ESAIM: Proceedings, Vol. 4, 1998, 21-57
Contrôle et Équations aux Dérivées Partielles
http://www.emath.fr/proc/Vol.4/

VARIATIONAL ASSIMILATION OF ALTIMETER DATA INTO A NON-LINEAR OCEAN MODEL: TEMPORAL STRATEGIES

J. BLUM

Projet IDOPT[1]
tour IRMA, LMC, BP 53X
38041 Grenoble Cédex, France

B. LUONG

Projet IDOPT[1]
tour IRMA, LMC, BP 53X
38041 Grenoble Cédex, France

J. VERRON

LEGI, UMR 5519
BP 53X, 38041 Grenoble Cédex, France

Key Words : Data Assimilation, Optimal Control, Oceanography, Inverse Problems Quasi-geostrophic Model..

AMS Subject Qualification : 35Q35, 49N50.

[1]IDOPT is a joint project between CNRS, Université Joseph-Fourier, Institut National Polytechnique de Grenoble and INRIA

Article published by EDP Sciences and available at http://www.edpsciences.org/proc or http://dx.doi.org/10.1051/proc:1998019

Abstract

In this paper, we explore the use of the adjoint method in the oceanographic context for assimilating satellite altimeter data. Experiments with simulated altimeter data are performed in a multi-layer quasigeostrophic ocean model. The interest is mainly with controlling the mesoscale eddy active ocean circulation observed in the mid-latitudes. Due to the surface nature of the observations, one key aspect in the success of assimilation is its ability to transfer the surface data information downwards to the deep flows. Test experiments are performed first, in a coarse resolution regional basin within which several eddies are interacting on the f-plane and second, in a high resolution basin size domain on the β-plane which mimics the Gulfstream-like behavior of mid-latitude jets and western boundary currents and the associated eddy system.

As a matter of fact, it is found that the length of the assimilation cycle is crucial to the success of this assimilation. Short assimilation cycles may be efficient in the control of surface flows but rather ineffective with respect to the downward penetration of information. Conversely, long assimilation cycles lead to rather coherent results in terms of efficiency on the vertical but this global efficiency is poor, especially identifying the initial control state.

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. However, the whole assimilation cycle must be long enough and larger than the vertical penetration time scale. A "sliced assimilation" strategy which satisfies these conditions is studied in which the assimilation period is divided into equal-length time-intervals. With this approach, an acceptable accuracy can be reached for the recovery of the final flow state. This meets the requirements of the "filtering" objectives of data assimilation and therefore the "forecast" purposes. A so-called "progressive assimilation" strategy is also studied in which successive iterations of increasing durations are performed in order to progressively improve the control of the initial state control variable. This more expensive strategy is more adequate for "smoothing" objectives.

1 Introduction

Over the past fifteen years, the new satellite techniques for observing the oceans, and especially the use of altimeter measurements, have greatly improved our knowledge of the oceans by allowing synoptic and continuous monitoring of the sea surface. Despite the short life of Seasat, its measurements of the sea-surface height have clearly demonstrated the feasibility and the usefulness of satellite altimetry. But it was with the availability of Geosat data between 1985 and 1989, and since August 1992, of Topex/Poseidon data, that the oceanographic community began intensive exploitation of these new observational sources. They have already given incomparable informations to study the general circulation of the ocean, to estimate the energy levels of the upper ocean, and to examine the local dynamics of different regions of particular interest, such as the tropical oceans, the Gulf Stream area, the Kuroshio extension and the Antarctic circumpolar current. The interested readers can for instance refer to the two special issues of the Journal of Geophysical Research on Geosat (Volume 95, Numbers C3 and C10 of March and October 1990) and to the recent one on Topex/Poseidon (Volume 99, Number C12, December 15, 1994). The new ERS satellite missions (ERS1 launched in July 1991 and ERS2 launched in April 1995) are the next stages in the setting up of an hopefully continuous, satellite observing system of the oceans.

In order to make a best use of these observations for validation and interpretation purposes and also sometimes with a view to predictive modelling objectives, ocean modellers have started to use data assimilation techniques. In this respect, meteorologists have a long term experience and operational weather prediction centers are currently operating increasingly efficient data assimilation schemes. In the ocean, interest for data assimilation quickly grew in the recent years but technical development and applications are still in their infancy. This interest is mainly due to the noticeable increase in data availability, especially due to satellites. These observational capabilities will render possible the setting up of oceanographical forecasting systems, just as it is for weather forecasting purposes. In this regard, the design of efficient assimilation techniques is compelling. However, the relative paucity of observations remains a major problem for the ocean together with the fact that the dominant scales in the ocean are much smaller than the atmospheric ones. As an example, there are about two orders of magnitude difference between the first Rossby radius of deformation of the two mediums. As a whole, the degrees of freedom of the ocean are widely undersampled in comparison to the atmosphere and this justifies even more the need for an optimal use of the observations at our disposal.

Several different data assimilation techniques have been considered in meteorology and oceanography. Two mains directions are clearly identifiable in the methodlogical approaches: the statistical methods issued from the statistical estimation theory and the variational methods originated in the optimal control theory (see for example the review by *Ghil and Manalotte-Rizzoli* [1991]). Of interest here is the variational one and more especially the so-called adjoint methods which are particularly efficient to mathematically solve the variational

problem. Adjoint methods seem to have been first introduced into meteorology by *Penenko and Obraztsov* [1976] but their use in numerical models had only occured recently owing in particular to *Lewis and Derber* [1985], *Le Dimet and Talagrand* [1986], *Talagrand and Courtier* [1987] for the atmospheric sciences and to *Thacker and Long* [1988], *Sheinbaum and Anderson* [1990], *Moore* [1991], *Schröter et al.* [1993], *Nechaev and Yaremchuk* [1994], *Morrow and De Mey* [1995] and others.

In this paper, the question of the use of variational assimilation in a mid-latitudes ocean model is addressed for the case of altimetric observations. In this context, a key issue is to assess the ability of assimilating surface-only altimeter data to control deep flows. The general matter of the observability of altimeter data into ocean models is indeed for a large part an unresolved issue. Several studies have shown that altimeter data can profitably be used for constraining, at least partially, the flow in the deep ocean. In the quasigeostrophic models with relatively few modes on the vertical, the problem can be properly solved even with simple methods [*De Mey and Robinson*, 1987; *Verron*, 1990, 1992; *Haines*, 1991]. With primitive equations models, especially high vertical resolution ones, the extrapolation of surface data is more delicate although several solutions have already been proposed [*Hurlburt*, 1986; *Mellor and Ezer*, 1991; *Ezer and Mellor*, 1994; *Cooper and Haines*, 1995; *Oschlies and Willebrand*, 1995]. Still much work is clearly required on this subject, in particular through exploring the comparative capabilities of the various existing methods in the same spirit as *Ikeda et al.* [1995] for example.

A second fundamental issue is to evaluate the impact of non-linearity on the efficiency of the assimilation procedure. Non-linearity is indeed an ubiquitous characteristic of the ocean dynamics and is even a prominent feature of mid-latitudes ocean currents. The formal generalization of linearized data assimilation methods to nonlinear systems is possible. The Kalman filter generalizes to the Extended Kalman filter. The variational methods formulation is in principle not limited to linear systems. With both approaches however, difficulties have been experienced in practice. In the EKF, it may not be justified to neglect the higher degree moments for the evolution of the error covariance [*Evensen*, 1992; *Gauthier et al.*, 1993; *Miller et al.*, 1994]. With variational methods, the cost function is no longer quadratic as soon as the model is non-linear. If it stays approximately quadratic in the neighborhood of the minimum all along the assimilation cycle, there might not have too much effects. But if there is a significant uncertainty on the initial solution for example, this quadratic approximation may not be valid and leads to various forms of obstacle to convergence (local minimums, conditioning, ...). The fact that the differences between the estimate state vector of the system and the real state vector are not too large and that local linearization is possible, is usually referred to as the tangent linear appproximation. The question of its validity is highly relevant to non-linear models. Rather limited amount of studies have been undertaken to investigate these limitations and possible cares in the framework of variational assimilation [*Gauthier*, 1992; *Tanguay et al.*, 1995]. In this latter paper, the authors examine a fully turbulent barotropic atmospheric flow on the β-plane. Interestingly,

they clearly show that the length of the assimilation period strongly influences the efficiency of the transfer of data through the spectrum of flow scales in the variational assimilation process.

In this paper, we place ourself in the context of variational data assimilation of altimeter data into stratified, non-linear ocean models. Academic configurations are being considered as well as simulated observations. The two major issues previously described are considered, namely how surface information from altimetric measurements can be propagated downwards to control the deep ocean and how variational assimilation is dealing with the flow non-linearities associated with the mid-latitudes mesoscale eddies.

In Section 2, we quickly present the quasi-geostrophic ocean model used and the basics of the variational assimilation by the adjoint method. In Section 3, the identical twin-experiment approach is described. In Section 4, 5 and 6, we successively discuss the results of various assimilation numerical experiments carried out according to different strategies in term of time sequencing; i.e. the standard assimilation, the so-called sliced assimilation and the progressive assimilation. A conclusion is presented in Section 7.

2 The model and the adjoint technique for assimilation.

The numerical ocean model used for this study is a quasigeostrophic model which is basically the one originally derived by *Holland* [1978]. This model has been shown to be able to realistically reproduce the statistical properties of mid-latitudes ocean circulations including the very energetic jet and mesoscale features typical of regions like the Gulf Stream and the Kuro Shio for example [*Schmitz and Holland*, 1982]. The basic mathematical formulation comes from the standard quasigeostrophic system of equations familiar to fluid geophysicists for the atmosphere and the ocean [*Pedlosky*, 1987]. In the present case it is used in its layered version: the stratification of the ocean is schematized in dividing the full ocean depth in N layers, each layer k is supposed of constant density ρ_k ($k = 1, \ldots, N$). It is then possible to define the reduced gravity $g'_{k+1/2}$ at the interface $k + 1/2$ between layers k and $k + 1$ by

$$g'_{k+1/2} = g\,\frac{\rho_{k+1} - \rho_k}{\rho_0} \qquad k = 1, \ldots, N-1$$

where g is the gravity and ρ_0 the average density. The k-layer depth at rest is denoted by H_k. The familiar β-plane approximation is used so that the Coriolis parameter f can be written $f(y) = f(y_0) + \beta(y - y_0)$, where y_0 corresponds to the basin mid-latitude and $\beta = df/dy(y_0)$.

The model equations are the classical quasigeostrophic equations which can be written:

$$\frac{D}{Dt}\left[\Delta\psi_k + f + \frac{f_0}{H_k}\left(h_{k+1/2} - h_{k-1/2}\right)\right] = F_k + D_k, \quad k = 1, \ldots, N \qquad (1)$$

where D/Dt is the Lagrangian operator $\partial/\partial t + u\,\partial/\partial x + v\,\partial/\partial y$. The model variables are the streamfunction ψ_k and the vorticity $\omega_k = \Delta\psi_k$ in each layer.

The horizontal components of the velocity V in the layer k are

$$u = -\frac{\partial \psi_k}{\partial y} \qquad \text{and} \qquad v = \frac{\partial \psi_k}{\partial x}$$

The variable $h_{k+1/2}$ is the interface deformation height. It could be expressed as

$$h_{k+1/2} = \frac{f_0}{g'_{k+1/2}}(\psi_{k+1} - \psi_k) \qquad k = 1, \ldots, N-1$$

with $h_{1/2} = 0$, and $h_{N+1/2} = h_B$ where h_B is the bathymetric height over the bottom. The forcing term F_k corresponds to the wind stress over the surface layer:

$$F_1 = \frac{1}{\rho_0 \, H_1} \, \text{curl}\tau$$

where τ is the wind stress over the ocean surface, and $F_k = 0$ for all $k > 1$. The subgrid scales are represented by a lateral friction of harmonic (Laplacian) or biharmonic type respectively

$$D_k = A_2 \, \Delta^2 \psi_k \qquad \text{or} \qquad D_k = -A_4 \, \nabla^6 \psi_k \qquad k = 1, \ldots, N-1$$

In the deepest layer, a bottom friction term is added which also contributes to the dissipation processes

$$D_N = A_2 \, \Delta^2 \psi_N - C_b \, \Delta \psi_N \qquad \text{or} \qquad D_N = -A_4 \, \nabla^6 \psi_N - C_b \, \Delta \psi_N$$

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

$$h = \frac{f_0}{g} \psi_{Obs}$$

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 Topex/Poseidon for example, the period of the satellite is 10 days and the ground track interval of 316 km at the equator (it decreases with the cosine of the latitude). Along the tracks, the spatial sampling is approximately every 7 km.

Such altimeter data are thus assimilated into the model using the adjoint method which is,as already said, directly stemmed from the optimal control theory: Equation (1) represents the so-called "state equation" in the terminology of optimal control (see *Lions*, [1968]) and is supposed to be a "strong constraint" of the assimilation. The result of the assimilation will therefore be an exact solution of the dynamical model or in other terms, the model is supposed to be a perfect representation of reality. The assimilation process consists of seeking to minimize some criterion error to be defined, between the model prediction and the observations, according to some control variable. This error is usually referred as the "cost-function" or the "distance criterion". A simple cost-function can be the mean-square difference between the model solution and

observed data. Rather classically in meteorology and oceanography, the control variable is chosen to be the flow state vector at the initial instant. It should also be boundary conditions on any freely tunable parameter of the model or possibly mixing of these.

In the present situation, the adjoint method is used to solve the variational problem. It will not be presented in great details here but more explanations can be found for example in *Luong* [1995] and *Luong et al.* [1997]. The adjoint method actually transforms a constrained minimisation problem in the physical model space into an unconstrained minimisation problem in the space of the control variables. The adjoint state of equation (1) is solved backwards in time and enables us to compute the gradient of the cost function. This gradient with regard to the control variable actually depends on the adjoint state at time $t = 0$. The adjustment of the control variable is progressively obtained by iterations until convergence where the minimum of the cost function is being reached. It must be said that running such variational assimilation is computationnally expensive since several iterations are necessary for the minimization process, each of this iteration requiring one integration of the direct model and one integration of the adjoint.

Thus, the control vector u is here chosen as being the initial state $\psi_k(t = 0)$ of the N layers which minimizes the trajectory of the model to the data. This would lead to the cost function:

$$J(u) = \sum_j \int_{\mathcal{L}_1} \Big[\psi_1(t_j) - \psi_{Obs}(t_j)\Big]^2 \, dx \tag{2}$$

where $\mathcal{L}_1$ is the surface layer, t_j the instants of observation, ψ_{Obs} the value of the surface streamfunction deduced from the altimetric measurements.

The ill-posedness of the inverse problem is well known, that is to say the instability of the optimal value u with respect to perturbations or errors on ψ_{Obs} [*Lavrentiev*, 1967]. This instability may classically be avoided by adding a second term to the cost-function as a "regularization" or "smoothing" term [*Tikhonov and Arsenin*, 1977]. Instead of minimizing the simple cost-function given by (2), one actually considers a cost-function of the type:

$$J(u) = \sum_j \int_{\mathcal{L}_1} \Big[\psi_1(t_j) - \psi_{Obs}(t_j)\Big]^2 \, dx + \epsilon \parallel R(u) \parallel^2 \tag{3}$$

where R is a Tikhonov regularization operator which is often written as a derivation operator at some order of the control vector. In general the derivation order is chosen large enough so as to insure the inverse problem stability. The derivation order is also clearly related to the smoothness of the control vector with regard to observations, i.e. to the control of small scales. In the present nonlinear problem, several types of regularization have been tested, the best one was found to be the L^2 norm of the relative vorticity, i.e. the enstrophy, and is used all along the paper [*Luong et al.*, 1997]:

$$J(u) = \sum_j \int_{\mathcal{L}_1} \Big[\psi_1(t_j) - \psi_{Obs}(t_j)\Big]^2 \, dx + \epsilon \sum_{k=1}^{N} \int_{\mathcal{L}_k} \omega_k^2(u) \, dx \tag{4}$$

The parameter ϵ in the cost-function $J(u)$ gives a measure of the influence of the regularization term, compared to the quadratic difference between observations and estimates. The choice of ϵ depends mostly on the observation errors and has great influence on the quality of the estimation: 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. Note that the physical dimension of ϵ is m^2 in the MKS system. To determine the optimal value, the generalized cross-validation (GCV) method [*Wahba*, 1980] can be used. In practice, we used it only occasionnally as it was found computationally prohibitive (typically 30 times the cost of one typical assimilation step). In other situations, the choice of ϵ was tuned empirically on the basis of the optimum values provided by the GCV method in similar cases and on the smoothness of the control flow vorticity fields.

The minimization of the cost-function may be realized using various procedures such as the conjugate-gradient methods or the quasi-Newton methods. In the present work, the $M1QN3$ quasi-newton limited memory algorithm by *Gilbert and Lemaréchal* [1989] was found to be the most efficient and versatile tool. In order to improve the efficiency of the algorithm, several preconditioning procedures have been tested which rely on the Hilbertian norm of the control space. As a conclusion of a specific investigation, best results were obtained by taking a norm for the control vector involving the velocity field and the streamfunction [*Luong et al.*, 1997].

As the state equation is non-linear, the cost-function $J(u)$ is not quadratic with respect to u. Therefore, the minimization algorithm is less inclined to provide a rapid convergence and even, in some cases, may converge towards local minima and prevents the global minimum of the cost-function from being reached. If multiple minima are present, this also means that the result will depend on the initial starting point of the procedure. In any case, the convergence of the minimization is slow and much less effective than in a corresponding linearized situation for example. This issue actually appears as the main obstacle to a satisfactory solution in our assimilation experiments. In relation to these difficulties, the definition of an adequate convergence criteria may not be a trivial task. In general, the stopping criteria was based on the relative error on the gradient of the cost function. This was found to be more effective than other criterions such as the ones based on the residuals of the control state vector or on the residuals of the cost function.

Error estimate of the control state vector is not a direct by-product of the calculation. It can be estimated using the Hessian of the cost function with respect to these control parameters (the variance-covariance matrix of the estimation error is actually the inverse of the Hessian matrix). But the calculation of the full Hessian matrix and its inversion are prohibitive in terms of computer time and memory size. Therefore, the full error analysis is not usually performed whereas it is part and parcel of the statistical estimation approaches such as the Kalman filtering for example. Note however that, the development of a Kalman filter is also confronted to a huge computational demand and the error estimates is necessarily and, often heavily, degraded at some point.

3 Description of the simulation experiments

The test simulation experiments have been performed in a square oceanic box assuming a schematic stratification. The main parameters of the model have been 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 m, 700 m and 4000 m. The basin has horizontal dimensions of 4000 km $\times$ 4000 km. Two main flow cases have been actually 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 in the Coriolis force avoids the formation of western boundary currents and associated resolution issues
  - the reduced gravity is $g'_{12} = 0.0357$ 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);
  - the wind stress curl is made of an alternated sinusoidal, 4 by 4 checkerboard pattern with a maximum amplitude of 4×10^{-2} m^2/s^2 and a zero average;
  - the lateral friction A_4 has an amplitude of 10^{-5} m^4/s and is parameterized as a biharmonic;
  - the bottom friction coefficient is $\epsilon = 5 \times 10^{-8}$ 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:
  - the reduced gravity is $g'_{12} = 0.0357$ m /s^2 and $g'_{23} = 0.0162$ m /s^2;
  - the wind stress curl has a double-gyre sinusoidal structure with an amplitude of 10^{-4} m^2/s^2;
  - the lateral friction A_4 has an amplitudes of 10^9 m^4/s and is parameterized as a biharmonic;
  - the bottom friction coefficient is $\epsilon = 10^{-7}$ s^{-1};
  - 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 twenty years of oceanic time. All further assimilation experiments are performed over time sequences far beyond this transitory spin-up phase.

Figure 1a and b shows an example of the surface flow fields for some arbitrary realization of Case 1 and Case 2 respectively, at some instant of the statistically steady-state. 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 bounary 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 similarily to the Gulfstream (or other western boundary currents) generation process. In the first case, the eddies are irrealistically large in size (about 1000 km) with regard to ocean mesoscales eddies. Typical correlation size for those eddies was of 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 Gulfstream system for example. The eddies are typically of 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 at a first stage the only real manner to validate an assimilation technique as with real observations there is no way to fully assess the performance of an assimilation experiment. In such 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. These pseudo-data are possibly degraded by adding some noise or error fields. 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 were supposed to be obtained on every gridpoint of the model with a time sampling of 1.5 day. Simulated surface data are then provided as observations for the cost function $J(u)$. The initial guess of the assimilation experiment is chosen for the three layers of the model to be completely decorrelated from the "true" solution. (In other term, 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, is 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 5,000 to about 120,000.

4 A standard assimilation strategy

For the first part of this study, the so called Case 1 which is much less bulky has been considered in order to save computing time.

In a first stage, it is very much illustrative to look at the results of assimilation experiments in which the length of the assimilation sequence is chosen arbitrarily. Figure 2 relates for example to an assimilation experiment which lasts for .5 month. It actually shows the Rms errors between the assimilation experiment streamfunction field and the reference streamfunction field in each layer. These errors are non-dimensionalized by the running Rms error of the reference flow:

$$\mathrm{Rms}_k = \sqrt{\frac{\int_{\mathcal{L}_k} (\psi_i^{Ass} - \psi_i^{Ref})^2 dx}{\int_{\mathcal{L}_k} (\psi_i^{Ref})^2 dx}} \qquad k = 1, 2, 3$$

The clear conclusion of Figure 2 (see also Table 1) is that although the surface layer is reasonably well identified, there is no good control in the bottom layer which stays quite far from the reference simulations that they are supposed to reproduce. The identification error of the initial state of layer 1 is not better than 8% (9% at the final state) when the convergence of the minimization is set by a stopping criterion of 10^{-3} for the convergence error of the gradient. In the first layer the error stays roughly at this level of 8 or 9% with a slight decrease in the mid-time interval. Conversely, the error decrease, although very slow, is slightly better in the bottom layers as the time is going.

Figure 3 shows the same results when the assimilation experiment is extended up to a much longer interval of 8 months. It is observed now that a better identification of the bottom layers is possible since the Rms errors can decrease down to some 12 or 13% at the final state (Table 1). But the surface layer where the data are taken is now more poorly identified. All layers have now a similar behavior with a strong variation in time of the Rms error convergence. Overall, the assimilation efficiency is relatively acceptable with regard to the final flow state while the initial state which is the control variable, is actually not controlled (the Rms errors are more than 100% for all layers!). One could see a possible paradox in this result: a poor identification of the initial state may anyway lead to a correct control of the final state. In any case, there is clearly no good optimization of the overall model trajectory.

A complementary result relative to the same experiment than Figure 3 is shown in Figures 4 a and b where the streamfunction correlation functions between the reference flow field and, the initial flow state and the final flow state respectively, are drawn as a function of the numbers of iterations for the minimization process. One can see that the correlation with regard to the initial state slowly increases with the number of iteration in the surface layer but not

significantly for the deep layers (Figure 4a). Quite differently, the correlation with respect to the final state quickly converges towards a strong correlation of more than 90% together for the three layers (Figure 4b). A limited number of iterations may be sufficient for reaching an acceptable convergence for the final state while a lengthy minimization process is required to reach an acceptable convergence of the actual control variable, i.e. the initial state. Obviously, the weight of recent data is dominating the adjustment process. In addition, only the first iterations of the minimization are able to provide for a quick control of the final flow state.

The experiments shown in Figures 2 and 3 point out the crucial role of the duration of the assimilation cycle. If this period is short, the assimilation is uneffective for the control of the deep flow with regard to the control of the surface flow. If the assimilation period is long, the assimilation may still be inoperative as the initial conditions cannot be precisely retrieved. The deep flow may be at least partly controlled in this case but the identification process of the initial flow state by itself performs poorly. Interpretation of this behaviour for long assimilation periods can clearly be related to the time-scale of the flow non-linearity. This decorrelation or predictability time-scale is a function of the spatial scales. In practice, it can be mostly associated to the more active mesoscale eddies: typical scales for these eddies in Case 1 are 1000 km for the length scale and 10 cm/s for the velocity leading to a typical turn-over time of 100 days. After a time which is of the order of this decorrelation scale, the flow has "forgotten" its past. Therefore, there is a lack of consistency in the data when one considers a time sequence which is larger than the decorrelation scale. In these situations, the global property of the variational approach to optimize the trajectory over a given time sequence is likely to fail.

From the numerical point of view, this corresponds to a failure of the minimization process to converge towards the true minimum. As an illustration, Figure 5 shows a section of the cost-function as a function of the distance between the control state and the reference state. It clearly shows the possibility of convergence towards a local minimum of the cost function. Additional understanding can be gained looking at the Hessian matrix. A spectral decomposition of the Hessian of the cost function (without the regularization term) has been computed at the optimum state by the Lanczos method [*Parlett and Scott*, 1979]. This latter method requires the evaluation of the product Hessian $\times$ vector which is computed by the second order adjoint equations [*Wang et al.*, 1990]. This decomposition reveals interesting information. The spectrum of the Hessian (not shown) clearly splits in two parts. The first part corresponds to large eigenvalues which amplify the corresponding modes, the second to small eigenvalues which damp the corresponding modes. The dimension of the "amplification" space is much smaller than the one of the "damping" space. As a consequence, the state variables are very sensitive to a small number of eigenmodes of the control space and they are almost completely decorrelated with the other eigenmodes. Moreover, these eigenvalues vary exponentially with the

length of the assimilation interval. Therefore, having long assimilation periods strongly conditions the minimization algorithm.

There are several conclusions at this stage. Firstly, the efficiency of the variational assimilation of altimeter data into such a non-linear ocean model is not obtained by a straight application of the method although the situation is much propitious as only simulated data are used and that these data are assumed to be perfect and sampled much more favorably than real data will be. Secondly, the effective convergence of the assimilation process would seem to be 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, (ii) but, it must be long enough for the deep flow control to be effective. An efficient procedure is possibly to sample the assimilation in an adequate number of sub-sequences: The variational assimilation would be achieved on each of these sub-sequences, the duration of which will satisfy the first criterion. After several such stages, the total time sequence should be long enough for the second criterion to be satisfied. Thirdly, the issue of the final objective for conducting the assimilation must be raised. Traditionnally, one speaks of "smoothing" and "filtering" in data assimilation problems. "Smoothing" aims at reconstructing the best possible trajectory for the system over all the observational time-interval under consideration. Quite differently, "filtering" wants to make an optimal use of past observations to estimate as precisely as possible the only state of the system at present time. Forecasting may be issued from such a filtering analysis. An adequate convergence on the final state vector in the terminology of this paper will clearly meet the filtering objective.

In some way, the requirement of relatively short assimilation intervals may seem penalizing and contradictory to the basic idea of variational assimilation that is to optimize the model trajectory over a given period, possibly the observational period. This is also rather contradictory to smoothing objectives. Also importantly, there might be some inadequacy between these time-intervals and the data collection. In the ocean, over a typical assimilation period thus imposed by the previously defined constraints, data may be very sparse. In general, this can prevent us from taking the full advantage of a long time series of data. These results are of potential importance for satellite altimetric measurements in which the periodicity of the observational sampling is usually not much smaller than the predictability scales of the extra-tropical ocean. For example, the periods of Geosat, Topex/Poseidon and ERS1 are respectively 17, 10 and 35 days whereas the predictability scale for the mid-latitude oceanic mesoscales is typically of the order of a month or less [*Brasseur et al.*, 1995].

As a reference, we have shown in Figure 6 (see also Table 1) the corresponding results in exactly the same conditions than Figure 3 except that a simple nudging technique is used for assimilation [*Verron*, 1990]. We just recalled that the nudging consists of relaxing the model predictions towards the observations

with a simple Newtonian relaxation term. It concerns only the equation in the upper layer which is written:

$$\frac{D}{Dt}\left[\Delta\psi_1 + f + \frac{f_0}{H_1}\left(h_{1+1/2} - h_{1/2}\right)\right] = F_1 + D_1 - R(\psi_1 - \psi_{Obs})$$

The tuning of the relaxation term is determined empirically. In the experiment shown, the relaxation factor is chosen as $R = 6 \times 10^{-7} m^{-2}.s^{-1}$. This value typically corresponds to a relaxation time-scale of 5 days for 100 km size eddies. The convergence of the nudging assimilation is rather slow but after a few months the Rms errors has fallen down to about 10% in the three layers. The main advantage of the nudging method is its small computing cost as it only adds a few percent to the regular model integration time. These results give a measure of the level of performances that can be obtained with a suboptimal method like nudging. At this stage of our investigations, standard variational assimilation without careful considerations of a strategy for the time sequencing does not perform any better than the nudging (Table 1).

5 A sliced assimilation strategy

As we previously said, a quite natural alternative worthwhile investigating to solve the previous difficulties is to subdivide the assimilation time period into several sequences, each of these sequences being smaller than the decorrelation scale and larger than the penetration scale. With regard to this general approach of splitting the assimilation period into several sub-intervals, several strategies are conceivable. The simplest trivial approach is to split the period of assimilation into several equal sub-periods and to make sequentially the assimilation over each subperiod. The initial guess of one sub-interval is the final state of the previous sub-interval. The "prior knowledge" that is contained in the successive first guesses is thus improved with time. This is the strategy investigated in this chapter and referred as to the sliced assimilation.

Several examples of this strategy are shown in Figures 7 for 4 series of variable length sub-sequences and on a global time period of 8 months. (Figure 3 can be reminded as a reference for the whole period. Note also that initial state convergence of Figure 7a is slightly different from Figure 2 as the convergence criteria are different.) The convergence for the error on the final state vector degrades slowly as the basic time-interval increases, although it stays relatively acceptable for a time-interval of up to 2 months (see also Table 2). The global convergence is better for the shortest basic time-interval case, i.e. the scenario $16 \times .5$ month. Although the deep flows do not converge in the first stages, after a typical duration of 3 months the convergence is reached as for the upper layers. The time is long enough for the penetration of information downward in the deep layers to be effective. In the shortest time-interval cases, the general convergence of the assimilation is very much improved compared to a standard

assimilation, especially and as expected, with regard to the final state and to filtering objectives. But the time slicing also allows for the minimizing process to converge in acceptable conditions in each sub-sequence. The control of the initial state in each sub-interval is improved after several sub-intervals, as the corresponding first guesses are of better quality. However, there is no way, with such a sliced assimilation strategy to improve the identification of the first initial state, i.e. the truly initial first guess. As a consequence, the smoothing problem is still not adequately solved.

Due to the progressive non-linear decorrelation, all the observations have not the same weight in the minimization process. The most recent observations are "adjusted" first and then the other observations when the time is going backward. This has the practical consequence that even a limited number of iterations may lead to a good adjustment of the final state (also remind Figures 4). This suggests a transitory strategy for assimilation experiment in which a first guess of good quality is not available. In that case indeed, it is suggested to require a not too stringent convergence criteria for the minimization. Similarily, one should not be too strict in term of convergence if the assimilation sequence is short, as the surface information has not enough time to propagate downwards in the deep layers. In some ways, the surface layer is being decoupled from the bottom during the assimilation process: it is "pulled out" towards the observations but there is no control for the vertical dimension to be dynamically adjusted. An accurate convergence of the minimization may then be unnecessary for the control of the final flow state vector.

The experiments shown in Figures 8 are very revealing in this regard (see also Table 3). It concerns a sliced assimilation experiment divided in 16 slices of .5 month each. In a first experiment of this type (Figure 8a) the minimization process is conducted over 1000 iterations i.e. a rather stringent criterion for the convergence. In the second experiment (Figure 8b), a rough convergence is tolerated as the criterion only specifies 15 iterations for the minimization. The regularization term must be adjusted in time. As already said, the systematic use of an optimal adjustment following the GCV method was found prohibitive. Therefore, a simpler approach was adopted in practice here: the regularization coefficient was divided by two every month starting from an initial value of $\epsilon = 10^{-6}m^2$. As expected, a better identification of the initial state is obtained for a more stringent convergence criterion. But surprisingly, a similar convergence is obtained in all layers for the final state even when the convergence criterion is less strict. The more constrained minimization has allowed a better identification of the initial first guess in the case of Figure 8a but this has not been really favorable for the whole period assimilation. The explanation of this probably lies in the fact that, during the first iterations, the minimization algorithm works essentially on data which are at the end of the assimilation interval. A "perfect" minimization is not more favorable with regard to the deep flow identification as on a short time-interval , we have seen that surface information alone cannot allow a good recovery of the corresponding deep flows.

It is then better to keep track of the longer time progressive adjustment of the deep flows which is contained in the successive "first guesses" by performing a "rough" minimization. Again, one should raise here the issue of the final objective to be reached in this assimilation process. But clearly, the objectives of filtering and forecasting types can be attained with only a loose convergence of the minimization process that is also a much less expensive computational burden.

6 A progressive assimilation strategy

The disadvantage of the sliced assimilation is that the trajectory of the identified flow undergoes discontinuities and that the identification on the control vector does not fully exploit all the available information over the whole observational time sequence. Therefore, a second approach has been considered for dividing the global time period of the assimilation sequence, which consists of mixing an iterative process to the slicing. We call this the progressive assimilation strategy. The assimilation is first made over some short sub-period; the identified initial state is then chosen as the first guess for the identification over a longer period. This procedure is then continued until all the data available are taken into account. This procedure is more adapted to an improved identification of the initial state which is the actual control variable and to the smoothing problem.

The results of this procedure are represented on Figure 9 a, b and c and in Table 4 for the three layers of the model in a case of a progressive assimilation for successively .5, 1, 2, 4 and 8 months of iterative assimilation. This iterative method enables us to propagate the surface information to the bottom, the results being better and better as the length of assimilation increases. Figures 9 show clearly the iterative process by decomposing for each layer the convergence in time-sequences. 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. Although this progressive assimilation procedure is more time-consuming than the sliced one, the Rms on the initial state is much better. Figure 10 synthesizes the overall convergence of the full cycle. The control is excellent for the final state but also very good for the initial state contrary to all the other strategies investigated so far. With this strategy, one can consider that the identification and smoothing problems have been adequately solved and the model trajectory optimized all over the assimilation period.

We have seen from the investigations on the standard assimilation that the shorter the assimilation interval is, the larger the attraction basin of the functional will be. The progressive assimilation strategy is effective since the iteration process forces the successive initial states to stay close to the attraction basin of the functional even though the length of the assimilation interval is

increased. The convergence towards the global optimum of the functional may then be favored. However, the time sequencing of a given progressive assimilation strategy cannot necessarily satisfy such maintenance of the successive initial states within the true attraction basin. This evidently depends on the successive size of the time-intervals. In practice, we have chosen each time-interval as the double of the previous one and convergence was always reached.

7 A large size problem application

The experience gained from the previous investigations with the relatively small (computational) size problem of Case 1 has been used for considering the Case 2 with a high resolution. This situation is in many respects representative of a basin-scale situation such as the North Atlantic ocean. The computational burden is now similar to the realistic size problem and more especially the eddy field is now relatively realistic in terms of spatial and temporal scales and more generally of statistical properties. The computing cost is now a limiting factor and only a series of successive experiments employing the sliced assimilation strategy has been run. A series of 36 sliced assimilation sequences of ten days was actually performed. The overall assimilation has been then conducted for a period of one year (actually 360 days). The regularization factor has been slowly decreased with time as indicated in Table 5. The minimization criterion is always 10^{-3} but it is reached after a variable number of iterations, roughly from a typical 500 in the first sequences to about 100 in the last ones.

Table 5 presents the results for the final flow Rms errors at different stages in time. The error decrease is rapid in the first three months but becomes rather slow after that. The error in the upper layer goes down to 1% after 2 months and stays roughly at this level. The convergence in the intermediate layer is always slower than in the deep layer but this was also observed in the Case 1. It can be interpreted by the fact that the control of the second baroclinic mode is slower than the control of other modes.

Figures 11 shows the adjustment in the three layers during the last month of the experiment which is thus divided in three sub-sequences of 10 days. A slight improvement of the Rms error is still perceptible as the time goes but only in the deep layer.

The one-year experiment has been compared to the performances of nudging in equivalent conditions. The results are also shown in Table 5 for the final state errors in the three layers. The variational approach is now indisputably much more efficient than the nudging approach.

8 Conclusion

This paper shows that the problem of controlling an oceanic flow field under the only constraint of surface information can be solved by the adjoint method for the QG model. However, this solution does not follow directly from a single application of the method and imposes some constraints on the practical achievement of the assimilation. The main constraint lies in the duration of the assimilation sequence. Indeed there is a need for a sufficient time interval for the information to propagate downwards to the deep ocean circulation, but in the opposite way a too long time sequence penalizes the assimilation process because of the decorrelation scale associated with non-linearity. An optimal approach seems therefore to split the assimilation process into several time sub-intervals.

Various temporal strategies can then be considered to perform this assimilation. Actually, the choice of a strategy must depend strongly on the objectives to be reached. A "sliced assimilation strategy" in which the whole assimilation cycle is divided into several equal sub-intervals may be sufficient for filtering and forecasting objectives. In that cases indeed, one especially aims at defining some "initialization" field for prediction (which would be the "final" state in the terminology of this paper) but not a global optimization along the overall observational time sequence. A sliced assimilation may indeed lead to an adequate convergence of the final flow state after several time sub-intervals.

With regard to filtering objectives and forecasting, an alternative approach may also be thought of. Given the numerical cost of the adjoint method and the difficulty for this method to cope with the problem of local minimums, it may indeed be worthwhile to adopt in realistic problems a two-stage strategy in which a sub-optimal method like nudging is used to rough out the flow solution and provide for a prior guess not exceedingly far from the truth. Then, the adjoint method would be applied to refine the convergence of the very final solution.

The objective of global optimization, typical of smoothing problems, may be better reached using the so called "progressive assimilation strategy" that is an iterative assimilation process which is performed with time-intervals of progressively increased durations. The pure control problem aiming at the proper identification of the control variable, namely the initial state vector, is adequately solved with such a progressive assimilation strategy. Note that the initial state vector stays as the control variable for all sequences, contrary to the sliced assimilation strategy where successive initial state vectors are used as successive control variables. Actually, the progressive assimilation strategy gives an accurate result for all purposes: the identification objective is by far the most constraining objective and is satisfied once the other objectives of filtering and smoothing are already satisfied. It is also the most computationally expensive approach which requires typically two times more calculations than

the sliced assimilation. It can cost even more since a sliced assimilation strategy often requires only a rough convergence of the minimization algorithm which is not the case for a proper identification in a progressive assimilation strategy.

Acknowledgements. We are indebted to J. C. Gilbert and C. Lemaréchal (INRIA, Rocquencourt) for having provided to us their efficient minimization algorithm. The support of ORSTOM at Nouméa, New Caledonia, where one of the authors had a visiting position during the completion of this work is also acknowledged. The calculations were made possible by the computer facilities at the Centre Grenoblois de Calcul Vectoriel (CGCV-CEA) and at the Institut de Recherche et de Ressources en Informatique Scientifique (IDRIS-CNRS).

		Layer 1	Layer 2	Layer 3
Standard assimilation (.5 month)				
	Initial state	8%	33%	46%
	Final state	9%	25%	38%
Standard assimilation (8 months)				
	Initial state	120%	132%	117%
	Final state	20%	13%	12%
Nudging assimilation (8 months)				
	Initial state	140%	120%	120%
	Final state	13%	7%	9%
Sliced assimilation (16×.5 month)				
	Initial state	18%	37%	41%
	Final state	.2%	1.5%	.5%
Progressive assimilation (.5+1+2+4+8 months)				
	Initial state	4%	7%	8%
	Final state	1%	.6%	.6%

Table 1: Case 1. Residual Rms errors of the initial state and of the final state for some selected assimilation experiments discussed in text.

		Layer 1	Layer 2	Layer 3
Sliced assimilation (16×.5 month)				
	Initial state	18%	37%	41%
	Final state	.2%	1.5%	.5%
Sliced assimilation (8×1 months)				
	Initial state	33%	48%	53%
	Final state	1%	2%	2%
Sliced assimilation (4×2 months)				
	Initial state	60%	67%	63%
	Final state	4%	3%	4%
Sliced assimilation (2×4 months)				
	Initial state	85%	94%	91%
	Final state	12%	8%	7%
Sliced assimilation (1×8 months)				
	Initial state	120%	132%	117%
	Final state	20%	13%	12%

Table 2: Case 1. Residual Rms errors of the initial state and of the final state for various scenarios studied of sliced assimilation experiments.

		Layer 1	Layer 2	Layer 3
Sliced assimilation 1000 iterations used for the minimization				
	Initial state	7%	33%	46%
	Final state	.5%	9%	2%
Sliced assimilation 15 iterations used for the minimization				
	Initial state	37%	70%	119%
	Final state	.5%	4%	2.5%

Table 3: Case 1. Effects of the minimization convergence criteria. Residual Rms errors of the initial state and of the final state for the sliced assimilation experiment 16×.5 month with two different convergence criteria.

		Layer 1	Layer 2	Layer 3
Progressive assimilation (After .5 month)				
	Initial state	6%	33%	45%
	Final state	7%	25%	37%
Progressive assimilation (After 1 months)				
	Initial state	4.5%	23%	24%
	Final state	6%	15%	15%
Progressive assimilation (After 2 months)				
	Initial state	4.5%	14%	10%
	Final state	4%	8%	3%
Progressive assimilation (After 4 months)				
	Initial state	4%	8%	9%
	Final state	2%	1.5%	1%
Progressive assimilation (After 8 months)				
	Initial state	4%	8%	8%
	Final state	1%	.6%	.6%

Table 4: Case 1. Residual Rms errors of the initial state and of the final state at the successive stages of the progressive assimilation experiment.

		Layer 1	Layer 2	Layer 3
Sliced assimilation				
	Regularization factor (m^2)			
after 1 month	1×10^{-7}	5%	30%	41%
after 2 months	5×10^{-8}	5%	23%	32%
after 3 months	2×10^{-8}	1%	19%	24%
after 6 months	2×10^{-8}	1%	13%	11%
after 9 months	5×10^{-9}	1%	12%	10%
after 12 months	1×10^{-9}	1%	11%	9%
Nudging				
after 12 months		26%	41%	51%

Table 5: Case 2. Residual Rms errors for some successive final states during a one year assimilation divided in 36 slices of ten-day sliced assimilation. Comparison with a nudging assimilation experiment using a nudging factor of $1 \times 10^{-7} m^{-2}.s^{-1}$.

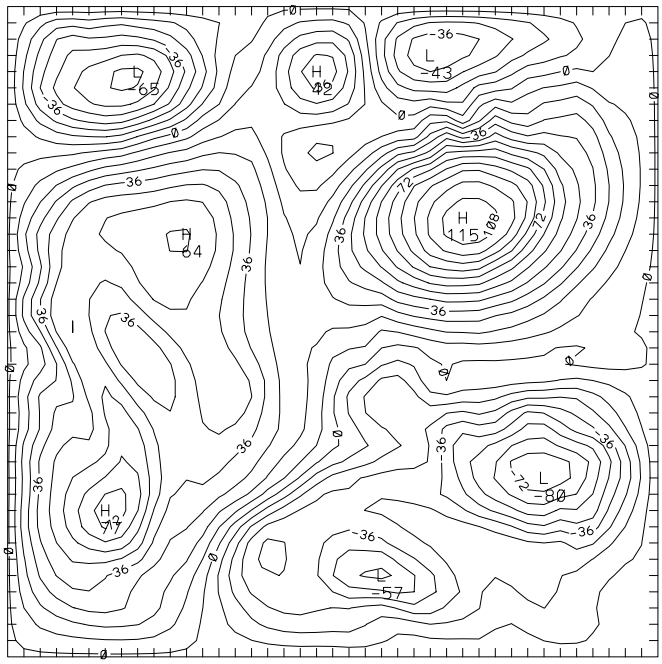

Figure 1a: Example of realization for the upper surface streamfunction in flow Case 1.

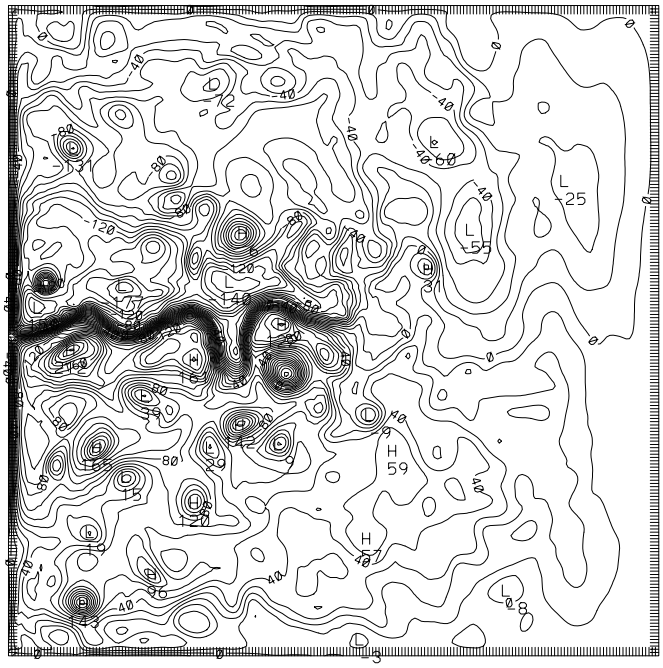

Figure 1b:Example of realization for the upper surface streamfunction in flow Case 2.

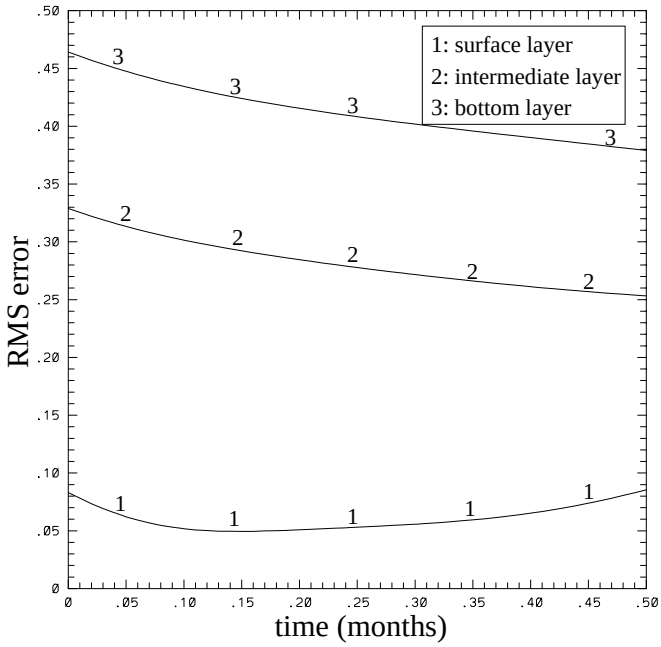

Figure 2: Rms errors for the standard assimilation over half a month.

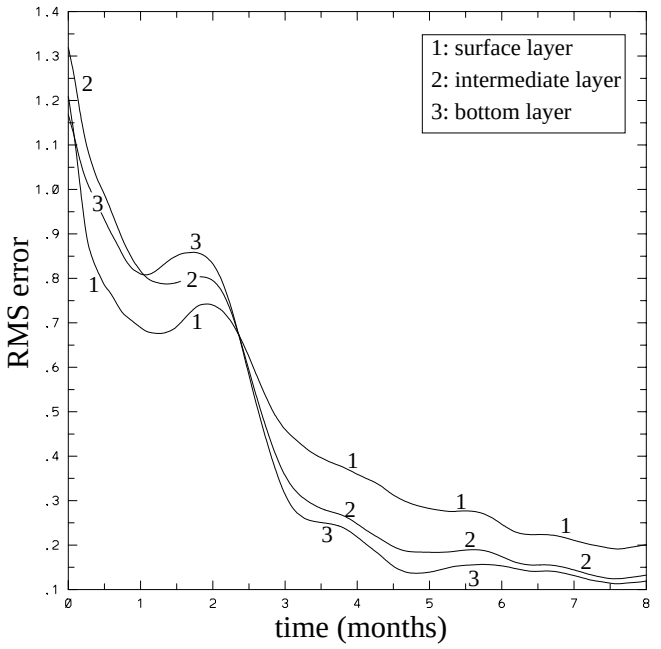

Figure 3: Rms errors for the standard assimilation over 8 months.

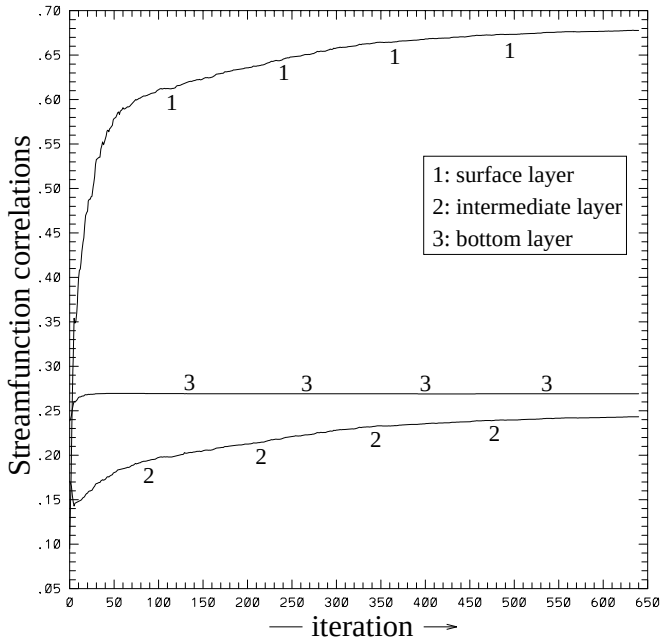

Figure 4a: Streamfunction correlations between the reference and the identified initial state.

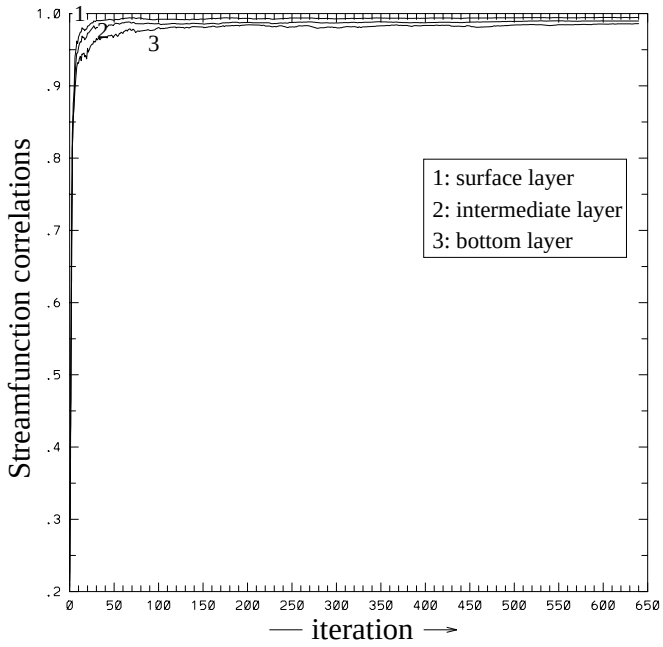

Figure 4b: Streamfunction correlations between the reference and the identified final state.

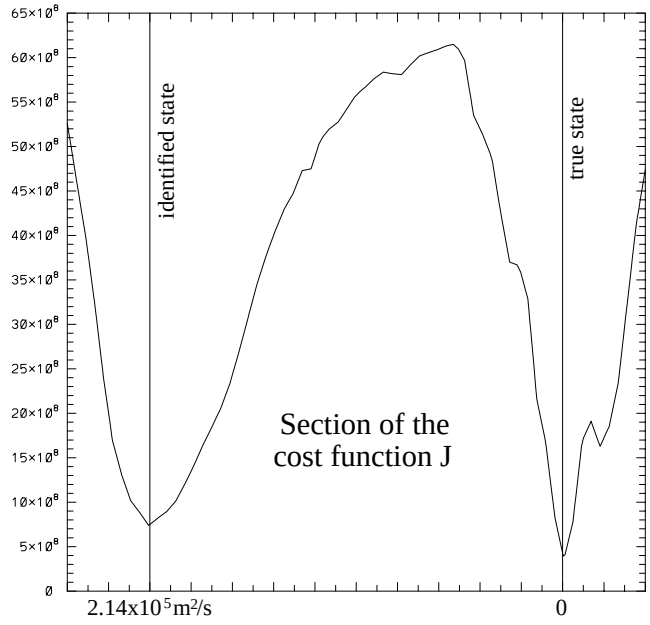

Figure 5: Section of the cost function.

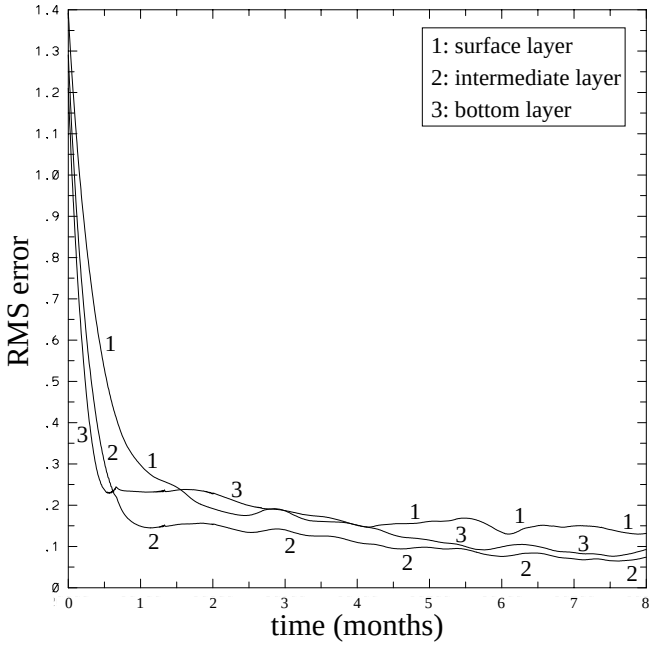

Figure 6: Rms errors for a nudging assimilation over 8 months.

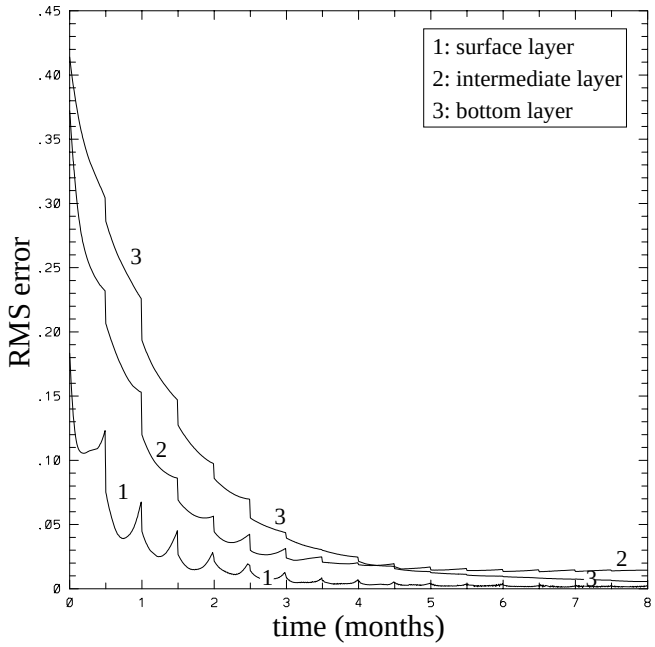

Figure 7a: Rms errors for a sliced assimilation
with 12 sequences of 0.5 month.

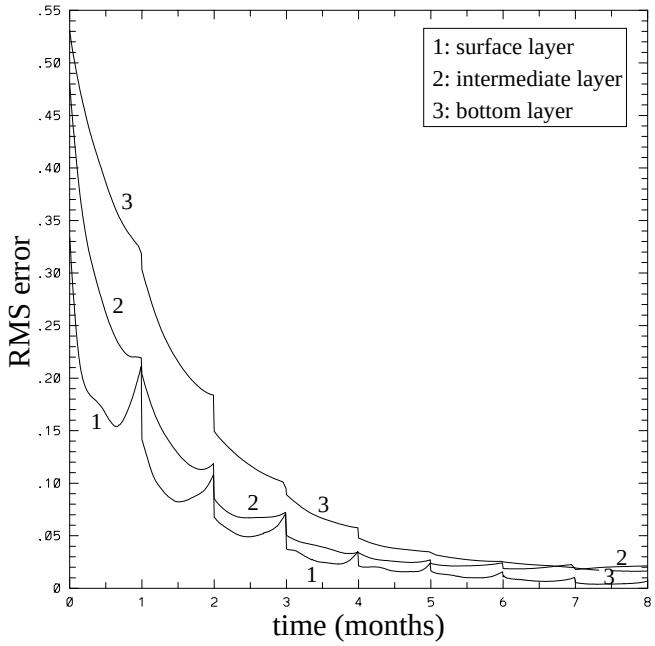

Figure 7b: Rms errors for a sliced assimilation
with 6 sequences of 1 month.

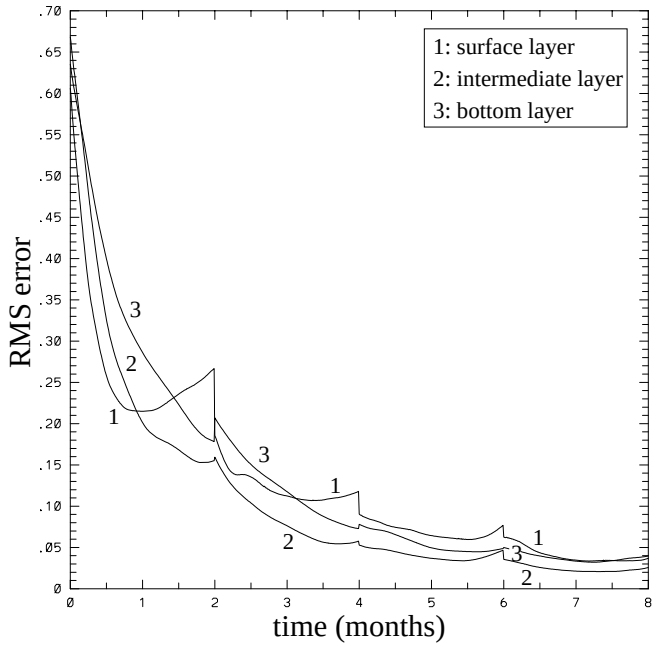

Figure 7c: Rms errors for a sliced assimilation with 3 sequences of 2 months.

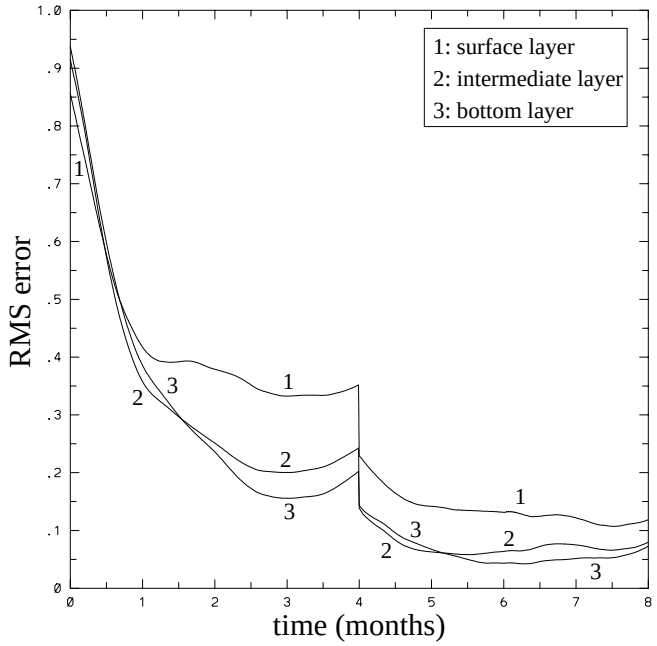

Figure 7d: Rms errors for a sliced assimilation with 2 sequences of 3 months.

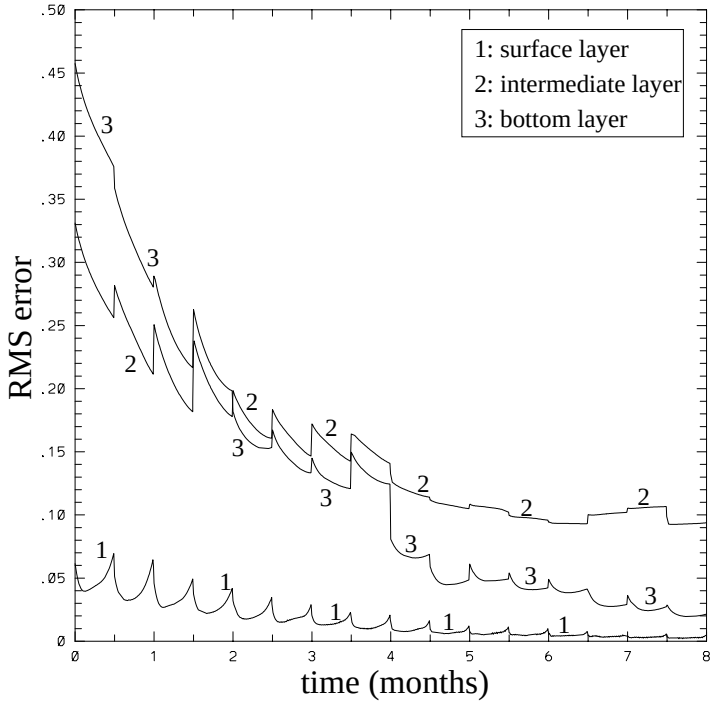

Figure 8a: Rms errors for a 0.5 month sliced assimilation with 1000 iterations.

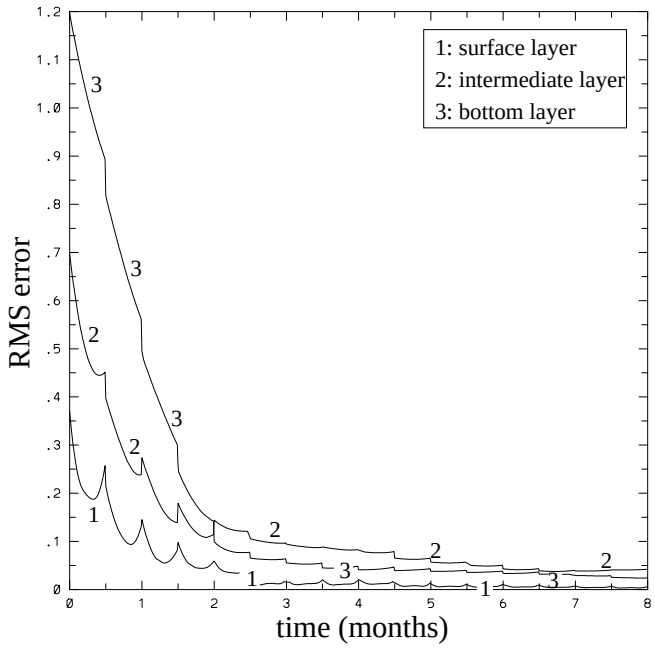

Figure 8b: Rms errors for a 0.5 month sliced assimilation with 15 iterations.

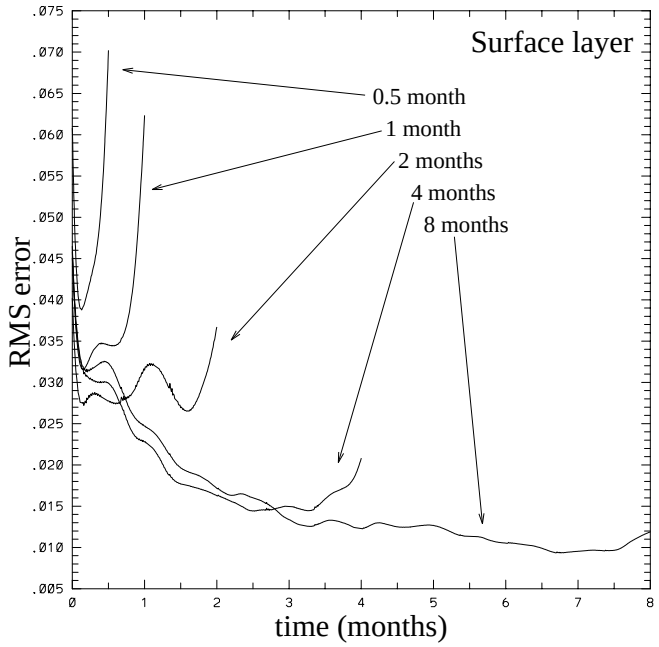

Figure 9a: Rms errors for a progressive assimilation with .5, 1, 2, 4 and 8 months.

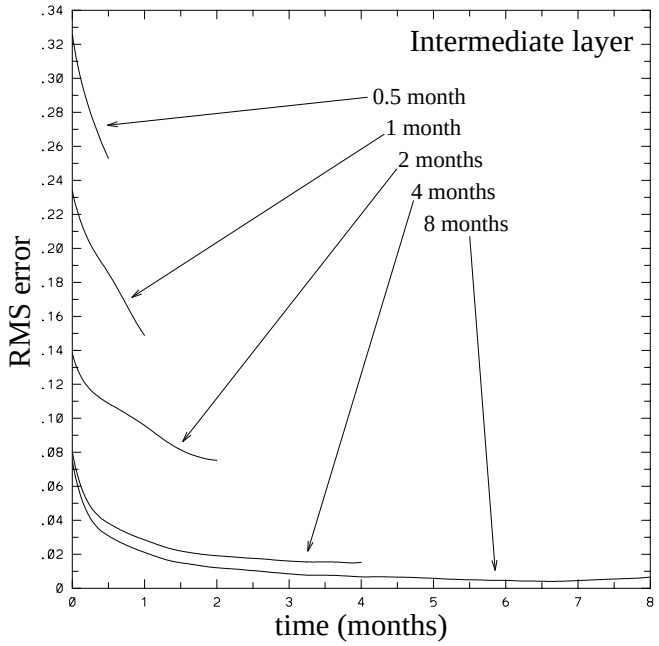

Figure 9b: Rms errors for a progressive assimilation with .5, 1, 2, 4 and 8 months.

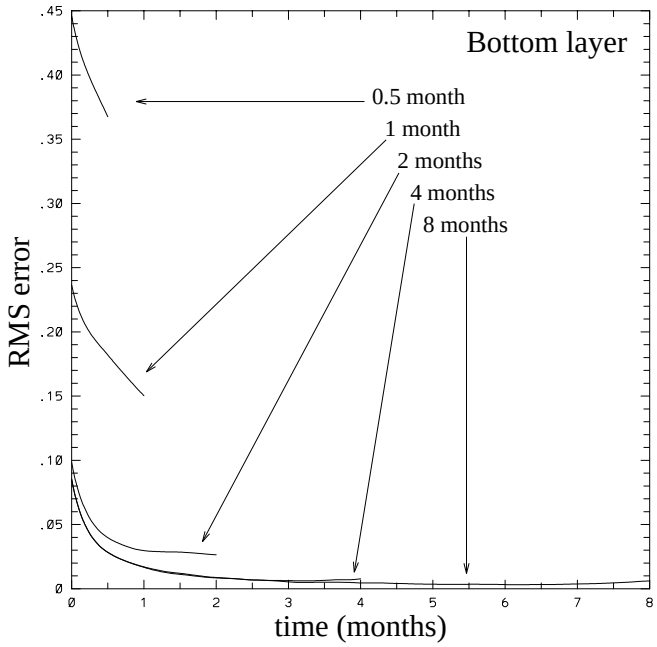

Figure 9c: Rms errors for a progressive assimilation with .5, 1, 2, 4 and 8 months.

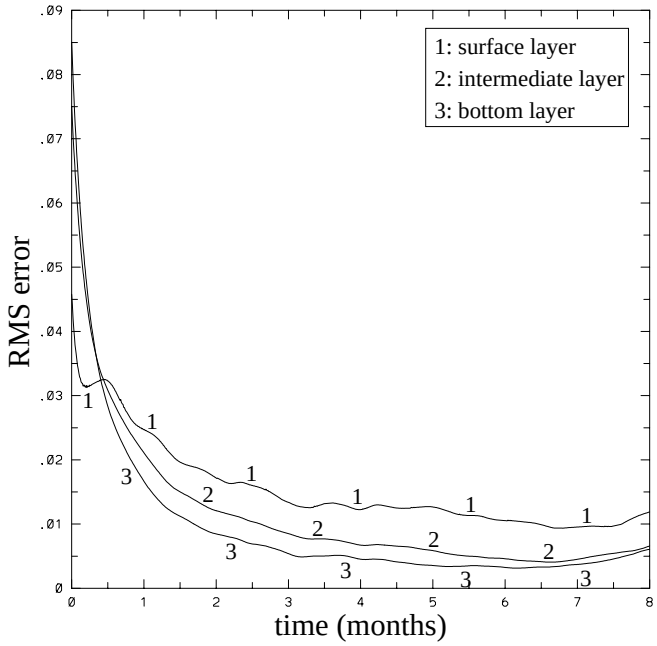

Figure 10: Same figure as before with only the final convergence and three layers together.

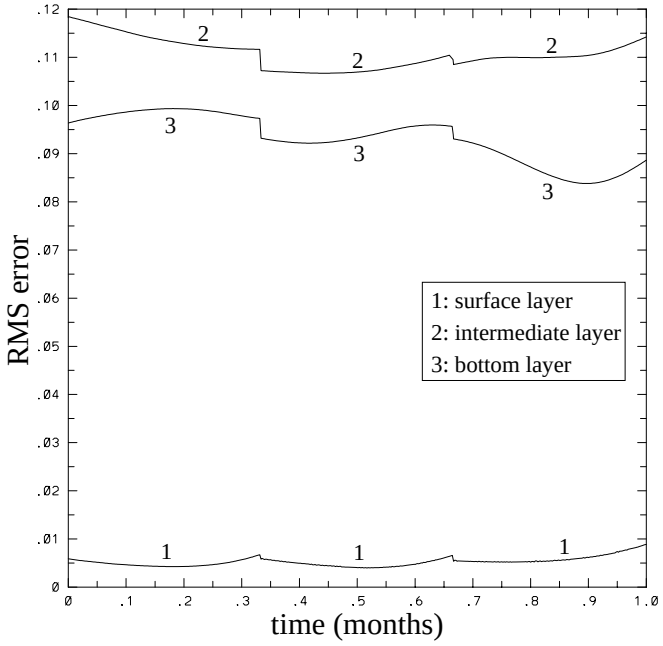

Figure 11: Rms errors during the last month for a sliced assimilation with 36×10 days.

References

Brasseur P., E. Blayo et J. Verron, 1995: Predictability experiments in the North Atlantic ocean: outcome of a QG model with assimilation of Topex/Poseidon altimeter data. Submitted to the *J. Geophys. Res.*.

Cooper M. et K. Haines, 1995: Altimetric assimilation with water property conservation. Submitted to the *J. Phys. Oceanogr.*.

De Mey P. and A. R. Robinson, 1987: Assimilation of altimeter eddy fields in a limited-area quasi-geostrophic model. *J. Phys. Oceanogr.*, 17, 2280-2293.

Evensen G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, 97, 17,905–17,924.

Ezer T. et G. L. Mellor, 1994: Continuous assimilation of Geosat altimeter data into a 3D primitive equation Gulf Stream model. , *J. Phys. Oceanogr.*, 24, 832-847.

Gauthier P., 1992: Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model. *Tellus*, 44A, 2-17.

Gauthier P., P. Courtier and P. Moll, 1993: Assimilation of simulted wind lidar data with a Kalman filter. *Mon. Wea. Rev.*, 121, 1803-1820.

Ghil, M., and P. Manalotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.*, 33, 141-266, 1991.

Gilbert J. C. and C. Lemaréchal, 1989: Some numerical experiments with variable storage quasi-Newton algorithms. *Mathematicals programming*, B45, 503-528.

Haines K., 1991: A direct method for assimilating sea surface height data into ocen models with adjustments to the deep circulation. *J. Phys. Oceanogr.*, 21, 843-868.

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, 363-392.

Hurlburt H. E., 1986: Dynamic transfer of simulated altimeter data into subsurface information by a numerical ocean model. *J. Geophys. Res.*, 91, 2372-2400.

Ikeda M., G. Evensen and L. Cong, 1995: Comparison of sequential updating, Kalman filter and variational methods for assimilating Rossby waves in the simulated Geosat altimeter data into a quasi-geostrophic model. *J. Mar. Syst.*, 6, 15-30.

Lavrentiev M. M., 1967: Some improperly posed problems of mathematical physics. Springer, New-York.

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

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

Lions J. L., 1968: Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles. Eds. Dunod.

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

Luong B., J. Blum and J. Verron, 1997: A variational method for the resolution of a data assimilation problem in oceanography, to appear in *Inverse Problems*.

Miller R. N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*

Mellor G. L. et T. Ezer, 1991: A Gulf Stream model and an altimetry assimilation scheme. JGR, 96, C5, 8779-8795.

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.

Morrow R. and P. De Mey, 1995: An adjoint assimilation of altimetric, surface drifter and hydrographic data in a QG model of the Azores current. Submitted to *J. Geophys. Res.*

Nechaev and M. I. Yaremchuk, 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), 865-879.

Oschlies A. et J. Willebrand, 1995: Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic ocean. *J. Phys. Oceanogr.*, submitted.

Parley B. N. and D. S. Scott, 1979: The Lanczos algorithm with selective orthogonalization. *Mathematics of Computation*, 33 (145), 217-238.

Pedlosky J., 1987: Geophysical fluid dynamics. Springer-Verlag. New York Heidelberg Berlin.

Penenko V. V. and N. N. Obraztsov, 1976: A variational initialization method for the fields of the meteorological elements. *Soviet Meteorol. Hydrol.*, 11, 1-11.

Schmitz, W. J. and W. R. Holland, A preliminary comparison of selected numerical eddy-resolving general circulation experiments with observations, *J. Mar. Res.*, 40, 75-117, 1982.

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

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

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

Tanguay M., P. Bartello and P. Gauthier, 1995: Four-dimensional data assimilation with a wide range of scales. *Tellus*, 47A, 974-997.

Thacker and Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, 93, 1227-1240.

Tikhonov A. and V. Arsenin, 1977: Solutions of ill-posed problems. Wiley, New-York.

Verron, J., 1990: Altimeter data assimilation into an ocean model: sensitivity to orbital parameters, *J. Geophys. Res.*, 95 (C7), 11443-11459.

Verron, J., 1992: Nudging altimeter data into quasigeostrophic ocean models, *J. Geophys. Res.*, 97 (C5), 7479-7491.

Wahba G., 1980: Cross validation and constrained regularization methods for mildly ill posed problems. Proceedings of the Internationnal Conference on Ill Posed Problems, M.Z. Nashed (ed.). Academic Press, NewYork.

Wang Z., Navon I. M. and F. X. Le Dimet, 1990: The second order adjoint analysis: Theory and Applications. *Meteorol. Atmos. Phys.*, 50

Figure captions

Figure 1. Examples of realization for the upper surface streamfunction in flow Case 1 (a) and flow Case 2 (b).

Figure 2. Case 1. Convergence of the Rms errors in the three model layers for a standard assimilation experiment over half a month. The regularization factor is $\epsilon = 10^{-7}m^2$. The minimization criterion is 10^{-3} and this requires approximately 500 iterations. Figure 3. Case 1. Convergence of the Rms errors in the three model layers for a standard assimilation experiment over 8 months. The regularization factor is $\epsilon = 10^{-7}m^2$. The minimization criterion is 10^{-3} and this requires approximately 640 iterations.

Figure 4. Case 1. Streamfunction correlations between the reference flow field and the initial flow state (a) or the final flow state (b) as a function of the number of iterations for the minimization process. The regularization factor is $\epsilon = 10^{-7}m^2$.

Figure 5. Case 1. Section of the cost function as a function of the distance between the control state and the reference state for the case experiment of Figure 3.

Figure 6. Case 1. Convergence of the Rms errors in the three model layers for an assimilation experiment with the nudging technique over eight months. The nudging factor is $6 \times 10^{-7}m^{-2}.s^{-1}$.

Figure 7. Case 1. Convergence of the Rms errors in the three model layers for a sliced assimilation experiment in which the time sequences are of 16 sequences of .5 month (a), 8 sequences of 1 month (b), 4 sequences of 2 months (c) and 2 sequences of 4 months (d). The regularization factor is $\epsilon = 10^{-7}m^2$. The minimization is stopped at 100 iterations.

Figure 8. Case 1. Convergence of the Rms errors in the three model layers for a sliced assimilation experiment in which the time slices are .5 month long. The number of iterations for the minimization is 1000 (a) and 15 (b). The regularization coefficient is ajusted for each time-interval: it starts from $\epsilon = 10^{-6}m^2$ and it is divided by two every month.

Figure 9. Case 1. Convergence of the Rms errors for a progressive assimilation experiment with a successive series of .5, 1, 2, 4 and 8 months assimilation sequences: Surface layer (a), intermediate layer (b) and bottom layer (c). The regularization factor is $\epsilon = 10^{-7}m^2$. The minimization criterion is 10^{-3}.

Figure 10. Case 1. Same figure as before at the final convergence for all three layers together. The regularization factor is $\epsilon = 10^{-7}m^2$. The minimization criterion is 10^{-3}.

Figure 11. Case 2. Convergence of the Rms errors during the last month of the sliced assimilation experiment 36×10 days. The regularization factor is $\epsilon = 5 \times 10^{-9}m^2$. The minimization criterion is 10^{-3}.