A variational method for the resolution of a data assimilation problem in oceanography

Bruno Luong, Jacques Blum, Jacques Verron

A variational method for the resolution of a data assimilation problem in oceanography

Bruno Luong†, Jacques Blum† and Jacques Verron‡

† IDOPT Project (INRIA/CNRS/UJF/INPG), Laboratoire de Modélisation et Calcul/IMAG, BP 53, 38041 Grenoble Cedex 9, France

‡ Laboratoire des Écoulements Géophysiques et Industriels, UMR 5519 CNRS, BP 53X, 38041 Grenoble Cedex, France

Abstract. We consider the assimilation of satellite altimetric data into a general circulation model of the ocean at basin scale. The satellite observes only the sea-surface height of the ocean. With the assimilation of these data, we aim at reconstructing the four-dimensional space–time circulation of the ocean including the vertical. This problem is solved using the variational technique and the adjoint method. In the present case, a strong constraint approach is assumed, i.e. the quasi-geostrophic ocean circulation model used is assumed to be exact. The control vector is chosen as being the initial state of the dynamical system and it should minimize the mean-square difference between the model solution and the observed data. The assimilation procedure has been implemented and has the ability to transfer the surface data information downward to the deep flows, and hence to reconstruct the oceanic circulation in the various layers used to describe the vertical stratification of the ocean. The paper points out more specifically the crucial influence of the choice of the norm in the vector control space on the convergence speed of the optimization algorithm. Furthermore, various temporal strategies to perform the assimilation are presented and discussed with regard to their ability to properly control the initial state (which is the actual control variable) and the final state.

1. Introduction

The world's oceans play a major role in our global environment and more especially in the Earth's climate. Observations of these oceans have undergone extensive development in recent years because of the advent of new satellite techniques and especially the use of altimeter measurements. This has greatly improved our synoptical knowledge of the oceans. With the availability of Geosat satellite data and more recently of Topex/Poseidon and ERS1/2 satellite data, the oceanographic community began intensive exploitation of these new observational sources. They have already given incomparable information for the study of general ocean circulation, for estimation of the energy levels of the upper flows, and for examination of the local dynamics of different regions of particular interest, such as the Gulf Stream area, the Kuroshio extension, the Antarctic circumpolar current and the manifestations of climatic variability such as El Niño in the tropical Pacific ocean. At the same time, the modelling of the ocean system has also undergone a considerable development mostly because of formidable increases in computing capabilities. Together with climatic studies, an emerging prospect for oceanography is the challenge of the so-called operational oceanography, i.e. a similar operational system for the oceans to the one which is currently exploited in meteorological centres for weather prediction.

At the interface between the two major components of oceanographic science i.e. observations and models, lies the domain of so-called data assimilation (DA). DA covers all the mathematical and numerical techniques which allow us to blend as optimally as possible all the sources of information coming from theory, models and other types of data. (Clearly these techniques may not only apply in oceanography but to other disciplines.) DA allows us to recreate the time–space structure of a system from a set of information which has, in general, a large disparity in nature, in space–time distribution and in accuracy. There are two main categories of DA methods. Variational methods based on the optimal control theory and statistical methods based on the theory of optimal statistical estimation. The prototype of the first class which is actually of interest here is the adjoint method which was first introduced in meteorology by Penenko and Obraztsov (1976). However, the actual use of this method in numerical models is relatively recent for atmospheric sciences (see Lewis and Derber 1985, Le Dimet and Talagrand 1986, Talagrand and Courtier 1987) and even more so in the ocean (see Thacker and Long 1988, Sheinbaum and Anderson 1990, Moore 1991, Schröter *et al* 1993, Nechaev and Yaremchuk 1994). The prototype of statistical methods is the Kalman filter whose introduction in oceanography dates back roughly a decade (see, for example, Ghil 1989 and the review by Ghil and Malanotte-Rizzoli 1991). The Kalman filter was extended to nonlinear cases (Jazwinski 1970, Gelb 1974) but it has been mostly applied in oceanography to quasi-linear situations of the tropical oceans (Gourdeau *et al* 1992, Fukumori *et al* 1993, Fukumori 1995, Cane *et al* 1996, Verron *et al* 1998).

All optimal DA techniques encounter major difficulties in practice for computing reasons: memory size and computing costs. The full Kalman filter would, in principle, require the manipulation of $(N \times N)$ matrices where N is the state vector dimension which is typically 10^7 or 10^8 in an oceanic problem. The adjoint method often requires several hundred iterations for the minimization process to converge, thus implying an equivalent number of model runs.

Another crucial issue in realistic ocean problems, especially true in the mid-latitude oceans, is the handling of nonlinearity (Miller 1994). The extended Kalman filter which relies on the linear tangent assumption may be inadequate (Evensen 1992). The adjoint method, in principle, does not assume linearity but local linearization for the minimization fails in certain situations and necessitates complex strategies to circumvent the difficulty as was shown, for example, by Blum *et al* (1998).

The nature and the distribution of the observations available in oceanography strongly govern, as well, the assimilation efficiency. As was said previously, today the best observational network is provided by satellites which measure, for example, the sea-surface height along the satellite ground tracks. In spite of their unequalled density, these measurements pose specific problems, the main one being the ability to access or control the three-dimensional ocean circulation (or four-dimensional if time is added) from data which are obtained at the surface of the ocean only.

In this paper, we focus our interest on the use of the variational adjoint method in a relatively simple ocean model in order to try to reconstruct the four-dimensional ocean system from altimetric surface observations of the ocean. Here, the variational method uses the strong constraint hypothesis, i.e. the ocean circulation model is assumed to be exact. The assimilation process is carried out by an identification of the initial state of the dynamical system which minimizes a cost function. This cost function is the mean-square difference between the observations and the corresponding model variables. The functional will be minimized using a numerical unconstrained optimization method. In the present work, we lean towards the limited memory BFGS (Broyden–Fletcher–Goldfarb–Shanno) method algorithm (see Gilbert and Lemaréchal 1989). From the optimal control theory

(see Lions 1968), the gradient vector is obtained analytically from the adjoint state, which can be interpreted as the Lagrange multiplier of the model equations. This adjoint state itself is a solution of the adjoint system of the linearized partial derivative equations, which govern the dynamics of the flow-state, and hence is a backward system of PDE. If the control space is the initial state space, then the gradient vector depends only on the adjoint state at the initial instant, which is in fact the final instant of the backward adjoint system. In practice, the nonlinearities characterizing the mid-latitude ocean dynamics and the huge size of the problem have hindered considerably utilization of this method in a numerical model.

In this paper, this approach is shown to be able to control the mesoscale eddy active ocean circulation and to have the ability to transfer the surface data information downwards to the deep flows. Furthermore, we analyse how the choice of the norm of the control space affects the variational assimilation procedure. The identification of the initial state cannot be performed by a simple resolution of the variational problem over the whole data assimilation period. So, a judicious choice for the norm combined with an appropriate temporal strategy is useful in order to obtain convergence of the algorithm to the correct initial state.

The organization of this paper is as follows. Section 2 concerns the description of the assimilation problem with the quasi-geostrophic ocean model, the adjoint equations and the computation of the gradient of the cost function. In section 3, some numerical experiments on the assimilation data problem using several choices of norm will be given and commented on. In section 4, several temporal strategies will be discussed with respect to improving the efficiency of the algorithm. Finally, general conclusions will be given in section 5.

2. Description of the data assimilation problem

2.1. Equations of the oceanic model

The circulation test model used for this study is a layered, quasi-geostrophic ocean model (see Holland 1978). The dominant (geostrophic) dynamical balance between rotational effect and pressure gradient is the primary balance which governs most large-scale geophysical flows. The quasi-geostrophic model describes the temporal evolution of such geostrophic dynamical equilibrium. It is an approximate model with regard to the full primitive equation model, in particular because thermodynamics are discarded. However, it has been shown to be able to realistically reproduce the statistical properties of mid-latitudes ocean circulation including the very energetic jet and mesoscale features typical of regions like the Gulf Stream and the Kuroshio (Verron *et al* 1992, Blayo *et al* 1994).

The model system is composed of N coupled equations which result from the conservation law of the potential vorticity. These equations are written as

$$\frac{D_k(\theta_k(\Psi)+f)}{Dt} + \delta_{k,N} A_1 \Delta \Psi_N - A_2 \Delta^2 \Psi_k = F_k \qquad \forall k = 1, \ldots, N \text{ in } \Omega \times]0, T[\tag{1}$$

where

- $\Omega \subset \mathbb{R}^2$ is the circulation basin and $]0, T[$ is the time interval;
- N is the number of layers (1 for the surface layer and N at the bottom);
- Ψ_k is the stream function in layer k;
- $\theta_k(\Psi)$ is the potential vorticity in layer k. It is the sum of the dynamic vorticity and

the thermal vorticity in layer k

$$\begin{pmatrix} \theta_1(.) \\ \vdots \\ \theta_N(.) \end{pmatrix} = [\Delta . - [W].]$$

where $[W]$ is a $(N \times N)$ tridiagonal matrix whose coefficients depend only on the physical parameters

$$W_{k,k-1} = -\frac{f_0^2}{H_k g'_{k-\frac{1}{2}}} \qquad W_{k,k+1} = -\frac{f_0^2}{H_k g'_{k+\frac{1}{2}}} \qquad W_{k,k} = \frac{f_0^2}{H_k}\left(\frac{1}{g'_{k-\frac{1}{2}}} + \frac{1}{g'_{k+\frac{1}{2}}}\right)$$

where f_0 is the Coriolis force at the reference latitude of Ω, H_k is the depth of layer k at rest and $g'_{k+\frac{1}{2}} = g\,(\rho_{k+1} - \rho_k)/\rho$ is the reduced gravity, g is the Earth gravity, ρ_k is the fluid density in the layer k and ρ is the reference density.

• f is the Coriolis force. In the β-plane approximation, it varies linearly with respect to the latitude: $f(x, y) = f_0 + \beta y$, where (x, y) are Cartesian coordinates of the current point in Ω.

• $D_k./Dt$ designates the Lagrangian particular derivative operator in layer k. It is also expressed as

$$\frac{D_k.}{Dt} = \frac{\partial .}{\partial t} - \frac{\partial \Psi_k}{\partial y}\frac{\partial .}{\partial x} + \frac{\partial \Psi_k}{\partial x}\frac{\partial .}{\partial y}$$

or by

$$\frac{D_k.}{Dt} = \frac{\partial .}{\partial t} + J(\Psi_k, .)$$

where $J(.,.)$ is the Jacobian operator

$$J(\varphi, \xi) = \frac{\partial \varphi}{\partial x}\frac{\partial \xi}{\partial y} - \frac{\partial \varphi}{\partial y}\frac{\partial \xi}{\partial x}.$$

• The two terms $\Delta^l(.)$, $l = 1, 2$ in equation (1) represent, respectively, the bottom and the lateral friction dissipations. Sometimes, the lateral friction, parametrized as a biharmonic $-A_4\Delta^4\Psi_k$, is preferred.

• F_k is the forcing of the dynamical flow. F_k is the surface wind stress applied to the surface ($F_k = 0, \forall k \geqslant 2$).

2.2. Layer-mode transformation of the stream functions

The layer-mode transformation is, in addition to natural physical considerations, required by the formulation of the boundary conditions which we will give in section 2.3.

The coupling matrix $[W]$ can be diagonalized in $\mathbb{R}$ as:

$$[W] = [P].\mathrm{diag}(\lambda_1, \ldots, \lambda_N).[P]^{-1}$$

where $0 = \lambda_1 < \lambda_2 \leqslant \cdots \leqslant \lambda_N$ and $[P]$ is the similarity transformation matrix. The vector

$$\begin{pmatrix} \Phi_1 \\ \vdots \\ \Phi_N \end{pmatrix} \stackrel{\text{def}}{=} [P]^{-1}.\begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{pmatrix}$$

is called the mode vector of the stream functions.

2.3. Space boundary conditions

The boundary conditions follow directly from the mass conservation law (see Holland 1978). They are written as

$$\Psi_k(t) = C_k(t) \text{ on } \partial\Omega \quad \forall t \in [0, T], \quad \forall k = 1, \ldots, N, \text{ such that}$$
$$\text{if } \Phi_k = [P^{-1}.\Psi]_k$$
$$\text{then } \Phi_1 = 0 \text{ in } \partial\Omega \times [0, T]$$
$$\text{and } \int_\Omega \Phi_k(t)\, \mathrm{d}s = 0 \qquad \forall t \in [0, T], \quad \forall k \geqslant 2 \tag{2a}$$
$$\Delta\Psi_k(t) = 0 \text{ in } \partial\Omega \times]0, T[\qquad \forall k = 1, \ldots, N. \tag{2b}$$

The equations of the direct model are hence (1), (2*a*) and (2*b*). In this respect, we recall the following existence and uniqueness result due to Bernier (1992).

Theorem 2.1. *For $\theta(t = 0) = \theta_0$ in H^{-1}, system (1), (2a) and (2b) admits a unique solution in $C([0, T], H^{-1}) \cap L^2(0, T, L^2) \cap L^2_{\text{loc}}(0, T, H^1)$. The semigroup $G(t)$ from H^{-1} in H^{-1}, $G(t)\theta_0 = \theta(t, x)$, associated to these equations, is such that there exists a maximal attractor A which is bounded in L^2, compact and connected in H^{-1} and whose basin of attraction is the whole space H^{-1}. This attractor has finite Hausdorff and fractal dimensions. For θ_0 in H^1, system (1), (2a) and (2b) admits a unique solution in $C([0, T], H^1) \cap L^2(0, T, H^2)$.*

2.4. Assimilation data

The satellite altimeter observes the sea-surface height or dynamical topography, h, which is, in the framework of quasigeostrophy, proportional to the upper layer stream function

$$h^{\text{obs}} = \frac{f_0}{g}\Psi_1^{\text{obs}}. \tag{3}$$

The measurements are performed along the ground tracks of the satellite thus undersampling the sea-surface height in time and space according to the satellite orbit characteristics. In the case of the Topex/Poseidon satellite for example, the period of the satellite is 10 days and the ground track interval is 316 km at the equator (it decreases with the cosine of the latitude). Along the tracks, spatial sampling is approximately every 7 km. In the framework of this paper, the undersampling problem has been momentarily discarded and data are assumed to be perfectly observed everywhere on the surface. As was mentioned earlier, our aim is to blend the surface data with the corresponding variable (Ψ_1 in this case) of the ocean model for the purpose of improving the numerical reconstruction of the oceanic circulation, especially ones of the deep ocean flow.

2.5. Cost function and adjoint equation

The control vector is chosen as being the initial state of all layers. The strong constrained variational problem is to find $u = (\Psi_k(t = 0))_{k=1,\ldots,N}$ which minimizes the following least-square functional or cost function

$$\mathcal{J}_\varepsilon(u) = \frac{1}{2}\sum_{j=1}^{n}\int_\Omega (\Psi_1(t_j) - \Psi_1^{\text{obs}}(t_j))^2\, \mathrm{d}s + \frac{\varepsilon}{2}\|R(u)\|^2 \tag{4}$$

where

• u is assumed to belong to a certain admissible control space, $\mathcal{U}_{\text{ad}}$, of functions satisfying relation (2*a*);

• $t_1, \ldots, t_n \in [0, T]$ are the observation instants. In the present case, the surface stream function Ψ_1^{obs} is assumed to be observed at every instant t_j.

• $\Psi_1(t_j)$ are the corresponding quantities deduced from the solution of the so-called state equation (1), (2*a*) and (2*b*) in the terminology of optimal control theory, with the initial condition $(\Psi_k(t=0))_{k=1}^N = u$.

• The second term $\varepsilon/2\|R(u)\|^2$ in the cost function (4) is Tikhonov's regularization term (see Tikhonov and Arsenin 1977). The presence of this term is necessary because the data are noisy and the inverse problem is ill-posed. In the present nonlinear problem, several types of regularizations have been tested, the best one was found to be based on the potential vorticity

$$\|R(u)\|^2 = \sum_{k=1}^{N} H_k \left[\int_\Omega ((\Delta\Psi_k)(0) - [W]_k.(\Psi)(0))^2 \, \mathrm{d}s \right]. \tag{5}$$

The parameter ε in the cost function $\mathcal{J}_\varepsilon$ gives a measure of the influence of the regularization term compared to the quadratic difference between the observations and estimates. The choice of ε depends mostly on the observation errors: if ε is too large, too smooth fields are obtained which are inconsistent with the measurements; if ε is too small, the result is unrealistically noisy. To determine the optimal value, the generalized cross validation (GCV) method (see Wahba 1980) was used successfully.

The numerical minimization of the cost function $\mathcal{J}_\varepsilon$ can be realized by using various procedures such as the conjugate gradient methods or the quasi-Newton methods. In the present work, the M1QN3 algorithm (see Gilbert and Lemaréchal 1989) was found to be the most efficient and versatile tool.

The gradient vector of the functional is obtained by solving backwards in time the system of the following quasi-geostrophic adjoint equations

$$\begin{aligned} &\frac{\partial}{\partial t}\theta_k^t(\Lambda) - \Delta J(\Psi_k, \Lambda_k) - [W^t]_k J(\Psi, \Lambda) - J(\Lambda_k, \theta_k(\Psi) + f) \\ &\qquad + \delta_{k,N} A_1 \Delta\Lambda_N - A_2\Delta^2\Lambda_k = E_k \text{ in } \Omega\times]0, T[\qquad k = 1, \ldots, N \end{aligned} \tag{6}$$

where

• $\Lambda = (\Lambda_1, \ldots, \Lambda_N)^t$ is the N-uplet of the adjoint state. It is the unknown variable of the system (6).

• $\theta_k^t(\Lambda) = -\Delta\Lambda_k + [W^t]_k\Lambda$.

• $J(\Psi, \Lambda) = (J(\Psi_1, \Lambda_1), \ldots, J(\Psi_N, \Lambda_N))^t$.

• E_k is the derivative of the unconstrained functional with respect to Ψ_k

$$E_k(t) = \begin{cases} 0 \qquad \text{if } t \neq t_j \\ \text{else, the Dirac function } \delta_{t_j} \text{ multiplied by the spatial function} \\ \begin{cases} (\Psi_1(t) - \Psi_1^{\text{obs}}(t)) - \dfrac{1}{|\Omega|}\left[\displaystyle\int_\Omega (\Psi_1(t) - \Psi_1^{\text{obs}}(t)) \, \mathrm{d}s \right] & k = 1 \\ \dfrac{P_{11}\cdot P_{1k}^{-1}}{|\Omega|}\left[\displaystyle\int_\Omega (\Psi_1(t) - \Psi_1^{\text{obs}}(t)) \, \mathrm{d}s \right] & k \neq 1. \end{cases} \end{cases}$$

The space boundary conditions satisfied by the adjoint state Λ are:

$$\begin{aligned} &\Lambda_k(t) = C_k'(t) \text{ in } \partial\Omega \qquad \forall t \in [0, T], \quad \forall k = 1, \ldots, N, \text{ such that} \\ &\text{if } \chi_k = [P^t.\Lambda]_k \\ &\text{then } \chi_1 = 0 \text{ in } \partial\Omega \times [0, T] \end{aligned}$$

$$\text{and} \int_{\Omega} \chi_k(t)\,\mathrm{d}s = 0 \quad \forall t \in [0,T] \quad \forall k \geqslant 2 \tag{7}$$
$$\Delta\Lambda_k(t) = 0 \text{ in } \partial\Omega\times]0,T[\qquad \forall k = 1,\ldots,N.$$

We recall the following existence and uniqueness result for the adjoint state and the existence result for the optimal control (see Luong 1995).

Theorem. *For $\theta^t(\Lambda)(t = T)$ in H^{-1}, system (6) and (7) admits a unique solution in $C([0,T],H^{-1}) \cap L^2(0,T,L^2)$. Furthermore, if $\varepsilon > 0$ and if $\Psi_1^{\text{obs}}(t_j) \in H^1$ $\forall j$ then there exists an optimal control $u \in H^1$ which minimizes the cost function $\mathcal{J}_\varepsilon$.*

The mathematical problem of identifiability of the initial state from the surface data (or observability) will deserve more consideration in the future.

2.6. Computation of the gradient vector from the adjoint state

Numerical computation of the direct equations (1), (2*a*) and (2*b*) and the adjoint equations (6) and (7) is performed by using the leap-frog scheme in time and the second-order finite differences in space.

At the initial instant $t = 0$, the N functions $(\overline{\Psi}_k)_{k=1,\ldots,N}$ defined by

$$\overline{\Psi}_k \overset{\text{def}}{=} \Psi_k - \Psi_{k|_{\partial\Omega}} \qquad \forall k = 1,\ldots,N \tag{8}$$

are equal to zero on the boundary $\partial\Omega$. The set of all these functions is in a one-to-one relation to the set of all functions verifying the first relation (2*a*) of the space boundary conditions. The values of $\overline{\Psi}_k$ at the nodes which are located strictly inside the domain Ω will be used for the control of the functional $\mathcal{J}_\varepsilon$. After these nodes have been numbered, the admissible control space $\mathcal{U}_{\text{ad}}$ will be identified to $\mathbb{R}^m$ where m is the number of degrees of freedom. If $\mathcal{U}_{\text{ad}}$ is equipped with the Euclidean inner product $\| \,.\, \|_E$ of $\mathbb{R}^m$, then the gradient vector of the first term of $\mathcal{J}_\varepsilon$ is given from the adjoint state Λ, solution of (6) and (7), at an instant ($t = 0$) by the formula

$$\boldsymbol{\nabla}\mathcal{J}_0 = [H](-\Delta. + [W].)[H^{-1}] \begin{pmatrix} \overline{\Lambda}_1^0 \\ \vdots \\ \overline{\Lambda}_N^0 \end{pmatrix} \tag{9}$$

with $[H] = \text{diag}(H_1,\ldots,H_N)$, H_k is the depth of layer k at rest, and $\overline{\Lambda}_k^0 = \Lambda_k^0 - \Lambda_{k|_{\partial\Omega}}^0$. The gradient vector of Tikhonov's term is obtained by taking directly the derivative of (5) with respect to the control variable u.

2.7. Experimental strategy

The test simulation experiments are performed in a square oceanic box assuming a layered schematic stratification. The main parameters of the model are chosen to be typical of the mid-latitudes. The Coriolis parameter is $f_0 = 9.3 \times 10^{-5}$ s^{-1}. The stratification configuration chosen is three-layered with depths of 300, 700 and 4000 m. The basin has horizontal dimensions of 4000 km $\times$ 4000 km. Two main flow cases have actually been investigated.

• Case 1 considers a simple oceanic box in which a few eddies interact on the f-plane and has a relatively coarse resolution. Assuming no gradient to the Coriolis force avoids the formation of western boundary currents and associated resolution issues:

(i) the reduced gravities are $g'_{12} = 0.0357$ m s^{-2} and $g'_{23} = 0.0162$ m s^{-2}, respectively, at the interface between layers 1 (surface) and 2 and the interface between layers 2 and 3 (bottom);

(ii) the wind stress curl is made of an alternated sinusoidal, (4×4) checkerboard pattern with a maximum amplitude of 4×10^{-2} m^2 s^{-2} and a zero average;

(iii) the lateral friction A_4 has a magnitude of 10^{-5} m^4 s^{-1} and is parametrized as a biharmonic;

(iv) the bottom friction coefficient is $A_1 = 5 \times 10^{-8}$ s^{-1} and is parameterized as a linear drag for the vorticity;

• Case 2 considers a full oceanic basin on the β-plane and has a high numerical resolution of 20 km in both horizontal directions. The stratification parameters are fairly standard as are the forcing and dissipation conditions:

(i) the reduced gravities are $g'_{12} = 0.0357$ m s^{-2} and $g'_{23} = 0.0162$ m s^{-2};

(ii) the wind stress curl has a double-gyre sinusoidal structure with an amplitude of 10^{-4} m^2 s^{-2};

(iii) the lateral friction A_4 has a magnitude of 10^9 m^4 s^{-1} and is parametrized as a biharmonic;

(iv) the bottom friction coefficient is $A_1 = 10^{-7}$ s^{-1};

(v) the Coriolis parameter β is 2×10^{-11} m^{-1} s^{-1}.

In both configurations, the model flow is forced until a statistically steady-state situation is reached. This normally takes about 20 years of oceanic time. All further assimilation experiments are performed over time sequences far beyond this transitory spin-up phase.

In the first case, the absence of β-effect leads to spatial organization without any preferred direction. In the second case, the β-effect promotes the formation of two western boundary currents cyclonic in the northern gyre and anticyclonic in the southern gyre. These two currents converge at the mid-latitude to form a strong eastward current flowing in the open ocean, quite similar to the Gulf Stream (or other western boundary currents) generation process. In the first case, the eddies are unrealistically large in size (about 1000 km) with regard to ocean mesoscale eddies. Typical correlation time scale for those eddies was about 100 days. In the second case, conversely, the general features of the flow patterns are realistic from a statistical point of view even with this oversimplified geometry. The transport and energy properties of the main current and of the eddies are quite similar to the ones observed for the real Gulf Stream system, for example. The eddies are typically 200 km in diameter with a time correlation scale of a few tens of days.

The experimental approach is to perform a series of 'twin-experiments' with simulated data. It is thought to be the first stage of validating an assimilation technique, as with real observations there is no way to fully assess the performance of an assimilation experiment. In a twin-experiment investigation, two series of experiments are conducted in parallel: one reference experiment from which pseudo-data are extracted, one assimilation experiment which uses these pseudo-data and which is further compared to the reference experiment. It is possible to sample it in order to mimic the real spatio-temporal distribution of satellite altimeters. In the present tests, we have not considered this sampling problem *per se*. Data are supposed to be obtained on every gridpoint of the model with a time sampling of 1.5 days. Simulated surface data are then provided as observations for the cost function $\mathcal{J}_\varepsilon(u)$. The first guess of the assimilation experiments is chosen for the three layers of the model to be completely decorrelated from the 'true' solution. (In other words, it is an arbitrary point of the control space.) It is chosen, for example, as a very distant (in time) realization of the model. The results of the identification process, i.e. of the assimilation experiment, are then compared to the reference experiment. Note that, from the assimilation

point of view, there is a noticeable change in the order of magnitude of the control problem from case 1 to case 2 as the state vector size increases from about 5000 to about 120 000.

3. Numerical experiments using different Hilbertian norms

In this section, case 1 is considered and investigations are performed on various norms of the control space based on the stream function L^2 norm, its derivatives or a combination of the previous. Assimilation experiments have been performed with numerical quasi-Newton minimization of the cost function (4) using the four following norms for the control vector:

(a) B_1 norm:

$$\|\Psi\|^2_{B_1} = \sum_{k=1}^{N} \int_\Omega \overline{\Phi}_k^2 \, \mathrm{d}s$$

(b) B_2 norm:

$$\|\Psi\|^2_{B_2} = \sum_{k=1}^{N} \int_\Omega \|\nabla \Phi_k\|^2 \, \mathrm{d}s$$

(c) B_3 norm:

$$\|\Psi\|^2_{B_3} = \sum_{k=1}^{N} \left(\int_\Omega \|\nabla \Phi_k\|^2 \, \mathrm{d}s + \lambda_k \int_\Omega \overline{\Phi}_k^2 \, \mathrm{d}s \right)$$

(d) B_4 norm:

$$\|\Psi\|^2_{B_4} = \sum_{k=1}^{N} \int_\Omega (\Delta \Phi_k)^2 \, \mathrm{d}s.$$

In the expressions of these norms, $\overline{\Phi}_k$ represents the mode stream functions from which are deduced its scalar constants on the domain boundary. In the B_3 norm, λ_k are eigenvalues of the coupling matrix $[W]$.

From the mathematical point of view, B_1 is simply the L^2 norm, B_2 the H^1 semi-norm, B_3 a weighted H^1 norm and B_4 a H^2 semi-norm. From the physical point of view, B_1 can be related to the pressure of the flow field, B_2 to the kinetic energy of the flow, B_3 to the total energy of the flow and B_4 to the enstrophy.

The numerical method for the large scale minimization carried out in the context of this study is an unconstrained BFGS quasi-Newton method with limited memory. The M1QN3 algorithm by Gilbert and Lemaréchal (1989) is used for our experiments. This algorithm uses the BFGS formula which calculates recurrently an approximation of the Hessian at the current point of the minimization sequence. This BFGS formula requires only the knowledge of the gradient vector $\boldsymbol{G}$. Let us now point out the modification of the gradient due to the change of norm. Among these four norms, it is the discretized stream function norm, B_1, which corresponds to the 'natural' Euclidean norm. Each component G_i of the gradient vector $\boldsymbol{G}$ relatively to the norm B_1 represents the limit

$$\lim_{\alpha \to 0} \frac{\mathcal{J}_\varepsilon(\boldsymbol{X} + \alpha \boldsymbol{e}_i) - \mathcal{J}_\varepsilon(\boldsymbol{X})}{\alpha} \qquad \forall i = 1, \ldots, m$$

with $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m\}$ being the canonical basis of $\mathbb{R}^m$. In the continuous case, this limit is obtained from the adjoint state by formula (9). If we equip the control space with an inner product associated with an arbitrary norm $\|.\|_B$, and if we define the corresponding Grammian matrix

$$[B] \in \mathcal{M}(\mathbb{R}^m) \qquad \text{with } B_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle_B \tag{10}$$

then the new gradient vector, denoted by $\boldsymbol{G}_B$, of the functional $\mathcal{J}_\varepsilon$ relatively to the norm $\|.\|_B$ can be deduced from the initial gradient $\boldsymbol{G}$, corresponding to B_1 norm, by the formula:

$$\boldsymbol{G}_B = [B]^{-1} \cdot \boldsymbol{G}. \tag{11}$$

As the gradient vector of the cost function will vanish in the minimization process, then formula (11) proves clearly that a change of norm is equivalent to preconditioning on the left-hand side of the problem $\boldsymbol{\nabla}\mathcal{J}_\varepsilon = 0$ through the matrix $[B]^{-1}$. In the present situation, this relation (11) can be interpreted, in the terminology of PDE, as follows: if $\boldsymbol{G}_{B_2}$ denotes the gradient vector associated with the norm B_2, by identification of the gradient vector with the $N(=3)$ stream functions, then $\boldsymbol{G}_{B_2}$ is the solution of the following elliptic equation

$$\begin{array}{rcll} -\Delta G_{B_2,k} & = & G_k & \text{in } \Omega \\ G_{B_2,k} & = & 0 & \text{on } \partial\Omega \end{array} \qquad \forall k = 1, \ldots, N.$$

In the same way, $\boldsymbol{G}_{B_3}$ satisfies the following equation:

$$\begin{array}{rcll} -\Delta G_{B_3,k} + \lambda_k G_{B_3,k} & = & G_k & \text{in } \Omega \\ G_{B_3,k} & = & 0 & \text{on } \partial\Omega \end{array} \qquad \forall k = 1, \ldots, N.$$

As for the B_4 norm, the initial state control space $\mathcal{U}_{\text{ad}}$ could be imposed to satisfy the second relation ($2b$) of space boundary condition in addition to the first relation ($2a$). Once that is settled, then the gradient vector $\boldsymbol{G}_{B_4}$ will be the solution of the fourth order elliptic equation:

$$\begin{array}{rcll} -\Delta^2 G_{B_4,k} & = & G_k & \text{in } \Omega \\ G_{B_4,k} & = & 0 & \text{on } \partial\Omega \\ \Delta G_{B_4,k} & = & 0 & \text{on } \partial\Omega \end{array} \qquad \forall k = 1, \ldots, N. \tag{12}$$

It is clear that this supplementary boundary condition is necessary to ensure well-posedness of equation (12).

Therefore, at each iteration step of the minimization algorithm, the use of a 'non-Euclidean' norm requires us to solve an additional elliptic equation to deduce the corresponding gradient vector. From the computational point of view, the cost to solve this additional equation is negligible compared with the cost for integration of the forward direct model (1), ($2a$) and ($2b$), and of the backward adjoint equations (6) and (7).

In all these numerical experiments, illustrated below in figures 1–4, the initial estimated vector to start the minimization algorithm is the same. As mentioned earlier, this is arbitrarily chosen for the three layers of the model as being completely decorrelated from the 'reference state' (from which pseudo data are extracted). Tikhonov's regularization coefficient ε, is constant in all experiments. The issue is to evaluate the impact of various choices of norm on the efficiency of the minimization procedure.

Figures 1, 2, 3 and 4 relate respectively to the B_1, B_2, B_3 and B_4 norms. In the left pictures of these figures, the values of the two terms of the cost function $\mathcal{J}_\varepsilon$ (see formula (4)) are represented as a function of the iteration number: A-curve gives the magnitude of the quadratic difference term between the observations and the model estimates; B-curve gives the magnitude of Tikhonov's regularization term. In the right pictures, the root-mean-square (RMS) errors between the assimilated and the reference stream function fields in each layer 1, 2 and 3, are represented with respect to time for the last iteration of the minimization algorithm.

In addition, concerning the initial and final state convergences, we summarize in table 1 the RMS error values at these two initial and final instants and for the four norms. The values are normalized with regard to the reference flow and expressed in percentages.

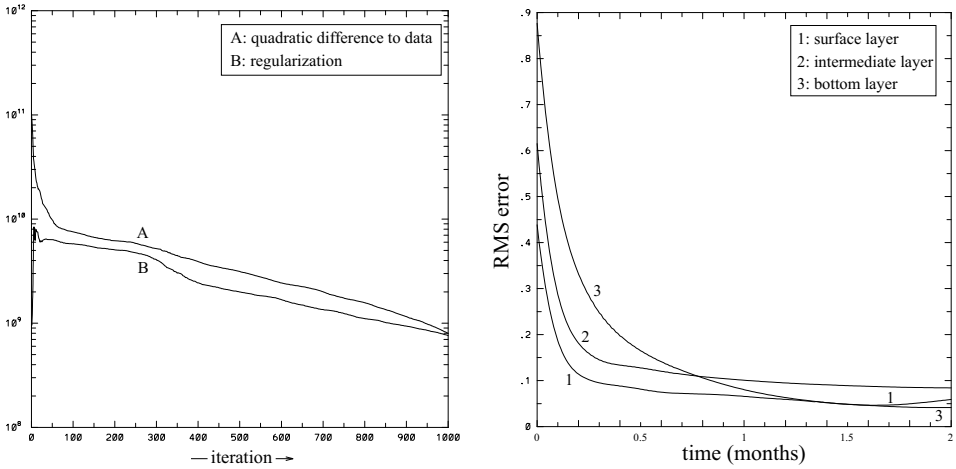

Figure 1. Minimization with the B_1 norm.

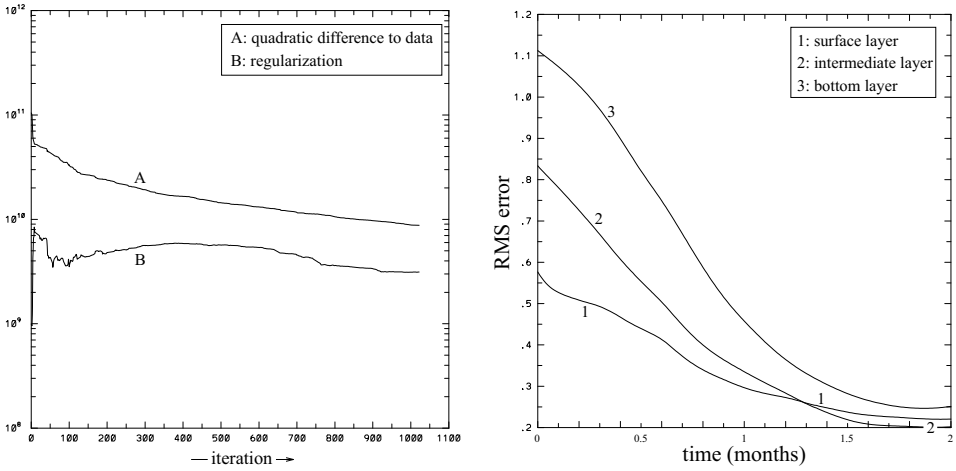

Figure 2. Minimization with the B_2 norm.

It appears clearly, especially from the table, that the B_3 norm (figure 3) enables the identified flow trajectory to be close to the 'true' flow trajectory, at the initial state as well as at the final state. Furthermore, the convergence speed of the minimization procedure, in this case, is relatively faster than for other experiments. For both B_1 and B_2 norms (figures 1 and 2), the minimization is rather inefficient and the convergence is slow. The identified trajectories remain too far from the reference trajectory, particularly at the initial state. Actually with the B_1 norm, if the iteration process is carried out further, at least up to 1600 iterations, the convergence seems to be reached with an optimum which is close to the one obtained with the B_3 norm. Finally, for the B_4 norm (figure 4), the convergence speed is irregular, and it turns out that the minimization algorithm prematurely ends, owing to an insufficiency of the decreasing speed of the cost function. More generally, it may seem natural from a physical point of view that the B_3 norm gives the best results, as it

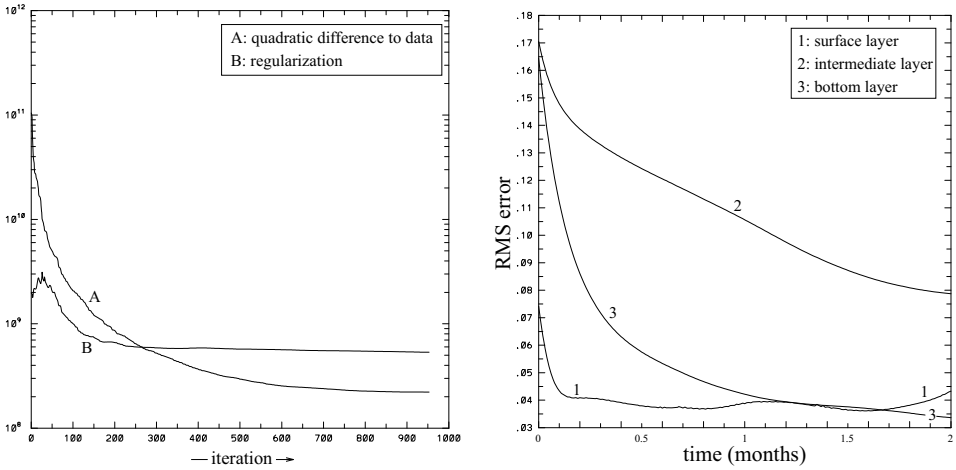

Figure 3. Minimization with the B_3 norm.

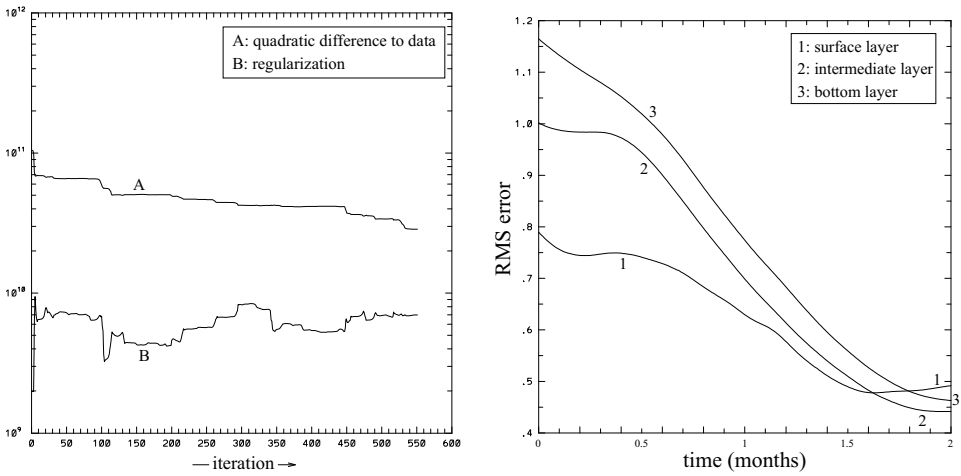

Figure 4. Minimization with the B_4 norm.

corresponds to the total energy of the system and hence should be more appropriate than a standard Euclidean norm.

4. Temporal strategies

In the case of a turbulent dissipative flow, it is well known that the space of states of the dynamical system can be split into two parts (see Temam 1988). The first part corresponds to unstable manifolds, which amplify the corresponding modes, the second to stable manifolds, which damp the corresponding modes. As a consequence, the state variables are very sensitive to a small number of modes of the control space and they are almost completely decorrelated with the other modes. Moreover these degrees of sensitivity increase with the length of the assimilation interval. This is a major drawback for the minimization algorithm

Table 1.

Norms	B_1	B_2	B_3	B_4
$t = 0$				
RMS ($k = 1$)	44%	58%	7%	79%
RMS ($k = 2$)	61%	84%	17%	100%
RMS ($k = 3$)	87%	112%	16%	116%
$t = T$				
RMS ($k = 1$)	6%	22%	4%	49%
RMS ($k = 2$)	9%	20%	8%	44%
RMS ($k = 3$)	4%	25%	3%	46%

when one has long assimilation periods even using an appropriate norm for the control space. The minimization algorithm modifies only some modes of the 'first-guess' and the process has difficulty in converging towards the desired state during the first few iterations.

On the other hand, if one splits the assimilation period in too small sub-intervals, over which the cost function is minimized in a sequential way, the propagation of the information from the surface to the bottom of the basin is not satisfactory. Therefore, surface information is not properly transmitted for controlling the deep flows. A good compromise was found using intervals, the length of which is a fraction of the decorrelation scale of the model dynamics.

The effective convergence of the assimilation process is a trade-off between two contradictory requirements for the length of the assimilation sequence: (i) it must be shorter than the predictability scale for the global trajectory optimization to be meaningful, but (ii) it must be long enough for the deep flow control to be effective. An efficient procedure is to sample the assimilation in an adequate number of sub-sequences. The variational assimilation is achieved on each of these sub-sequences, the duration of which satisfies the previous criterion. These results are of potential importance for altimetry in which the periodicity of the observational sampling is usually not much smaller than the predictability scale of the extra-tropical ocean.

With regard to this general approach of splitting the assimilation period in several sub-intervals, several strategies are possible. The first strategy is to split the period of assimilation in several sub-periods and to make sequentially the assimilation over each sub-period. The initial guess of one sub-interval is the final state of the previous sub-interval. Several examples are shown on figure 5 where the relative RMS errors between the reference state and the assimilation experiment are presented for each layer. Each experiment corresponds to a specific length of sub-sequences: 16 sequences of 0.5 month, eight sequences of 1 month, four sequences of 2 months, two sequences of 4 months and a standard minimization over 8 months. This experiment is carried out with the norm B_3 and only 15 iterations are used for each sub-sequence in time. A good compromise for the duration of the sub-sequences for the sequential strategy is about 2 months. This sequential strategy is poor concerning the identification of the initial state but gives a relatively good result for the estimation of the final state.

A second strategy is a progressive one: the assimilation is first made over some short sub-period. The identified initial state is given as the first guess for an identification over a double period. Hence, more and more data are assimilated in time, as the procedure goes on until all the available data are taken into account. The results of this procedure are represented on figure 6 for the three layers. This progressive method enables us to

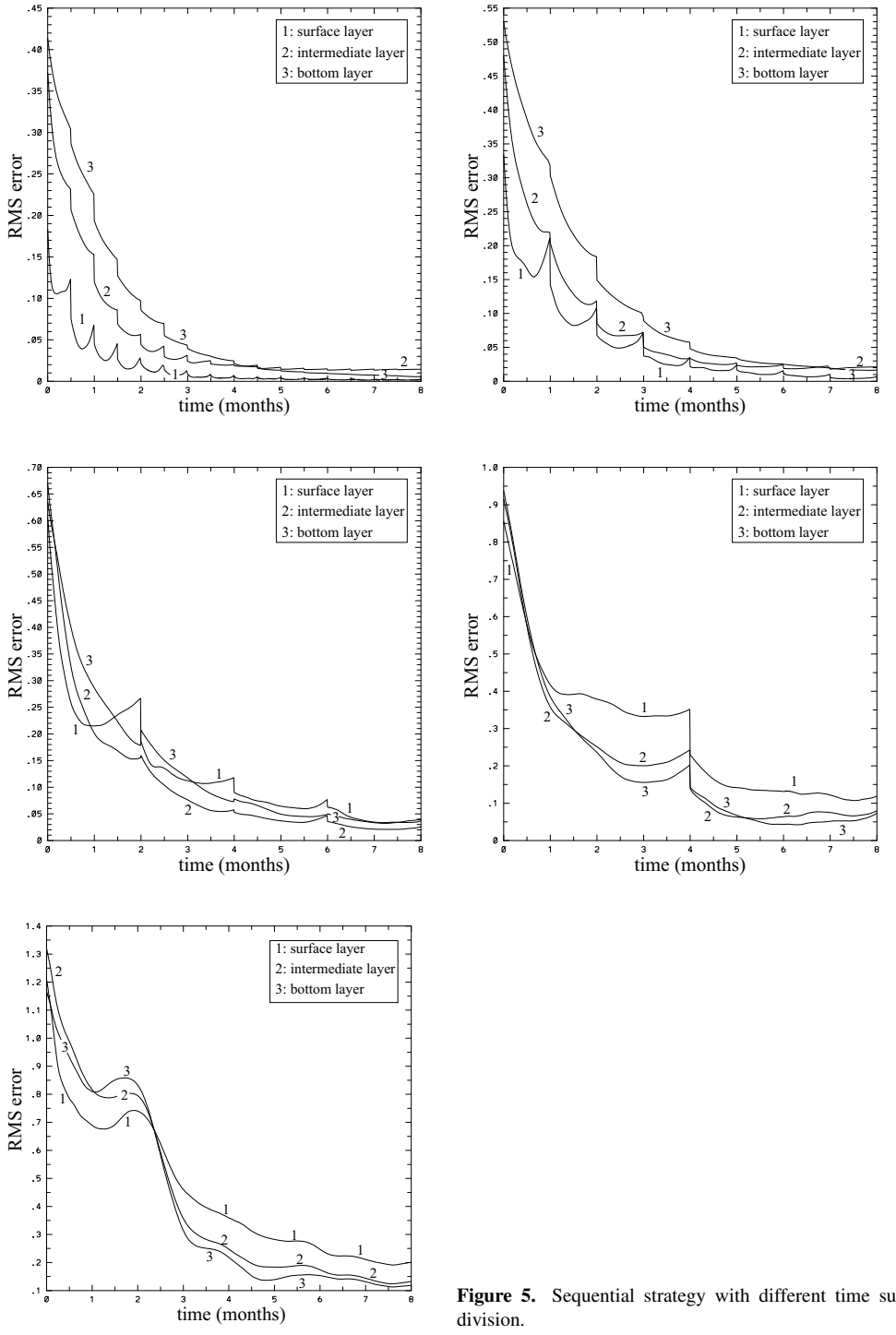

Figure 5. Sequential strategy with different time subdivision.

propagate the surface information to the bottom, the results being better and better as the length of assimilation increases. This progressive procedure is more time-consuming than

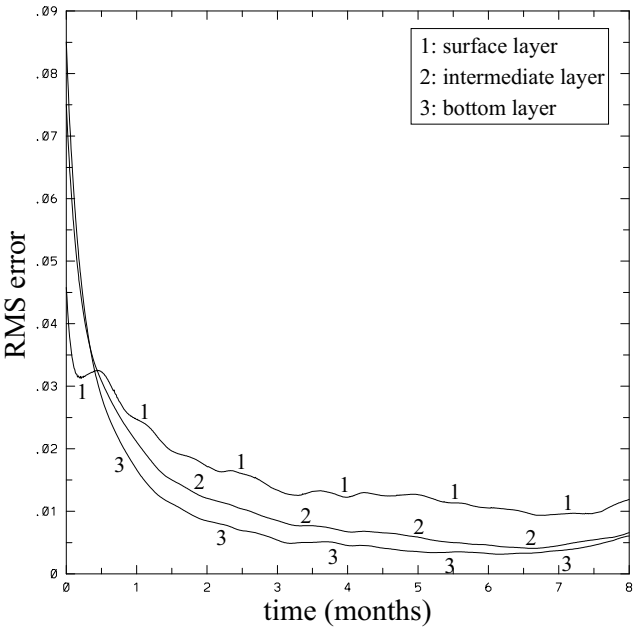

Figure 6. Final convergence using progressive strategy.

the sequential one. Typically the price to pay is two times larger. However, the RMS on the initial state is much better (less than 10%) while the RMS on the final state is still excellent (about 1%). Figure 7 completes this view of the progressive process by decomposing the assimilation process in time sequences and by representing the results for each layer. For example, the accuracy on the identification of the initial conditions for the deep layer improves from 45% for the first assimilation step to 8% for the fourth step.

The progressive approach gives better identification for the initial control state. However, the first sequential strategy may be sufficient if one only aims at properly recovering the flow fields at the final stage of the time sequence.

We conclude this section by giving some results relative to case 2, i.e. to a large size problem which has the dimension of a realistic problem, for example for the simulation of the North Atlantic circulations. Illustrations of the corresponding flows are given in figure 8. It is clear from figure 8 that, thanks to a correct choice of the norm of the control space and to the progressive assimilation strategy, the initial state is correctly identified, even in the bottom layer, whereas only surface data have been assimilated. In this experiment, each minimization iteration over the temporal assimilation period of 2 months consumes 83 s CPU time on a C98 Cray computer. A complete assimilation process needs an equivalent of 500 iterations to converge, consuming about 12 h CPU time.

5. Conclusion

This paper shows that the problem of identifying an oceanic flow field using only surface information can be solved by the variational adjoint method for the quasi-geostrophic ocean model. However, the solution does not follow directly from a simple application of the method and imposes some constraints on the practical achievement of the assimilation.

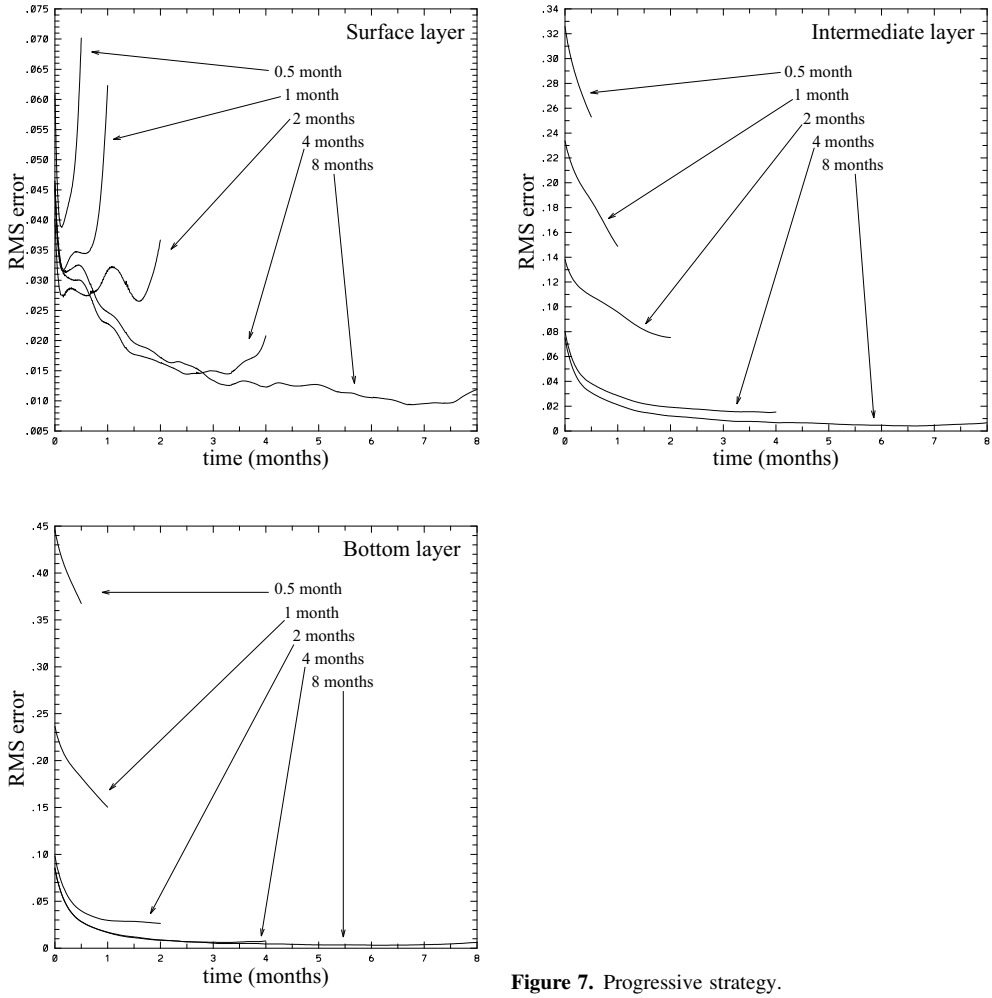

Figure 7. Progressive strategy.

The assimilation is formulated as an optimal control problem where the initial state is the control vector and where the cost function represents the quadratic difference between observations and model predictions. It is shown that the choice of the norm of control space is fundamental to obtaining good minimization convergence. A simple use of the Euclidean norm may fail to overcome various obstacles to convergence, speed of convergence and most probably local minima. In particular, this has important practical consequences for realistic size problems where the computing costs involved are considerable. In the present work, the norms tested belong to a relatively simple class (stream function and/or its derivatives L^2 norms). In the altimeter data assimilation context, we can claim that the norm which is associated to the total energy of the system is the correct norm for the identification of the initial state.

One key aspect in the success of assimilation is its ability to transfer the surface data information downwards to the deep flows. It is suggested that an efficient assimilation strategy can be constructed by dividing the global time sequence in several time sub-periods, the individual duration of which must be less than the typical predictability time scale of the flow. With this sequential approach, an acceptable accuracy can be reached for

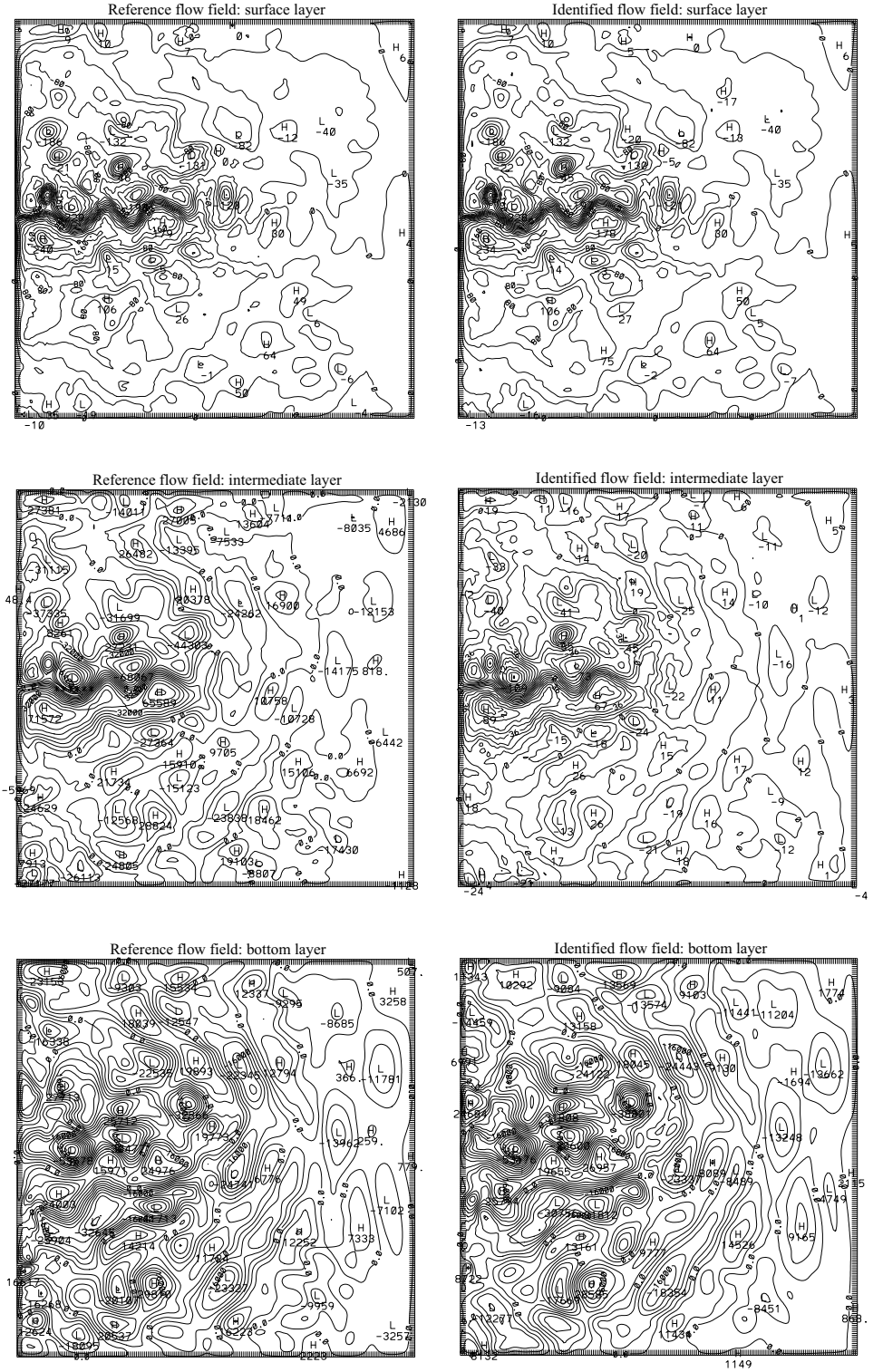

Figure 8. Comparison of the identified flow stream function fields to the reference flow fields.

the recovery of the final flow state, but the identification of the initial state is poor. The progressive strategy consists of assimilating data over time periods of increasing length, still aiming at identifying better and better the initial state. It is approximately twice as time consuming as the sequential one. However, in this approach, the identification is excellent for the final state as well as for the initial state.

In the present work, synthetic data are sampled using the whole surface layer. Generally, in the framework of the realistic oceanic data assimilation, the data are available only along ground tracks for time intervals corresponding to the satellite repeat period. However, in practice, the observation data are various in nature and should be combined together in the same functional to be minimized. Therefore, the optimal initial state would not be as well estimated because of the relatively small number of observations and their heterogeneous spatial distribution. Also, we notice that the performances of these methods have been assessed with a quasi-geostrophic model. It would be interesting to apply them to a more complicated model such as the primitive equation model and to check whether the initial condition for these equations can be reconstructed from the surface observations.

Acknowledgments

We are indebted to J C Gilbert and C Lemaréchal for providing us with their efficient M1QN3 minimization algorithm and for helpful discussions. The calculations were made possible by the computing facilities at the Centre Grenoblois de Calcul Vectoriel (CGCV-CEA) and the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS).

References

Bernier C 1992 Étude et parallélisation d'un code d'éléments finis pour la modélisation quasigéostrophique des circulations océaniques *Thèse doctorat* l'Université Joseph Fourier, France

Blayo E, Verron J and Molines J M 1994 Assimilation of Topex/Poseidon altimeter data into a circulation model of the North Atlantic *J. Geophys. Res.* **99** (C12) 24 691–705

Blum J, Luong B and Verron J 1998 Variational assimilation of altimeter data into a non-linear ocean model: temporal strategies *ESAIM Proc.* to appear

Cane M A, Kaplan A, Miller R N, Tang B, Hackert E C and Busalacchi A J 1996 Mapping tropical Pacific sea level: data assimilation via a reduced state Kalman filter *J. Geophys. Res.* **101** (C10) 22 599–617

Evensen G 1992 Using the extended Kalman filter with a multilayer quasigeostrophic ocean model *J. Geophys. Res.* **97** 17 905–24

Fletcher R 1986 *Practical Methods of Optimization* vol 1 (New York: Wiley–Interscience)

Fukumori I, Benveniste J, Wunsch C and Haidvogel D B 1993 Assimilation of sea surface topography into an ocean circulation model using a steady state smoother *J. Phys. Oceanogr.* **23** 1831–55

Fukumori I 1995 Assimilation of Topex sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific ocean *J. Geophys. Res.* **100** (C12) 25 027–39

Gelb A 1974 *Applied Optimal Estimation* (Cambridge, MA: MIT Press) p 374

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

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

Gilbert J C and Lemaréchal C 1989 Some numerical experiments with variable storage quasi-Newton algorithms *Math. Prog.* B **45** 407–35

Gourdeau L, Arnault S, Ménard Y and Merle J 1992 Geosat sea-level assimilation in a tropical Atlantic model using Kalman filter *Ocean. Acta* **15** 567–74

Holland W R 1978 The role of mesoscale eddies in the general circulation of the ocean—numerical experiments using a wind-driven quasigeostrophic model *J. Phys. Oceanogr.* **8** 363–92

Jazwinski A H 1970 *Stochastic Processes and Filtering Theory* (New York: Academic) p 376

Le Dimet F X and Talagrand O 1986 Variational algorithms for analysis and assimilation of meteorological observation: theoretical aspects *Tellus* **38A** 97–110

Lewis J M and Derber J C 1985 The use of adjoint equations to solve a variational adjustment problem with convective constraints. *Tellus* **37A** 309–22

Lions J L 1968 *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles* (Paris: Dunod–Gauthier–Villars)

Luong B 1995 Techniques de contrôle optimal pour un modèle quasigéostrophique de circulation océanique. Application à l'assimilation variationnelle des données altimétriques satellitaires *Thèse Doctorat* Université Joseph Fourier, France

Miller R N 1994 Perspectives on advanced data assimilaton in strongly nonlinear systems *Data Assimilation (NATO ASI Series, vol I 19)* ed P Brasseur and J C J Nihoul (New York: Springer) p 253

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

Nechaev V and Yaremchuk M I 1994 Application of the adjoint technique to processing of a standard section data set: world ocean circulation experiment section S4 along 67° S in the Pacific ocean *J. Geophys. Res.* **100** (C1) 865–79

Penenko V V and Obraztsov N N 1976 A variational initialization method for the fields of the meteorological elements *Sov. Meteorol. Hydrol.* **11** 1–11

Schröter J, Seiler U and Wenzel M 1993 Variational assimilation of Geosat data into an eddy-resolving model of the Gulf Stream area *J. Phys. Oceanogr.* **23** 925–53

Sheinbaum J and Anderson D L T 1990 Variational assimilation of XBT data. Part I *J. Phys. Oceanogr.* **20** 672–88

Talagrand O and Courtier P 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory *Q. J. R. Meteorol. Soc.* **113** 1311–28

Temam R 1988 *Infinite Dimensional Dynamical Systems in Mechanics and Physics* (New York: Springer)

Thacker W C and Long R B 1988 Fitting dynamics to data *J. Geophys. Res.* **93** 1227–40

Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-posed Problems* (New York: Wiley)

Verron J, Molines J M and Blayo E 1992 Assimilation of Geosat data into a quasigeostrophic model of the North Atlantic between 20° N and 50° N: preliminary results *Oceanologica Acta* **15** 575–83

Verron J, Gourdeau L, Pham D T, Murtugudde R and Busalacchi A J 1998 An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical Pacific: method and validation *J. Geophys. Res.* submitted

Wahba G 1980 Cross validation and constrained regularization methods for mildly ill posed problems *Proc. Int. Conf. on Ill Posed Problems* ed M Z Nashed (New York: Academic)