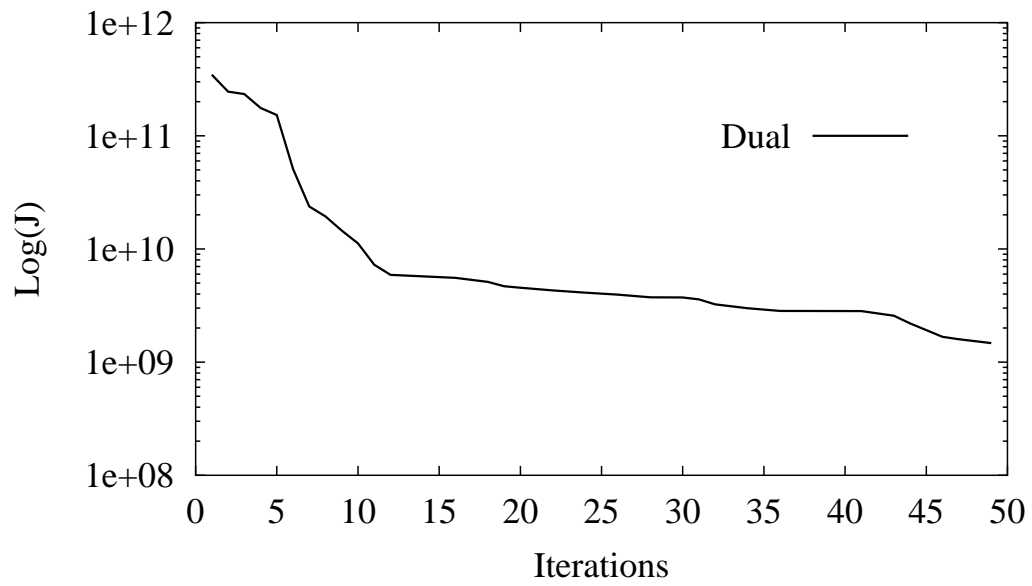
(b)

Figure 2. Evolution of the dual cost function (a) and its gradient (b) versus the iteration number in the minimization process.
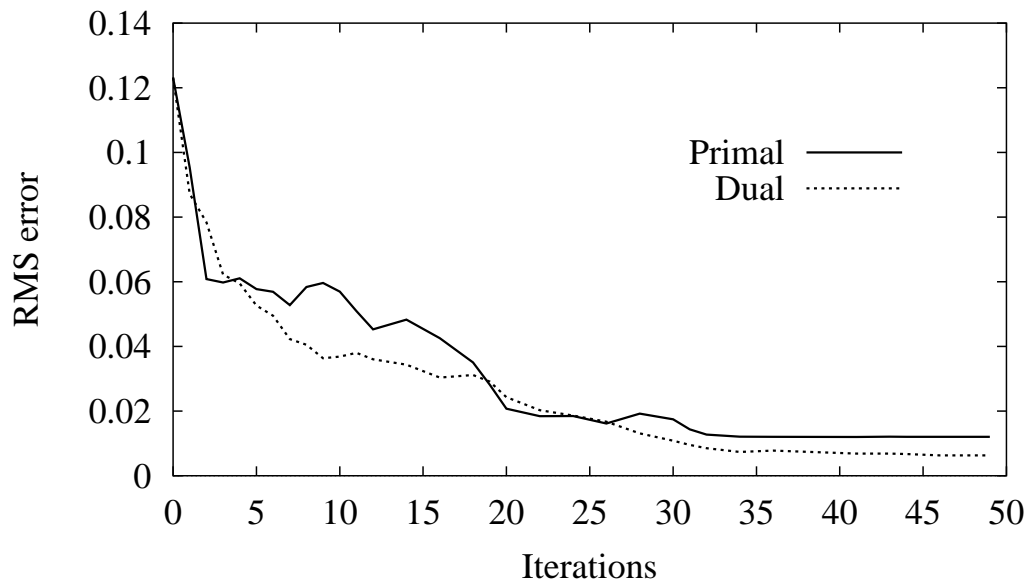
(a)

(b)

Figure 3. RMS difference between the identified and the exact initial states, for the linearized model (a) and the exact model (b).

(every 5 time steps), and are not noised.

We have compared the two methods, first using the linearized model, and then using the exact (nonlinear) model. In the linear case, the primal and dual cost functions are quadratic. The two approaches are mathematically equivalent, and hence should lead to the same minimum. This is no longer true for the exact (nonlinear) model. The minimization process is stopped after 50 iterations.

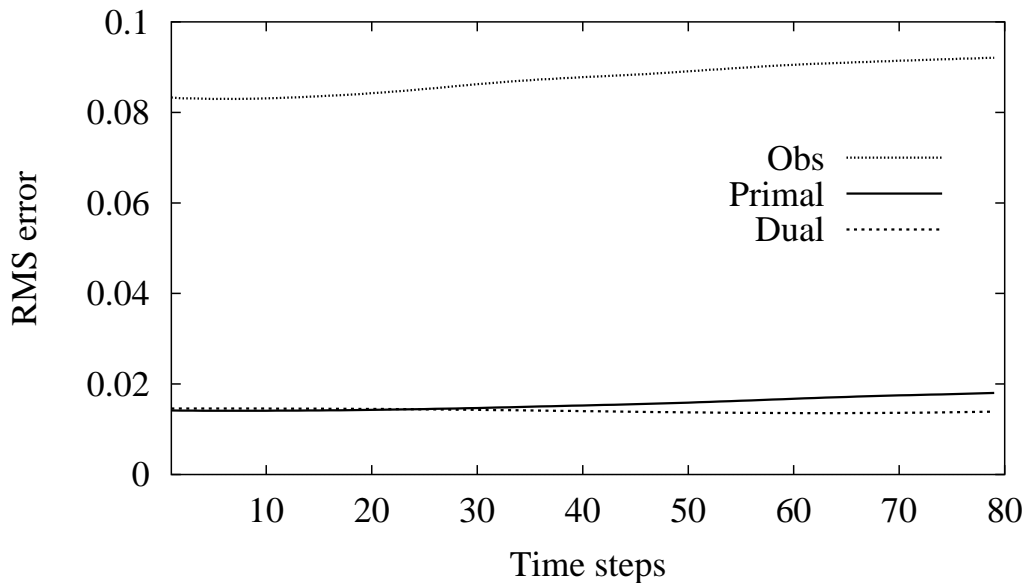Figure 3 shows for each method the RMS difference between the identified and the

Figure 4. Root mean square error between the exact solution and an identified solution over the entire assimilation period, using either the observations (dots), the primal solution (solid line), or the dual one (dashed line).

exact initial states, for the linearized model (a) and the exact model (b). The primal method gives slightly better results in the linear case, but the dual method seems to be better in the nonlinear one. We can see that the primal method is a little more sensitive to the presence of nonlinearities. It is well known in nonlinear optimization that a coarse discretization grid provides a much higher regularization than a fine grid for nonlinear problems, and hence gives better results. As the dimension of the dual space (i.e. the number of observations all along the assimilation period) is smaller than the primal one (i.e. the number of discretization points), this can partially explain the better behaviour of the dual method compared to the primal one in the nonlinear case. Finally, the differences between the linear and nonlinear cases can be explained by the loss of theoretical equivalence of the two methods in presence of nonlinearities.

5.3. Reconstruction of the initial state

We now return to the large dimension nonlinear model. The initial estimated vector at the beginning of the minimization process is chosen to be the reference state of the ocean one year before the assimilation period. In the primal approach, the minimization process is stopped after 40 iterations, each iteration consisting in one integration of the forward direct model (in order to compute $\mathcal{J}$) and one integration of the backward adjoint model (in order to compute $\nabla\mathcal{J}$). In the dual approach, the minimization process is also stopped after 40 iterations, each iteration consisting now in one integration of the backward adjoint equations and one integration of the direct nonlinear equations.

Figure 4 represents the root mean square error between the exact solution and an

identified solution over the entire assimilation period, using either the observations, the primal solution, or the dual one. The RMS error at time t for each of these solutions is :

$$rms(t) = \frac{\displaystyle\int_\Omega \left[\Psi_1^{sol}(t) - \Psi_1^{exact}(t)\right]^2 d\sigma}{\displaystyle\int_\Omega \left[\Psi_1^{exact}(t)\right]^2 d\sigma}. \tag{15}$$

The error reaches nearly 10% in the case of the trajectory resulting from the observations, and increases over time. This is due to the inherent non-linearities of the model, which propagate and increase the errors over time.

In the case of the two data assimilation methods, the RMS error is clearly smaller (by a factor of more than 5). This proves once again the usefulness of data assimilation, which makes it possible to reconstruct a trajectory with less than 2% RMS error using noised observations with a 10% RMS error. The RMS error of the primal method tends to increase over time, whereas the error of the dual solution is slightly larger at the beginning of the assimilation period but tends to remain constant over time. Moreover, as the dimension of the control vector u in the primal variational approach is $121,203$ versus $28,577$ in the dual one, the minimization of the dual cost function is faster than for the primal approach. By stopping the minimization of the dual cost function after 30 iterations, we obtain indeed almost the same decrease of the cost function as the one obtained for the primal cost function after 40 iterations. It therefore seems possible with the dual method to save time in the minimization process without any loss in the results, or equivalently to obtain better results in the same amount of time.

5.4. Sensitivity to the presence of an unknown term in the assimilation model

We now add an *unknown* term into the model equations (in all three layers). One of the main interests of the dual method is the taking into account of the model error. Hence, we now compare for each of the two methods the influence of the presence of an *error* term in the model equations used for the assimilation. The added term is layer and time independent, and at each time step, it spatially consists of a white gaussian noise corresponding to 2% or 5% of the other model terms. As the corrective term provided by the dual method is linked to the adjoint state (see equations (12)) and the model dynamics, an additive gaussian noise can be a quite large perturbation of the model. The corresponding covariance matrix Q is also set proportional to the identity matrix, with a weight equal to 10^2. This choice is justified by the RMS model error and the typical ratio between the model terms and the streamfunction values.

Figure 5 shows the RMS assimilation errors for the primal and dual methods, in presence of an additional term which noises the model by 2 or 5%. We can clearly see that the dual method is much less sensitive than the primal one. The deterioration of the primal solution is obvious and was predictable, because the primal algorithm cannot take into account the presence of an additional term in the equations, and the
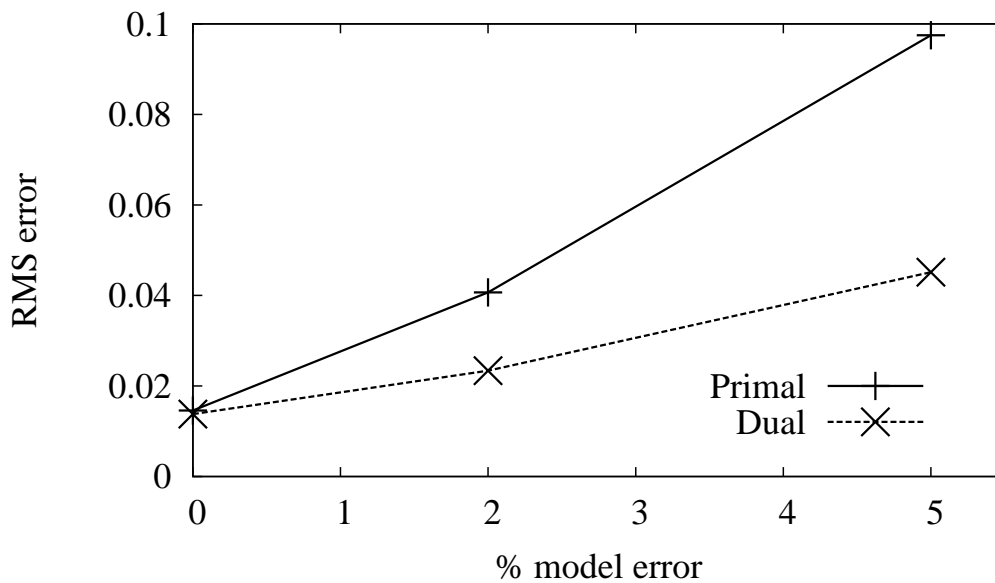
Figure 5. RMS assimilation errors for the primal and dual methods, in presence of an additional term which noises the model.

trajectories are inevitably disturbed. The dual method gives better results and seems to be less sensitive. The dual RMS error is indeed twice as small as the primal one. This illustrates one of the major advantages of the dual method. For no additional cost, it provides an additive (corrective) term in the direct model (see equations (12)). As the simulated observations were generated by the unnoised model, this corrective term (provided by the adjoint equations and the observations) partly reduces the impact of model noise, which is absolutely not the case in the primal method.

5.5. Sensitivity to the number of observations

We now focus on the influence of the number of available observations to assimilate. In the previous simulations, we had observations for every fifth gridpoint of the model and every fifth time step. We varied the availability of observations, from every space gridpoint and every time step to every twentieth gridpoint and time step. The dimension of the observation space varies then from 605 to 3, 272, 481 (i.e. 27 times more than the dimension of the state space).

Figure 6 shows the RMS difference between the identified solutions and the exact reference trajectory for the two methods, as a function of the number of available observations. There are many conclusions to be drawn from this figure. First, for both primal and dual methods, the error globally decreases strongly with the number of observations as long as the number of observations is less than the dimension of a state vector, and then remains constant or increases slightly when the number of observations becomes greater than the dimension of the space state. This may be due to the intrinsic over-determination of the problem (more observations than unknown factors to be determined), particularly in the case of the primal method, which makes
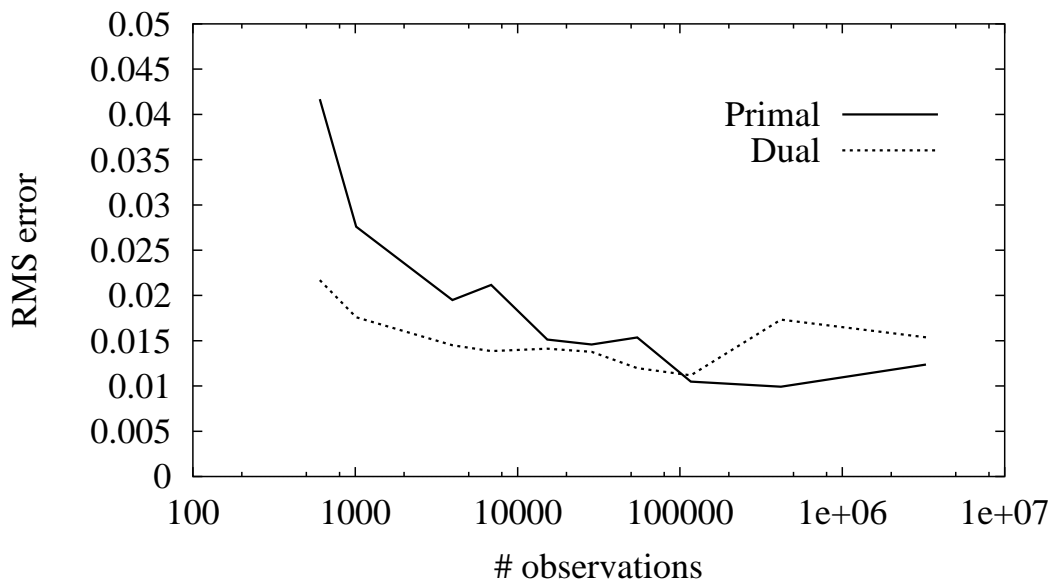
Figure 6. RMS difference between the identified solutions and the exact reference trajectory, as a function of the number of available observations.

the inverse problem relatively ill posed. The best configuration for the primal algorithm corresponds to comparable dimensions of the state and observation spaces. It has been shown in [26] that under some assumptions, the primal cost function is strictly convex if the system is fully observed, for any definite positive background matrix. Then the minimization of the primal cost function is somewhat optimal in such configuration.

We also notice that the dual method gives clearly better results than the primal method when the number of assimilated observations is small. In this case, the dimension of the observation space is quite small, and thus the minimization of the dual cost function quickly gives better results. The dual algorithm is clearly more powerful, as the corresponding error is smaller than the primal error by a factor of up to 2.

On the other side, when the number of observations is (much) larger than the size of the control vector, the primal method tends to be better. The minimization of the dual cost function may be harder because of the higher dimension of the observation space (we remind that the minimization process is stopped after a fixed number of iterations, whatever the number of observations is). Essentially, in this case, the primal method needs a large amount of observations to be efficient, while the dual method seems less sensitive to the variation of the number of available observations than the primal method.

5.6. Model error detection

In this section, we have performed twin experiments with the aim of identifying the initial condition using observations generated by a perturbed model but without any
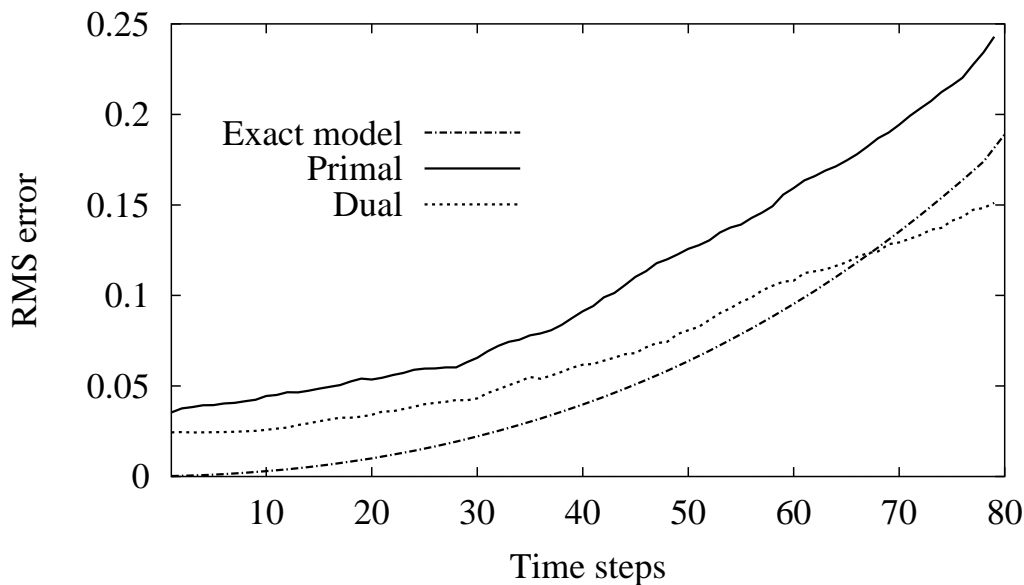
Figure 7. RMS norm of the difference between the reference trajectory and the identified trajectory versus time over the assimilation period.

observation error. We still have the same exact initial condition as before, and an experiment is run with a perturbed model (with a 2% model error; see section 5.4 for details about the noise structure), from which surface data are extracted at every fifth grid point of the model and every 5 time steps. These simulated data are not noised. From now on, we forget the model error, and we want to identify the exact initial condition using an *exact* model (without any additional term) and the data we have simulated.

The model error corresponds to the part of the equations governing the system in reality which remains unknown to this date. This enables us to generate observations derived from a different model, corrected by some a priori estimations of neglicted effects in the theoretical model. The main goal is to assimilate these observations generated by a *perturbed* model with a *theoretical* model. This allows us to determine if a variational algorithm is able to identify the solution of a model perturbed by an unknown error.

Figure 7 gives for each method the RMS norm of the difference between the reference trajectory (of the *real* model, with an unknown term) and the identified trajectory versus time over the assimilation period. We can also see on this figure the difference between the reference trajectory and the trajectory resulting from the integration of the theoretical model (with no additional term) with the exact initial condition.

It can be clearly seen that the difference between the reference trajectory to be reconstructed (resulting from the real model) and the computed trajectory resulting from the theoretical model (in which we ignore any additional term) increases strongly over time. From this remark we can deduce that, if in reality the model contains unknown terms, then not taking them into account decreases strongly the quality of the trajectories.

The primal identified solution differs from the theoretical model trajectory by a constant amount. The primal method indeed does not take into account any model error term (which exists in reality), and then gives worse results than integrating the theoretical model with the exact initial condition. On the other hand, the difference between the dual identified trajectory and the real exact solution increases less quickly than for the primal method, and moreover, at the end of the assimilation period, this difference is smaller than the difference betwen the theoretical exact trajectory and the real exact trajectory. This is the most interesting point because, by supposing that we know the exact initial condition (impossible in reality) but not the model error (which is the case in reality), we obtain worse results than by applying our dual algorithm without knowing the exact initial condition (or the model error). The dual method builds indeed a corrective term into the model using the observations. We can see this correction while looking at the difference between the direct equations solved in the primal approach (see equations $(1-2-3)$) and in the dual approach (see equation (12)). We can partially quantify the error reconstruction by measuring the relative standard deviation (RSD) between the different models. The RSD between the theoretical and perturbed models is 2% in this experiment. The RSD between the dual direct model (with the correction term provided by the adjoint state) and the perturbed model is only 1.2%. The keypoint is that the assimilated observations are derived from the perturbed model, and this systematic perturbation is then partly identified by the additive term provided by the adjoint forced by the observations (see equations $(11-12)$).

6. Conclusion

We have presented in this paper an extension to nonlinear situations of the dual variational data assimilation (4D-PSAS) scheme. We have compared this algorithm with the 4D-VAR in the particular case of a multi-layer quasi-geostrophic ocean model, which is a simple model, reproducing quite well the dynamics of the mid-latitude circulation. We have first studied the convergence of this algorithm in linear and nonlinear situations, and compared it with the 4D-VAR algorithm. The performance of these two algorithms are more or less the same, with a slight advantage for the 4D-PSAS. We have then studied the sensitivity of our algorithm to various perturbations. First, we have shown that the extended 4D-PSAS is much less sensitive to the presence of a perturbation term in the assimilation model. This is due to the intrinsic corrective term provided by the dual method in the model equations. We have then studied the impact of varying the number of available observations. The dual algorithm is less sensitive to this variation, but is also much more efficient than the 4D-VAR algorithm if the number of observations is very small, and less efficient if the number of observations is higher than the dimension of the state space. Finally, it is found that our algorithm is able to partly detect and identify the model error from the observations.

All these experiments, and in particular the last one, suggest that the nonlinear dual 4D-PSAS represents a promising method in oceanographic data assimilation. It allows

us (for no additional cost) to assimilate real data with a theoretical model, in which many physical effects have been neglicted. It is indeed possible to consider the model as a weak constraint. The nonlinear character of the equations remains a problem for the theoretical proof of convergence of this algorithm, and also for studying its comparison with the 4D-VAR. It is also necessary in the near future to apply this extended algorithm to a more complicated model based on the primitive equations, in a real configuration, and to study the behaviour of this scheme with real data.

Acknowledgments

References

[1] Amodei, L. Solution approchée pour un problème d'assimilation de données météorologiques avec prise en compte de l'erreur modèle. *C.R. Acad. Sci. Paris*, 321(II):1087–1094, 1995.

[2] Argo home page: `http://www.argo.ucsd.edu/index.html`

[3] Auroux, D. and J. Blum. Data assimilation methods for an oceanographic problem. In: *Multidisciplinary methods for analysis, optimization and control of complex systems*, Mathematics in Industry Series, Springer, 2004.

[4] Bennett, A. F. Inverse methods in physical oceanography. Cambridge University Press, Cambridge, 1992.

[5] Bennett, A. F. and M. A. Thorburn. The generalized inverse of a non-linear quasigeostrophic ocean circulation model. J. Phys. Oceanogr., 22:213–230, 1992.

[6] Bennett, A. F., B. S. Chua, and L. M. Leslie. Generalized Inversion of a Global Numerical Weather Prediction Model. Meteorol. Atmos. Phys., 60:165–178, 1996.

[7] Bennett, A. F., B. S. Chua, D. Ed. Harrison, and M. J. McPhaden. Generalized Inversion of Tropical Atmosphere-Ocean Data and a Coupled Model of the Tropical Pacific. J. Climate, 11:1768–1792, 1998.

[8] Bennett, A. F. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, Cambridge, 2002.

[9] Blayo, E., J. Verron, and J. M. Molines. Assimilation of Topex/Poseidon altimeter data into a circulation model of the North Atlantic. *J. Geophys. Res.*, 99(C12):24691–24705, 1994.

[10] Brasseur, P., J. Ballabrera, and J. Verron. Assimilation of altimetric observations in a primitive equation model of the Gulf Stream using a Singular Evolutive Extended Kalman filter. J. Mar. Syst., 22(4):269–294, 1999.

[11] Broyden, C. G. A new double-rank minimization algorithm. *Notices American Math. Soc.*, 16:670, 1969.

[12] Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi. Mapping tropical Pacific sea level: data assimilation via a reduced state Kalman filter. *J. Geophys. Res.*, 101(C10):22599–22617, 1996.

[13] Courtier, P. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, 123:2449–2461, 1997.

[14] Evensen, G. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, 97:17905–17924, 1992.

[15] Fukumori, I., J. Benveniste, C. Wunsch, and D. B. Haidvogel. Assimilation of sea surface topography into an ocean circulation model using a steady state smoother. *J. Phys. Oceanogr.*, 23:1831–1855, 1993.

[16] Fukumori, I. Assimilation of Topex sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific ocean. *J. Geophys. Res.*, 100(C12):25027–25039, 1995.

[17] Gelb, A. Applied Optimal Estimation. MA: MIT Press, Cambridge, 1974.

[18] Ghil, M. Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dyn. Atmos. Oceans*, 13:171–218, 1989.

[19] Ghil, M. and P. Manalotte-Rizzoli. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, 23:141–265, 1991.

[20] Gilbert, J.-Ch. and C. Lemaréchal. Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Prog.*, 45:407–435, 1989.

[21] Gourdeau, L., S. Arnault, Y. Ménard, and J. Merle. Geosat sea-level assimilation in a tropical Atlantic model using Kalman filter. *Ocean. Acta*, 15:567–574, 1992.

[22] Holland, W. R. The role of mesoscale eddies in the general circulation of the ocean. *J. Phys. Ocean.*, 8(3):363–392, 1978.

[23] Kalnay, E. Atmospheric Modelling, Data Assimilation and Predictability. Cambridge University Press, Cambridge, 2003.

[24] Jazwinski, A. H. Stochastic Processes and Filtering Theory. Academic, New York, 1970.

[25] Le Dimet, F.-X. and O. Talagrand. Variational algorithms for analysis and assimilation of meteorogical observations: theoretical aspects. *Tellus*, 38A:97–110, 1986.

[26] Le Dimet, F.-X., H.-E. Ngodock, B. Luong, and J. Verron. Sensitivity Analysis in Variational Data Assimilation. *J. Meteo. Soc. Japan*, 1B(75):245–255, 1997.

[27] Lewis, J. M. and J. C. Derber. The use of adjoint equations to solve a variational adjustment problem with convective constraints. *Tellus*, 37A:309–322, 1985.

[28] Lions, J. L. Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles. Dunod, 1968.

[29] Liu, D. C. and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Prog.*, 45:503–528, 1989.

[30] Louvel, S. Etude d'un algorithme d'assimilation variationnelle de données à contrainte faible. Mise en œuvre sur le modèle océanique aux équations primitives MICOM. *PhD Thesis*, Université Paul Sabatier Toulouse, 1999.

[31] Louvel, S. Implementation of a dual variational algorithm for assimilation of synthetic altimeter data in the oceanic primitive equation model MICOM. *J. Geophys. Res.* 106:9199–9212, 2001.

[32] Luong, B., J. Blum, and J. Verron. A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems*, 14:979–997, 1998.

[33] Mercator Océan home page: `http://www.mercator-ocean.fr`

[34] Moore, A.M. Data assimilation in a quasigeostrophic open-ocean model of the Gulf-Stream region using the adjoint model. *J. Phys. Oceanogr.*, 21:398–427, 1991.

[35] Nechaev, V. and M. I. Yaremchuk. Application of the adjoint technique to processing of a standard section data set: world ocean circulation experiment section S4 along 67S in the Pacific ocean. *J. Geophys. Res.*, 100(C1):865–879, 1994.

[36] Pedlosky, J. Geophysical fluid dynamics. Springer-Verlag, New-York, 1979.

[37] Pham, D. T., J. Verron, and M. C. Roubaud. A Singular Evolutive Extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, 16:323–340, 1998.

[38] Pham, D. T. .Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems *Month. Weather Rev.*, 129(5):1194–1207, 2001.

[39] Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four-dimensional assimilation. I: Experimental results with simplified physics.

Q. J. R. Meteorol. Soc., 126:1143–1170, 2000.

[40] Schröter, J., U. Seiler, and M. Wenzel. Variational assimilation of Geosat data into an eddy-resolving model of the Gulf Stream area. *J. Phys. Oceanogr.*, 23:925–953, 1993.

[41] Sheinbaum, J. and D. L. T. Anderson. Variational assimilation of XBT data. Part I. *J. Phys. Oceanogr.*, 20:672–688, 1990.

[42] Talagrand, O. and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. *Q. J. R. Meteorol. Soc.*, 113:1311–1328, 1987.

[43] Talagrand, O. Assimilation of observations, an introduction. *J. Met. Soc. Japan*, 75(1B):191–209, 1997.

[44] Thacker, W. C. and R. B. Long. Fitting dynamics to data. *J. Geophys. Res.*, 93:1227–1240, 1988.

[45] Uboldi, F. and M. Kamachi. Time-space weak-constraint data assimilation for nonlinear models. Tellus, 52(A):412–421, 2000.

[46] Veersé, F., D. Auroux, and M. Fisher. Limited-memory BFGS diagonal preconditioners for a data assimilation problem in meteorology. *Optimization and Engineering*, 1(3):323–339, 2000.

[47] Verron, J., J. M. Molines, and E. Blayo. Assimilation of Geosat data into a quasigeostrophic model of the North Atlantic between 20N and 50N: preliminary results. *Oceanol. Acta*, 15(5):575–583, 1992.

[48] Verron, J., L. Gourdeau, D. T. Pham, R. Murtugudde, and A. J. Busalacchi. An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical Pacific Ocean: method and validation. *J. Geophys. Res.*, 104:5441–5458, 1999.