Journal of Oceanography, Vol. 62, pp. 539 to 550, 2006

Efficiency of Reduced-Order, Time-Dependent Adjoint Data Assimilation Approaches

IBRAHIM HOTEIT[1]* and ARMIN KÖHL[2]

[1]*Scripps Institution of Oceanography, La Jolla, CA 92093, U.S.A.*
[2]*Institut für Meereskunde, Hamburg, D-20146, Germany*

(Received 23 August 2005; in revised form 12 April 2006; accepted 12 April 2006)

Applications of adjoint data assimilation, which is designed to bring an ocean circulation model into consistency with ocean observations, are computationally demanding. To improve the convergence rate of an optimization, reduced-order optimization methods that reduce the size of the control vector by projecting it onto a limited number of basis functions were suggested. In this paper, we show that such order reduction can indeed speed up the initial convergence rate of an assimilation effort in the eastern subtropical North Atlantic using *in situ* and satellite data as constraints. However, an improved performance of the optimization was only obtained with a hybrid approach where the optimization is started in a reduced subspace but is continued subsequently using the full control space. In such an experiment about 50% of the computational cost can be saved as compared to the optimization in the full control space. Although several order-reduction approaches seem feasible, the best result was obtained by projecting the control vector onto Empirical Orthogonal Functions (EOFs) computed from a set of adjusted control vectors estimated previously from an optimization using the same model configuration.

Keywords:
· Data assimilation,
· 4DVAR,
· adjoint method,
· order reduction,
· EOFs.

1. Introduction

Data assimilation aims at combining numerical models and observations to obtain an optimal description of the time-evolving state of a dynamical system. This technique is widely used in meteorology and has recently evolved into an important data synthesis element of physical oceanographic research. The theoretical framework of data assimilation for oceanographic problems is well established (Bennett, 1992; Wunsch, 1996) and two separate directions are usually followed in oceanographic applications, one being a filtering approach (e.g., Kalman filter), the other being a variational approach (e.g., Lagrange multiplier or adjoint method). Filtering methods proceed by incrementally correcting a model prediction each time new observations are available, based on prior information about uncertainties in the model and data. The variational approach is a specific case of the more general framework of the optimal control theory and belongs to the so-called "whole domain" approaches. Essentially, it consists of finding the model trajectory which best fits available data over a given period of time, taking model and data errors into account. This is accomplished by minimizing a cost function which measures the discrepancy between the model simulation and observations while adjusting a well-chosen set of uncertain model parameters, called control parameters. The variational problem is then solved as a constrained optimization problem while adding the model dynamics as constraints to the cost function. Ghil and Malanotte-Rizzoli (1991) and Wunsch (1996) describe applications in physical oceanography.

Although mathematically rigorous assimilation methods have been found to be very efficient in bringing a model closer to consistency with large multivariate data sets, their computational burden remains one of the main obstacles to using them in a full eddy-resolving setting or even to transition existing state-of-the-art technology to operational centers. For instance, filtering methods require the manipulation of huge-dimension covariance matrices of the estimation error. Attempts to reduce the computational burden of the Kalman filter have received considerable attention in the past (see, e.g., Fukumori and Malanotte-Rizzoli, 1995; Cohn and Tolding, 1996; Pham *et al*., 1997; Hoteit *et al*., 2002). Most proposed "sub-optimal" Kalman filters project the system state onto a low-dimensional subspace using an order-reduction op-

* Corresponding author. E-mail: ihoteit@ucsd.edu

erator which lead to low-rank representations of the error covariance matrices.

Because of the complexity of the variational optimization problem, the variational equations are solved implicitly and iteratively in practice, employing the gradients of the cost function and an optimization code (Wunsch, 1996). The most efficient way to compute the gradients of the cost function with respect to the control variables is to use the adjoint of the numerical model, which provides this information for the cost of a few model integrations. These gradients are then used in a descent algorithm to reduce the model-data misfit. The computational burden of this procedure is significant as usually many (ten to hundreds) iterations are needed to reach a satisfactory solution, each requiring one forward and one backward integration of the numerical model and its adjoint. Early attempts to reduce the computational burden of the variational methods have therefore focused on the simplification of the adjoint model; using a coarser grid in the adjoint model as proposed by Courtier *et al.* (1994), or through physical considerations and truncation of the adjoint model (e.g., Vogeler and Schröter, 1995).

Since the optimization is solved iteratively, the computational cost of a 4DVAR problem is also largely determined by the number of iterations required to bring the model more into consistency with the observations. Accelerating the convergence process is therefore another key issue for increasing the applicability of adjoint models for oceanographic problems. This problem is practically not present for low-dimensional linear problems, since it is usually possible to obtain a sufficient approximation for the preconditioning Hessian to improve the ratio between the dominant eigenvalue spectrum of the Hessian which characterizes the convergence speed of the optimization (Courtier, 1997). However, in nonlinear 4DVAR atmospheric and oceanic data assimilation applications with large dimensional control space, good approximations to the Hessian are harder to obtain and one way to improve the convergence rate of an optimization is to carry it in a smaller dimensional space. This is the idea behind the linearly-equivalent dual formulation (also called the Physical-space Statistical Analysis System (PSAS)*) of the variational adjoint approach (Courtier, 1997), which consists of performing searches for the optimum solution in the smaller data space rather than the large control space. Although the 4DVAR and PSAS methods were found to behave similarly when the simplest preconditioning strategies were employed (Golub and Van loan, 1996), the latter can lead to a superior convergence rate when the number of observations is much smaller than the number of control variables if sophisticated preconditioning methods are employed (Daley and Barker, 2000). However, given the large—and ever increasing—number of observations now available, the benefits of the dual approach may be rather limited.

Another way to reduce the dimension of the line search is to project the control vector onto a subspace of much smaller dimension using an order-reduction operator. Off course, this entails an approximation which in general will reduce the performance of the optimization because of the limited number of allowed search directions. This loss of accuracy can be insignificant, however, if a sufficiently representative reduced control space is available, as argued by Vidard *et al.* (2000) who adjusted the model error in the subspace spanned by the directions of the fastest growing perturbations. The authors also found that an order-reduction can be beneficial in preventing the model from fitting model noise and observation errors. Such an approach would also enable the use of the full Hessian matrix* in the optimization algorithm, and therefore naturally allow for improvements in performance. Recently, Durbiano (2001) and Robert *et al.* (2005) demonstrated the efficiency of this method in speeding up the optimization procedure while controlling the model initial conditions in a subspace generated by a set of Empirical Orthogonal Functions (EOFs). Although they found this application to be very effective in their twin experiments, the resulting EOFs may not be as effective when using real data since the variability of the corrections imposed by the real observations is often significantly different from that simulated by the model.

In this paper, an EOF approach has been adopted to test reduced-order optimization schemes in a realistic setting which usually proves to be much more complex and less forgiving than in a "twin-experiment" setting. Our goal is to build an optimization procedure that would speed up the convergence of ongoing global or regional synthesis approaches. As an example, the "Estimating the Circulation and Climate of the Ocean" (ECCO) consortium (see Stammer *et al.* (2002) for details) performs global assimilation efforts using the MIT model with a 1° spatial resolution and most available data during the period 1992 through 2002. In those efforts control variables are initial temperature and salinity fields and daily fields of surface momentum, heat and freshwater fluxes. Because of the large control vector and the associated large number of iterations required during an adjoint optimization, its computational cost is a factor of 100 to 1000 higher than regular forward simulations. Here, we essen-

*The PSAS algorithm is equivalent to the indirect representer method (Bennett, 1992).

*A diagonal approximation of the inverse Hessian matrix is often used as a preconditioner in Lagrange-multiplier approaches because of the large dimension of the system (Yang *et al.*, 1996).

tially use the same model as in the global approach, but investigate the effect of an order-reduction scheme only in a small sub-region of the sub-tropical Atlantic Ocean. In our present study we use only the atmospheric forcing fields as control variables, which are adjusted during the optimization to improve the consistency between the model and real (anomalies and mean) TOPEX/ERS sea surface height data, Reynolds sea surface temperature data, and Levitus temperature and salinity data. The more general case of including the model initial conditions in the control vector is straightforward and does not require any special treatment.

Hereafter, we first describe and discuss the application of an order-reduction approach in the context of 4DVAR methods before presenting and analyzing assimilation results of several numerical experiments.

2. The Reduced-Order 4DVAR Approach

To present the use of order-reduction in the context of the adjoint 4DVAR assimilation we essentially follow the notation of Fukumori and Malanotte-Rizzoli (1995). Consider a linear approximation $\tilde{\mathbf{u}}$ of the original control vector $\mathbf{u}$ in a smaller dimensional subspace

$$\mathbf{u} - \bar{\mathbf{u}} \approx \mathbf{E}\tilde{\mathbf{u}}, \tag{1}$$

where $\mathbf{E}$ is the order-reduction operator defining the approximation and can be thought of as a reconstruction operator that maps the reduced-order control vector $\tilde{\mathbf{u}}$ on the original control space. Note that this approximation is defined, without loss of generality, around some prescribed reference (mean) state $\bar{\mathbf{u}}$ in anticipation of linearizing nonlinear models around such a reference. Let $\mathbf{E}$ be the pseudo-inverse matrix $\mathbf{E}^+$ of $\mathbf{E}$ defined such that*

$$\mathbf{E}^+\mathbf{E} = \mathbf{I} \quad \text{(but in general, } \mathbf{E}\mathbf{E}^+ \neq \mathbf{I}\text{)}, \tag{2}$$

where $\mathbf{I}$ is the identity matrix. This operator projects $\mathbf{u}$ onto the subspace spanned by the columns of $\mathbf{E}$ defining the reduced-order control vector $\tilde{\mathbf{u}}$ as an identity

$$\tilde{\mathbf{u}} = \mathbf{E}^+(\mathbf{u} - \bar{\mathbf{u}}). \tag{3}$$

The variational assimilation problem can then be simplified to a so-called reduced adjoint method that consists of seeking the reduced-order control vector $\tilde{\mathbf{u}}$ that minimizes the cost function of the variational problem (Durbiano, 2001). Once this solution is found, the original control vector can be recovered using (1). The gradient of the cost function $\mathbf{J}$ with respect to $\tilde{\mathbf{u}}$ can be obtained from the "original" adjoint model by projecting the "full" gradient onto the reduced control space according to

$$\left(\nabla_{\tilde{\mathbf{u}}}\mathbf{J}\right)^T = \mathbf{E}^T\left(\nabla_{\mathbf{u}}\mathbf{J}\right)^T. \tag{4}$$

From a practical point of view, the advantage of this set-up is that only two simple steps have to be added to the entire optimization procedure, and in a very simple manner one can go from a full order to a reduced order and vice versa. The computational cost of these operations is negligible compared to the integration cost of the numerical model (forward and backward). Starting from an initial reduced-order control vector, the first step consists of reconstructing the full order control vector using the approximated formula (1). As in the original adjoint method, the forward and backward models are then integrated to compute the gradient of the cost function with respect to $\mathbf{u}$. In the second step, the solution of the adjoint model is projected onto the reduced control space using (4). A descent algorithm is then used to adjust $\tilde{\mathbf{u}}$ in the reduced space to improve the model's performance. No changes are therefore required for the model and its adjoint; only the descent algorithm is now applied in the reduced space, with significantly fewer control variables. The order-reduction operator and its inverse are applied to the resulting adjoint gradients and the control vector prior to and after each linear search step, respectively.

The natural choice for preconditioning and order reduction derives from the background (control) error covariance matrix $\mathbf{B}$ (e.g., Fujii and Kamachi, 2003; Robert *et al*., 2005). This can indeed be expected to improve the overall behavior of the assimilation system because of the use of more sophisticated statistics in the 4DVAR problem. Such a choice of $\mathbf{B}$, however, should be considered very carefully as accurate estimates of this matrix statistics are rarely, if ever, available in real applications. As will become clear in Section 4, inaccurate choice of $\mathbf{E}$ would strongly limit the performance of the optimization because the null space orthogonal to $\mathbf{E}$, which will never be explored by the optimization, might contain important information about the descent directions. In order to explore these limitations, it is assumed in the present study that $\mathbf{B}$ is known a priori (diagonal in our case, but can be non-diagonal as well) and the adjoint gradients are projected to an independent reduced control space. This setup actually enables one to conduct a straightforward comparison between the full and the reduced 4DVAR approaches, which is otherwise not possible if $\mathbf{B}$ was a priori parameterized in the reduced space. Most importantly, and as suggested in Subsection 4.3, with this formulation one can still consider a strategy similar

*If $\mathbf{E}$ was generated by normalized EOFs, $\mathbf{E}^+$ is simply its matrix transpose.

to the so-called multi-grid optimization techniques (Nash, 2000) in which the optimization is started in the reduced space to speed up the adjustments during the early optimization steps and then to continue the optimization in the full control space in order to capture the missing variability in the reduced subspace. This study does not focus on the estimation of the background covariance matrix but on the possibility of reducing the computational burden of a pre-defined 4DVAR problem, although the link between the two problems is fully acknowledged.

3. Choice of the Reduced-Order Control Space

An order reduction of the control vector dimension is feasible in oceanic and atmospheric models because only a few modes are needed to represent most of the observed variability of these systems (De Mey, 1997). Different methods to construct a reduced space are available in practice. For example, one may apply the optimization on a coarser grid (Courtier *et al.*, 1994; Fukumori and Malanotte-Rizzoli, 1995). Such a method, however, would only reduce the dimension of the system by a factor of 10 or 100. Another possibility, used by Menemenlis and Wunsch (1997), is to apply a series of transformations based on temporal, vertical, and horizontal filters. Fast Fourier and Wavelet transforms are also alternatives to constructing a low-dimensional approximation of the control vector. In this paper, a statistical approach (the EOF analysis) has been adopted. This approach has already been used by Cane *et al.* (1996) and Pham *et al.* (1997) in the context of reduced-order Kalman filtering and by Durbiano (2001) and Robert *et al.* (2005) in the context of the reduced-order adjoint method. It is basically used to extract the dominant spatial patterns (or modes) of a system that explain the greatest amount of variability of a given set of system vector realizations. The reader is referred to Preisendorfer (1988) for an exhaustive discussion of this technique. Adjusting primarily the gravest modes of the control space that carry a sizeable amount of the control variance in space and time plus the time-mean biases, is indeed expected to enhance the system capability for spreading the information contained in the observations, and therefore to dramatically enhance the performance of the optimization procedure.

Assuming that a set of r EOFs, $\mathbf{E} = (\mathbf{e_1}, \ldots, \mathbf{e_r})$, has been determined, an approximation formula of the control vector is

$$\mathbf{u} \approx \bar{\mathbf{u}} + \sum_{i=1}^{r} \tilde{u}_i \mathbf{e_i} = \bar{\mathbf{u}} + \mathbf{E}\tilde{\mathbf{u}}. \quad (5)$$

Following this representation, the new control parameters are the coordinates of the reduced control vector $\tilde{\mathbf{u}} = (\tilde{u}_0, \ldots, \tilde{u}_r)$. The dimension of this new control space is therefore equal to the number of EOFs, r, used to represent the variability of the full control vector.

The EOFs need to be carefully chosen in order to efficiently represent the variability of the control parameters. The choice of the set of vectors from which the EOFs are computed is therefore key to building an efficient reduced adjoint method. In our case, the atmospheric forcing fields are control variables and the EOFs could be computed from a time series of forcing fields (e.g., NCEP) assuming that the structure of the forcing error is correlated with the variances of the forcing itself. Alternatively, one can compute EOFs of the adjustments to the prior forcing fields, obtained from an earlier optimization. To the extent that the adjustments represent errors in the forcing or the model itself, the EOFs can be expected to better represent the uncertainties in the control parameters and therefore should be more efficient in reducing any model-data differences. It is important to notice that since the optimization is carried out over a given assimilation period [1 T] and the absolute surface forcing fields are time-varying,

$$\mathbf{u} = \left(Wind_{stress}\left(\left[1 \;\; T\right]\right) \vdots Heat_{flux}\left(\left[1 \;\; T\right]\right) \vdots Salinity_{flux}\left(\left[1 \;\; T\right]\right)\right)^T. \quad (6)$$

In the above identity, $V([1\ T])$ represents all forcing fields V from time 1 to time T. To determine an approximation formula as in (5), the EOF analysis should be computed from a sample of vectors of the same "format" as $\mathbf{u}$. Each of these vectors should therefore contain a set of forcing fields over a [1 T] period. In our experiments, for example, where the assimilation period is one year, a time series of N years of NCEP forcing provides a sample of N vectors (one vector per year) to compute the EOFs.

4. Experiments

We used the ECCO ocean general circulation model, which is derived from the MIT model (Marshall *et al.*, 1997). An adjoint code to the forward model was obtained using an automatic differentiation tool (Giering and Kaminski, 1998; Marotzke *et al.*, 1999). To test the reduced-order adjoint assimilation approach, we investigate a simple box of the sub-tropical North Atlantic Ocean extending from 10°N to 40°N and from 42°W to 4°E. Zero fluxes of volume, heat and salt were applied to the closed lateral boundaries. The model is set up on a 2° × 2° horizontal grid and 23 vertical levels, with the first 6 levels in the upper 100 meters. The model domain has a realistic bottom topography based on the ETOPO5 (1988) dataset. Free-slip bottom boundary conditions and no-slip boundary conditions at lateral walls are applied. Laplacian viscosity and diffusivities are imposed, with $\nu_h = 1 \times 10^4$

m^2/s and $\kappa_h = 10^2$ m^2/s and $\nu_v = 10^{-3}$ m^2/s and $\kappa_v = 10^{-5}$ m^2/s, in the horizontal and vertical, respectively. The surface mixed layer is modeled with the "KPP" code of Large *et al.* (1994). The time step is 1 hour. Atmospheric forcing consists of daily heat and fresh water fluxes, and twice-daily zonal and meridional wind stress components from the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) re-analysis project (Kalnay *et al.*, 1996). The model is started from rest and Levitus temperature T and salinity S fields. Given the model set up in a regional box with closed lateral boundaries, it provides meaningful simulations of the circulation of the eastern North Atlantic only for limited periods. All the following experiments are limited therefore to one year duration.

Assimilation experiments were carried out over a one-year period during which the model was constrained by TOPEX and ERS and sea surface height (*SSH*) data, monthly Reynolds surface temperature (*SST*) data, and by subsurface Levitus S and T data. To eliminate errors associated with uncertainties in the geoid, the mean and time-varying components of the model SSH were separately constrained to daily along-track TOPEX/ERS SSH anomalies and mapped TOPEX mean SSH minus the Earth Gravitational Model 1996 (EGM96) geoid, respectively. In all experiments, the control vector consists of heat flux, fresh water flux, and the wind stress, which were adjusted every two days. Increment forcing values on non-adjustable days were linearly interpolated from the adjusted ones. NCEP forcing fields were provided as first guess for the optimized parameters; the fields were also used in the cost function as constraints on the adjusted surface fluxes. Data and model errors are prescribed only along the diagonal of the error covariances and are the same as used by Stammer *et al*. (2002) in a global approach. They have been approximated by the error profiles for temperature and salinity taken from Levitus data and by 50% of the data variability for the SSH. For the wind stress, prior errors are provided as standard deviation (STD) of the differences between NCEP and QuickSCAT scatterometer wind fields. One-third of the local STD of the NCEP forcing was used as the prior error for the net heat and freshwater fluxes. The explicit form of the cost function was then

$$\begin{aligned}\mathbf{J} = & \sum_{t=2*days}\left[Forcing(t) - NCEP(t)\right]^T \mathbf{B}^{-1}\left[Forcing(t) - NCEP(t)\right] + \left[\overline{SSH} - \overline{SSH}_{geoid}\right]^T \mathbf{R}_{geoid}^{-1}\left[\overline{SSH} - \overline{SSH}_{geoid}\right] \\ & + \sum_{t=days}\left[SSH'(t) - SSH'_{TP/ERS}(t)\right]^T \mathbf{R}_{TP/ERS}^{-1}\left[SSH'(t) - SSH'_{TP/ERS}(t)\right] \\ & + \sum_{t=months}\left[SST(t) - SST_{Rey}(t)\right]^T \mathbf{R}_{Rey}^{-1}\left[SST(t) - SST_{Rey}(t)\right] \\ & + \sum_{t=months}\left[T(t) - T_{Lev}(t)\right]^T \mathbf{R}_{Lev:T}^{-1}\left[T(t) - T_{Lev}(t)\right] + \sum_{t=months}\left[S(t) - S_{Lev}(t)\right]^T \mathbf{R}_{Lev:S}^{-1}\left[S(t) - S_{Lev}(t)\right], \end{aligned} \quad (7)$$

where **B** and **R** are diagonal matrices. The descent directions toward the minimum of **J** have been determined using the Quasi-Newton M1QN3 optimization algorithm which has been developed by Gilbert and Le Maréchal (1989). The standard technique of preconditioning with the background error covariance matrix was applied for all experiments as described by Courtier (1997).

Results of assimilation experiments performed in 1993 are now presented. A summary of the different EOFs sets and the different assimilation experiments is provided in Tables 1 and 2.

Table 1. Summary of the different EOFs sets.

EOFs	Time series
TS1	NCEP forcings 1992 → 2001
TS2	ECCO forcings 1992 → 2001
TS3	1992's prior control estimates

4.1 *Sensitivity to the choice of EOFs*

To test the sensitivity of the performance of reduced-order optimizations to the choice of the EOFs-based order-reduction operator **E**, three distinct reduced-order control subspaces were determined by applying a multivariate EOF analysis on three different samples of control vectors. The first two sets of EOFs (TS1) and (TS2) were computed from: 10 years of NCEP forcing fields (which are used to force the model), and 10 years of ECCO global adjustments to the NCEP forcing fields, respectively. The latter have been optimized with a glo-

Table 2. Summary of the different assimilation experiments.

Experiment	Method	Number of EOFs	EOFs set
Sensitivity to choice of EOFs	Reduced	6	TS1
	Reduced	6	TS2
	Reduced	4	TS3
Sensitivity to number of EOFs	Reduced	2	TS3
	Reduced	4	TS3
	Reduced	6	TS3
Hybrid reduced adjoint method	Classical (C1)	—	—
	Reduced	4	TS3
	Reduced + Classical (H1)	4	TS3

bal state estimation procedure on a 2° × 2° global grid using the MIT model and a similar configuration to the one used in the current study as described by Stammer *et al.* (2002). Following the discussion in Section 2, the available NCEP and ECCO forcing provide two samples of only 10 vectors (one vector per year) to compute the EOFs. A third set of EOFs (TS3) was computed from a set of control vectors estimated by a prior optimization performed during 1992 using the same limited-area model and the same setup as the one used in our assimilation experiments. In this experiment, forty iterations were required to obtain an acceptable convergence using the full dimension of the control space. These three sets of EOFs are used in the following to assess the convergence rate of an optimization performed during 1993, and to study its sensitivity to the choice and the number of EOFs. Note that TS3 was computed from the optimized control vectors in 1992 to provide an "independent" set of EOFs for the experiments with the reduced-order 4DVAR approach in 1993. To account for approximately 90% of the total variance for each set, 6, 6 and 4 EOFs were retained from TS1, TS2 and TS3, respectively. The optimization was therefore performed in subspaces of dimension 6 and 4 instead of 284,700.

The decrease of the total cost function with the number of iterations is plotted in Fig. 1 for all three reduced 4DVAR experiments using TS1, TS2 and TS3 and evaluated against the one obtained with the classical 4DVAR run C1 (i.e., full control space). The figure shows that the performance of the reduced adjoint method depends closely on the choice of the reduced space and that the overall performance of the adjoint method is always superior when the optimization is performed in the full control space. Although a faster convergence rate is achieved during the first iterations (see below for more details), the cost-function decrease stagnates after only 8 iterations once optimization explored all descent directions generated by the few retained EOFs, suggesting that

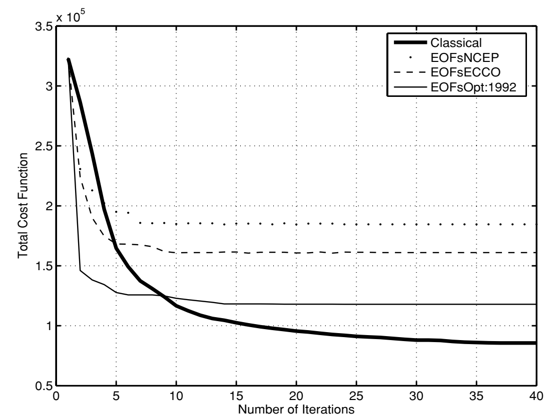

Fig. 1. Total cost function vs. the number of iterations as obtained with the reduced adjoint method using 4 EOFs computed from NCEP forcing (TS1), from ECCO forcing (TS2) and from prior 1992 control vectors (TS3).

a large fraction of the variability of the forcing error is not well represented by these functions. More precisely, the use of the EOFs computed from the ECCO forcing adjustments (TS2) clearly improves the assimilation results as compared to TS1. This is to be expected since the ECCO solution is already an estimate of the forcing errors and as such should be a better representation of the gravest forcing error modes, while the NCEP EOFs represent the entire forcing process variance. However, the overall performance of this method remains unsatisfactory, even for experiment TS2. Instead, the best assimilation results were obtained when the EOFs were computed from the prior optimized control vectors obtained from the same model setup in 1992. The use of these EOFs reduces the final cost function value by more than 35% compared to the other two sets of EOFs. This is because the TS3 approximation contains information from both the regional closed-boundary model and the assimilated

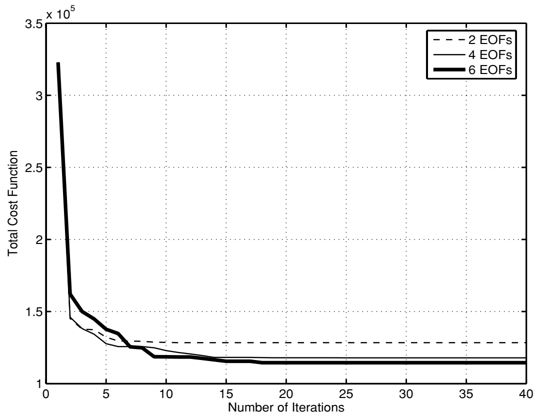

Fig. 2. Total cost function vs. the number of iterations as obtained with the reduced adjoint method using 2 EOFs, 4 EOFs and 6 EOFs computed from prior iterations in 1992 (TS3).

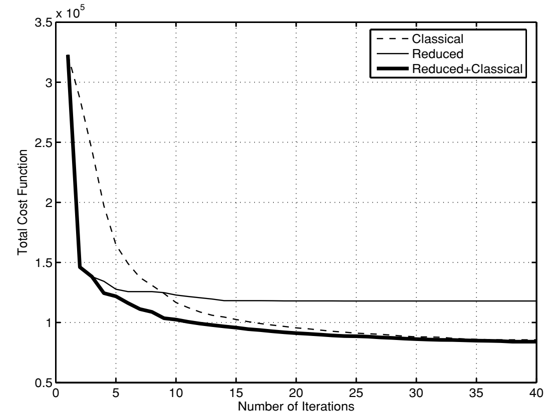

Fig. 3. Total cost function vs. the number of iterations as obtained with (i) classical adjoint method, (ii) the reduced adjoint method (using 4 EOFs from TS3), and (iii) the reduced adjoint method (using 4 EOFs from TS3) at the first 3 optimization iterations and then the classical adjoint method.

observations, providing a better representation of control vector modes.

4.2 Sensitivity to the number of retained EOFs

Although the previous reduced-order optimization strategy seems to provide some benefit during the first iterations, they obviously do not have sufficient degrees of freedom to bring the model as much into consistency with the data as the full optimization does. Instead, the decrease of the cost function in the reduced space stagnates quickly after all information contained in the gravest error modes has been used. In other words, the order-reduction eliminates corrections of the forcing on larger spatial scales and the success of the reduced adjoint method therefore depends on the number of EOFs that are actually included in the optimization effort.

To investigate the truncation effect of the reduced space on the final optimization results we investigate here the sensitivity of the convergence of the reduced adjoint method to the number of retained EOFs (dimension of the reduced space). In the following, all experiments are based on EOFs obtained from the control vectors of the previous optimization over 1992 (TS3). We successively use 2 EOFs, 4 EOFs and 6 EOFs, which account for approximately 76%, 90% and 93% of the error variance estimated during 1992, respectively.

Results are shown in Fig. 2 in terms of the evolution of the total cost function as a function of iterations performed. Not unexpectedly, the final performance of the optimization is shown to degrade significantly with fewer numbers of retained EOFs; however, the initial adjustments are fastest as the dimension of the reduced space decreases. Therefore the number of retained EOFs should be small enough to assure a fast convergence rate at the beginning of the optimization and large enough (up to a certain level, since the last EOFs generally represent noise) for an efficient representation of the variability of the full control space.

Despite a greatly improved convergence rate during the first iterations, the above experiments reveal that the overall performance of the reduced adjoint method calls for further improvements, even with 6 EOFs which account for 93% of the total variance of TS3, as compared to the classical solution of C1. The improvements are also shown to stagnate very rapidly as the number of retained EOFs increases. This suggests that the performance of the reduced-order optimization scheme will always be limited by the representativeness of the prior estimate of the control subspace.

4.3 A hybrid reduced adjoint method

Figure 3 suggests that a way to further improve the performance of the adjoint method is actually to start with a reduced-order approach to first adjust the control parameters on their graves modes, and to continue the optimization subsequently in the full control space to adjust important smaller-scale errors that are missing in the reduced space. The additional effect is that while a reduced-order approach relies on the prior error estimate and its scales, the full-state approach further adjusts the control variables according to the actual model setup and the actually used data constraints.

To study the performance of this hybrid strategy, an extra experiment (H1) was performed that starts with a reduced-order adjoint method for the first 3 optimization iterations using 4 EOFs (describing 90% of the total vari-

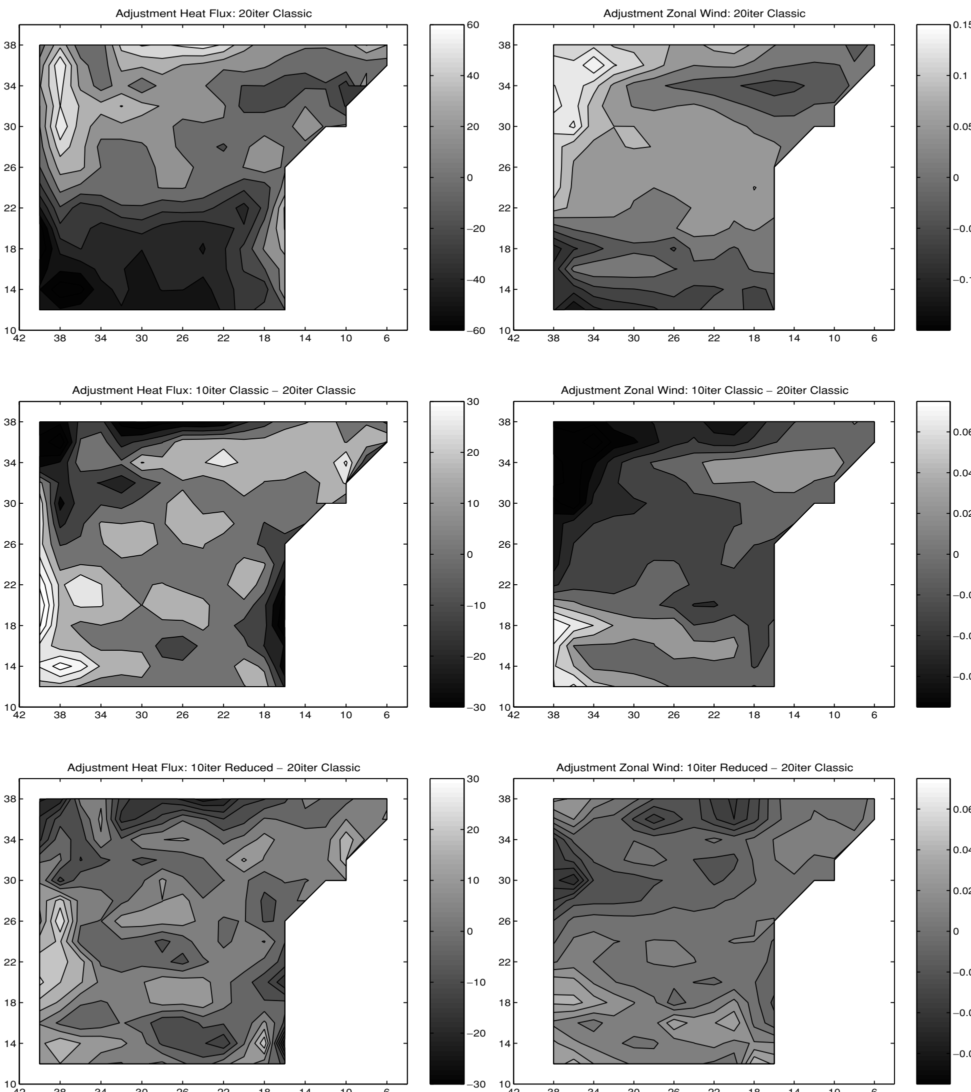

Fig. 4. *Left panel*: Time average adjustments to the heat flux (in W/m^2) after 20 iterations with the classical adjoint method (top), difference between average adjustments after 20 iterations and 10 iterations with the classical adjoint method (middle), and difference between average adjustments after 20 iterations with the classical adjoint method and 10 iterations with the reduced adjoint method (low). *Right panel*: Same as the left panel but for the time average adjustments to the zonal component of the wind stress in N/m^2.

ance of the prior estimates in 1992) and subsequently continues with the "classical" adjoint method in the full space. Results for H1 are included in Fig. 3. As can be seen from the figure, the classical adjoint method leads to a constant decrease of the cost function, reducing it by a factor of 4 after 40 iterations. The results of the hybrid approach show that by continuing the optimization in the full space after 3 initial reduced-space iterations, one can get a quite similar solution but with fewer iterations. As an example, the cost function value obtained with C1 after 20 iterations is reached in experiment H1 after only 10 iterations.

For a more detailed, quantitative comparison of the solutions from C1 and H1, Fig. 4 shows a plot of the time-mean adjustment to the heat flux and of the zonal component of the wind stress obtained after 20 iterations with C1 and compared them with similar fields obtained after 10 iterations with H1. The figure reveals that the corrections of the forcing fields from both runs are quite similar, with differences residing primarily at small spatial scales, probably not represented in the EOFs subspace. For comparison, the adjustments obtained after only 10 iterations of C1 (also shown) still look quite distinct from the results obtained after 20 or 40 iterations.

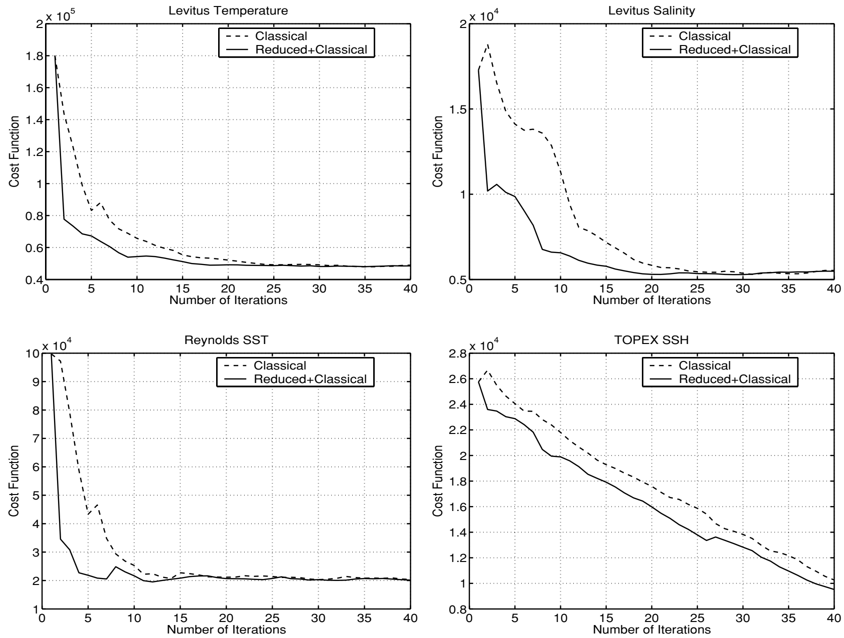

Fig. 5. Contributions of the individual data misfit terms to the total cost function vs. the number of iterations as they result from the classical adjoint method and the reduced adjoint method (using 4 EOFs) at the first 3 optimization iterations and then the classical adjoint method.

A more rapid convergence of the cost function was not only obtained for the total cost, but also for each individual contribution during experiment H1, including full depth potential temperature and salinity, SST and SSH. The decrease, however, is found to be different for each model state variable (Fig. 5). Quite clearly, the largest initial misfit is due to deviations of the model's temperature field from the Levitus and Reynolds data sets. Accordingly, the optimization tried to remove those large misfits during the first iterations. The associated adjustment of the model's temperature field seems to be stabilized after about 15 iterations with the classical adjoint method. However, more iterations were needed to correct errors in the salinity field and particularly in the SSH field. The improved convergence rate of H1 is particularly clear from the salinity misfit: fewer than 10 iterations are now required to reduce the initial salinity misfit to a level that was reached only after almost 20 iterations during C1: In H1 the rate of salinity improvement became comparable to the temperature's rate. This can be explained by an efficient propagation of the information extracted from the different data sets through a strong EOFs-coupling of the fresh water flux with the other control variables. In stark contrast, the slow adjustment of the *SSH* field in C1 did not change fundamentally during experiment H1.

Finally, Fig. 6 shows in the left column the difference between the time-mean Reynolds SST fields over 1993 minus the models SST obtained during C1 after 10 iterations. The lower two panels in the same column show similar differences by using model results from C1 after 20 iterations and from H1 after 10 iterations. The right column shows similar difference fields, but now from a meridional section of the salinity field along 32°W relative to the time-mean Levitus salinity (reference). The comparison of both variables illustrates that model-data differences obtained after 20 iterations from C1 and after 10 iterations from H1 look very similar and are both significantly smaller than differences found during iteration 10 of C1. This holds for surface and subsurface values.

5. Discussion

We have shown here that a reduced-order optimization method, which limits the size of the control vector of a 4DVAR ocean assimilation system by projecting it onto a small number of basis functions, can speed up the initial convergence rate of an assimilation effort in the eastern subtropical North Atlantic using real *in situ* and satellite data as constraints. The following findings are noteworthy:

Using a reduced-order control space, a significant speed-up of the convergence was found only during the initial first few iterations. Here the actual choice of EOFs was important for the convergence rate. In this paper we used three different sets of EOFs: one based on NCEP reanalysis forcing fields, one based on the control vector

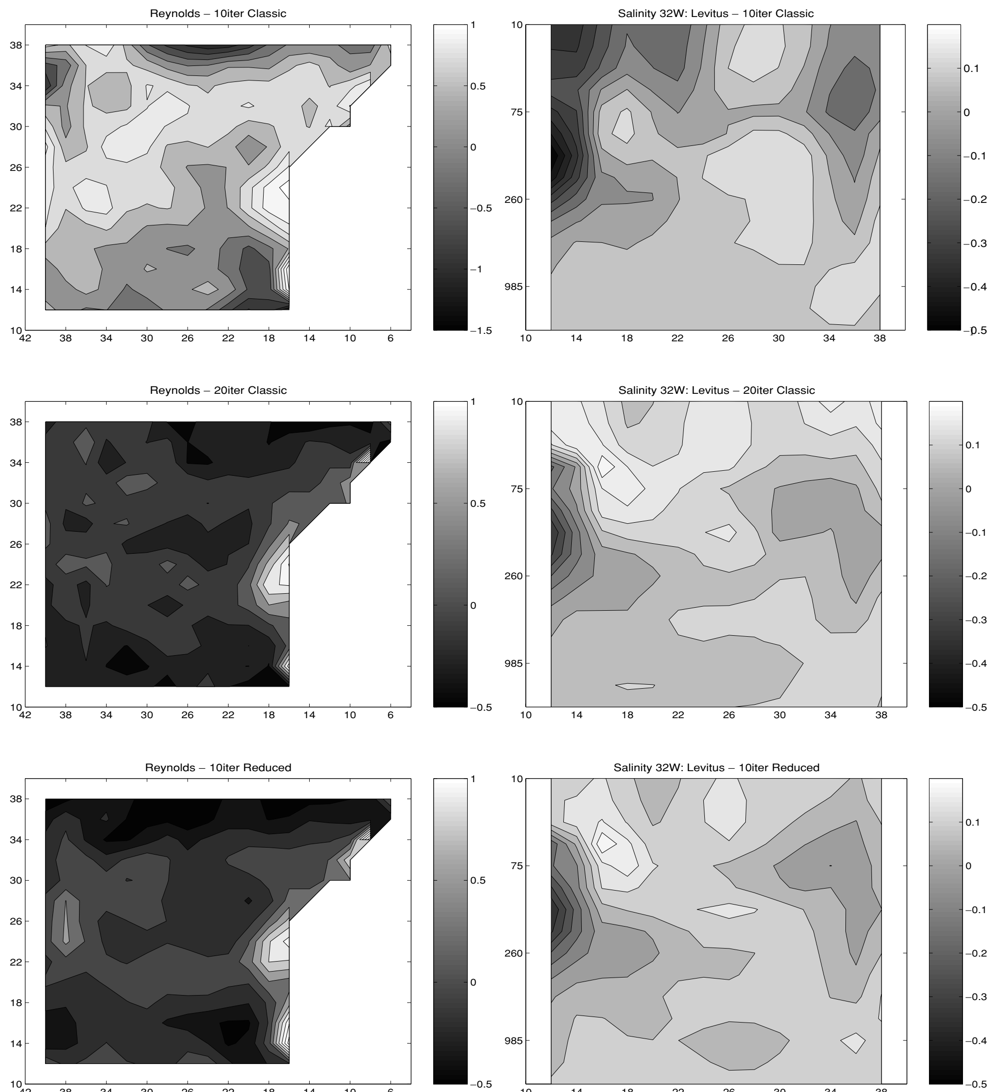

Fig. 6. *Left panel*: Difference of the time average between Reynolds and estimated SST after 10 iterations (top in °C), 20 iterations (middle) with the classical adjoint method, and 10 iterations with the reduced adjoint method (bottom). *Right panel*: Difference of the time average between salinity cross-section at 32°W from Levitus and estimated Salinity in PSU after 10 iterations (top), 20 iterations (middle) with the classical adjoint method, and 10 iterations with the reduced adjoint method (bottom).

of a previous global optimization using essentially the same model setup and data constraints, and one using the control vector obtained from a previous optimization using the same regional model setup. The best performance was obtained in the latter case, which accounts for both forcing errors and errors of the regional model set up alike. The results show that the chosen EOFs should give the best possible representation of the errors of the control parameters—in our case the surface forcing fields. We can conclude that using the gravest modes of the forcing variability as a prior estimate of the forcing errors was not a useful approach. To some extent this is due to the fact that the forcing variability and the forcing errors do not have the same space-time scales. It can also mean that internal model errors compensated by changing the surface forcing fields during the optimization have a significant effect on the parameter adjustments.

While the use of a reduced-order control subspace does speed up initial convergence rate of the optimization by adjusting the estimation parameters in the directions of the gravest modes of the control space, such an optimization never actually reached the quality of the full optimization. Instead, the optimization stagnated fairly early on and never reached a final solution that was any-where near the one obtained with a full control space. A full control space is also required to adjust finer scales of the control parameters representing smaller scales in the forcing errors over the open ocean or near boundaries. A hybrid strategy was therefore found to be most effective in assimilating real data while controlling atmospheric forcing fields: it starts in a reduced space and finishes the optimization subsequently in full space. Such an approach can be viewed as a way of obtaining the solution of the classical adjoint method, but with fewer iterations: roughly 50% of the iterations that were required to reach the quality of a convergence solution using the full control space have been saved in our simple setting using this approach. Although in this study the control vector was composed of atmospheric forcing fields only, results can easily be extended to include the model initial temperature and salinity conditions. Moreover, although the assimilation experiments were conducted here in a simple configuration of the North Atlantic ocean, they do demonstrate the effectiveness of the reduced-order optimization strategy.

The fact that the best result was obtained by projecting the control vector onto EOFs of the parameter errors estimated previously from an optimization using the same model configuration means that an important prerequisite for speeding up an optimization is a good prior knowledge of parameter errors and their structures in space and time. However, a serious drawback is that a previous assimilation experiment was needed. But while this seems impractical at first glance, in practice it is still very helpful: any sustained estimation effort, whether for the purpose of re-analysis or routine forecast, needs to extend a previous optimization in time or needs to improve a previous optimization by repeating it with a higher spatial resolution or with more complete data constraints. In all those cases it would still be helpful to use the EOFs of the previous control vector to obtain a significant speed-up of the optimization.

In some sense the existing ECCO efforts (Stammer *et al.*, 2002; Köhl *et al.*, 2006) have made use of this fact already by using the time mean forcing correction estimated during a previous optimization as priors for the parameter adjustments in a new optimization (on higher resolution or over longer periods). The extent to which the respective speed-up could be further enhanced by using several EOFs as proposed here still has to be evaluated on a global scale. But the ECCO experience can hold already as a first indication that the results obtained here in a small basin do have relevance for global ocean reanalyses as well.

Acknowledgements

Re-analysis surface forcing fields from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) were obtained through a computational grant at NCAR. The adjoint compiler TAF from the company FastOpt was used to create the adjoint general circulation model. Computational support through a NRAC grant from the National Partnership for Computational Infrastructure (NPACI) and through a grant from the National Center for Atmospheric Research is acknowledged. Supported in part through NASA grant NAG5-7857, through NSF grant OCE 9730071, and through two contracts with the Jet Propulsion Laboratory (958125 and 1205624). This is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean (ECCO) funded by the National Oceanographic Partnership Program under grants N00014-99-1-1049.

References

Bennett, A. (1992): *Inverse Methods in Physical Oceanography*. Cambridge University Press, 346 pp.

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

Cohn, S. E. and R. Tolding (1996): Approximate Kalman filters for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.

Courtier, P. (1997): Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123**, 2449–2461.

Courtier, P., J.-N. Thèpaut and A. Hollingsworth (1994): A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.

Daley, R. and E. Barker (2000): NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*, **129**, 869–883.

De Mey, P. (1997): Data assimilation at the oceanic mesoscale: A review. *J. Met. Soc. Japan*, **75**(1B), 415–427.

Durbiano, S. (2001): Vecteurs caractéristiques des modèles océaniques pour la réduction d'ordre en assimilation de données. Ph.D. thesis, University of Grenoble.

ETOPO5 (1988): Digital relief of the surface of the earth. worldwide bathymetry/topography data announcement 88-mgg-02. Technical report, National Geophysical Data Center.

Fujii, Y. and M. Kamachi (2003): A reconstruction of observed profiles in the Sea East of Japan using vertical coupled temperature-salinity EOF modes. *J. Oceanogr.*, **59**, 173–186.

Fukumori, I. and P. Malanotte-Rizzoli (1995): An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.*, **100**, 6777–6793.

Ghil, M. and P. Malanotte-Rizzoli (1991): Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141–266.

Giering, R. and T. Kaminski (1998): Recipes for adjoint code construction. *ACM Trans. On Math. Software*, **24**, 437–474.

Gilbert, J. C. and C. Le Maréchal (1989): Some numerical experiments with variable storage Quasi-Newton algorithms.

Math. Program., **45**, 407–435.
Golub, G. and C. Van loan (1996): *Matrix Computations*. 3rd ed., The Johns Hopkins University Press, Baltimore, U.S.A., 694 pp.
Hoteit, I., D.-T. Pham and J. Blum (2002): A simplified reduced-order kalman filtering and application to altimetric data assimilation in tropical pacific. *J. Mar. Syst.*, **36**, 101–127.
Kalnay and co-authors (1996): The NCEP/NCAR re-analysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
Köhl, A., D. Stammer and B. Cornuelle (2006): Interannual to decadal changes in the ECCO global synthesis. *J. Phys. Oceanogr*. (in press).
Large, W., J. McWilliams and S. Doney (1994): Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, **32**, 363–403.
Marotzke, J., R. Giering, Q. K. Zhang, D. Stammer, C. N. Hill and T. Lee (1999): Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Res.*, **104**, 29,529–29,548.
Marshall, J., A. Adcroft, C. Hill, L. Perelman and C. Heisey (1997): A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, **102**, 5753–5766.
Menemenlis, N. and C. Wunsch (1997): Linearization of an oceanic general circulation model for data assimilation and climate studies. *J. Atmos. Oceanic Tech.*, **14**, 1420–1443.
Nash, S. G. (2000): A multigrid approach to discretized optimization problem. *J. Comp. Appl. Math.*, **14**, 99–116.
Pham, D. T., J. Verron and M.C. Roubaud (1997): Singular evolutive Kalman filter with EOF initialization for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323–340.
Preisendorfer, R. (1988): Principal component analysis in meteorology and oceanography. *Elsevier Sci. Publ.*, **17**, 425 pp.
Robert, C., S. Durbiano, E. Blayo, J. Verron, J. Blum and F.-X. Le Dimet (2005): A reduced-order strategy for 4D-Var data assimilation. *J. Mar. Syst.*, **57**, 70–82.
Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C. N. Hill and J. Marshall (2002): The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.*, **107**(C9), 3118.
Vidard, P. A., E. Blayo and F.X. Le Dimet (2000): 4D variational data analysis with imperfect model. *Flow. Turbul. Combust.*, **65**, 489–504.
Vogeler, A. and J. Schröter (1995): Assimilation of satellite altimeter data into an open ocean model. *J. Geophys. Res.*, **100**, 15951–15963.
Wunsch, C. (1996): *The Ocean Circulation Inverse Problem*. Cambridge Univ. Press, New York, 442 pp.
Yang, W. Y., I. M. Navon and P. Courtier (1996): A new Hessian pre-conditioning method applied to variational data assimilation experiments using NASA general circulation models. *Mon. Wea. Rev.*, **124**, 1000–1017.