



HAL
open science

Assimilation of sea-surface temperature and altimetric observations during 1992-1993 into an eddy-permitting primitive equation model of the North Atlantic Ocean

Charles-Emmanuel Testut, Pierre Brasseur, Jean-Michel Brankart, Jacques Verron

► To cite this version:

Charles-Emmanuel Testut, Pierre Brasseur, Jean-Michel Brankart, Jacques Verron. Assimilation of sea-surface temperature and altimetric observations during 1992-1993 into an eddy-permitting primitive equation model of the North Atlantic Ocean. *Journal of Marine Systems*, 2003, 40-41, pp.291-316. 10.1016/S0924-7963(03)00022-8 . hal-00212078

HAL Id: hal-00212078

<https://hal.science/hal-00212078>

Submitted on 12 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean

Charles-Emmanuel Testut*, Pierre Brasseur, Jean-Michel Brankart, Jacques Verron

Laboratoire des Ecoulements Geophysiques et Industriels (LEGI), UMR 5519, CNRS, B.P. 53X, 38041 Grenoble Cedex, France

Sea-surface temperature (SST) and sea-surface height (SSH) observations collected from space between October 1992 and December 1993 have been assimilated into a realistic primitive equation model of the North Atlantic Ocean circulation at eddy permitting resolution. The assimilated SST data originate from AVHRR observations gathered and processed within the NASA Pathfinder project; the altimetric data consist of SSH maps computed as the sum of a time-invariant dynamic topography and gridded sea-level anomalies obtained by combining Topex/Poseidon and ERS altimeter data. The assimilation scheme is a reduced-rank Kalman filter derived from the Singular Evolutive Extended Kalman (SEEK) methodology [J. Mar. Syst. 16 (1998) 323], in which the error statistics is represented in a subspace of small dimension. The error subspace is initialized with a truncated series of Empirical Orthogonal Functions (EOFs) of the system variability. The analysis algorithm includes a mechanism to update the forecast error statistics adaptively using all pertinent informations from the innovation vector. Hindcast experiments have been conducted with a $1/3^\circ$ model of the North Atlantic basin forced with ECMWF atmospheric reanalyses. The impact of the data assimilated during 1993 is assessed by examining how observed (SSH and SST) and nonobserved variables (such as velocity and thermohaline properties in the interior of the ocean) are modified by the assimilation scheme. Finally, the validation of the hindcast experiments with independent XBT measurements is performed in order to evaluate the objective skill of the procedure. The various diagnostics demonstrate the positive impact of the satellite data to hindcast the upper ocean circulation at eddy permitting resolution and the capacity of the scheme to estimate the geographic distribution of the forecast error.

Keywords: Data assimilation; North Atlantic Ocean; Operational oceanography; SEEK

1. Introduction

Assimilation algorithms are at the hearth of operational ocean prediction systems that will be operated

in the future on a routine basis, at resolutions fine enough to represent oceanic eddies.

The theoretical framework of data assimilation in meteorology and oceanography is now well established: variational methods seek to minimize the misfit between data and model simulations by optimization of well-chosen sets of control parameters, while sequential methods proceed by intermittent

* Corresponding author.

E-mail address: testut@hmg.inpg.fr (C.-E. Testut).

blending of observations and model solutions according to their respective accuracy.

The generalized inverse (Bennett, 1992) or adjoint variational methods (Courtier, 1997; Rabier et al., 2000) can be used to optimize the initial conditions and the atmospheric forcings of an ocean simulation, assuming that the model describes the dynamics perfectly (e.g., Lee and Marotzke, 1998; Wenzel et al., 2001; Stammer et al., 2003) or unperfectly (e.g., Bennett et al., 1998). In order to ensure rapid convergence of the minimization process, the model dynamics has to be linear or weakly nonlinear. These methods have been applied successfully in the context of equatorial (Weaver et al., 2002) or global models at coarse resolution. In the case of strongly nonlinear flows, specific temporal strategies (Luong et al., 1998) or iterative algorithms (Chua and Bennett, 2001) are needed to cope with the problem of local minima.

Using the sequential estimation theory, ensemble Kalman filters and smoothers, which propagate the error statistics by simulating a limited set of model trajectories simultaneously, have been investigated in the context of nonlinear ocean dynamics (Evensen, 1994; Burgers et al., 1998; Evensen and van Leeuwen, 2000). Like variational algorithms, the practical implementation of these methods show that the numerical resources needed to optimally combine nonlinear ocean models with observations are equivalent to, at least, a few tens of model integrations (Brusdal et al., 2003). Considering the computing power available today, and the need to operate high-resolution ocean models at global scales routinely (e.g., every week), cheaper assimilation methods are required to demonstrate the feasibility of operational ocean prediction systems.

Several studies have examined how simplified representations of the estimation error statistics can reduce the computational burden of the conventional Kalman filter with nonlinear models in academic configurations (e.g., Fukumori and Malanotte-Rizzoli, 1995; Pham et al., 1998; Voorrips et al., 1999), while applications of reduced-order Kalman filters into realistic models of the Tropical Pacific Ocean (Fukumori, 1995; Cane et al., 1996; Verron et al., 1999) have demonstrated the usefulness of the concept in quasi-linear regimes.

Concerning the operational prototypes in use today, most of them assimilate data with suboptimal interpolation methods wherein the background error

covariances are derived from fairly simple schemes (De Mey et al., 2002; Bell et al., 2000; De Mey and Benkiran, 2001; Smedstad et al., 2003). These simplified methods are easy to set up and computationally efficient, but they require external estimates of error covariances. As a result, the assimilation increments may not be consistent dynamically, and relevant statistically. A better use of the data in those systems can thus be expected from “advanced methods”, which aim at more dynamically and statistically consistent error covariances and state estimates. In the framework of reduced-order Kalman filters, this goal can be achieved by specifying multivariate error subspaces with dynamically balanced covariances, and by using all pertinent informations from the model and the data to propagate the error statistics during the assimilation period.

This issue has motivated the present study, which examines the implementation of an advanced statistical method into a realistic primitive equation model of the North Atlantic. A reduced-order Kalman filter derived from Singular Evolutive Extended Kalman (SEEK) (Pham et al., 1998) is used to assimilate sea-surface temperature (SST) and sea-surface height (SSH) data collected from space between October 1992 and December 1993. Potentially, other data types such as surface salinity (Durand et al., 2002) and hydrographic profiles should also be included to meet operational requirements. However, a first objective of the present work will be to assess, in a pre-operational context, the benefit of satellite data assimilation by comparison with independent in situ data.

The analysis step of the SEEK filter is achieved in a subspace of small dimension which contains the dominant directions of the background error (Pham et al., 1998; Ballabrera et al., 2001). The error subspace used here is initialized with a truncated series of three-dimensional, multivariate Empirical Orthogonal Functions (EOFs) of a previous model simulation; this technique allows us to specify error covariances consistently with the model dynamics. In addition, the SEEK algorithm includes an adaptive mechanism to update the error subspace with all pertinent informations left in the innovation vector after the analysis step (Brasseur et al., 1999; Brankart et al., 2003). A second objective will be to examine the capacity of the scheme to propagate the error statistics in a consistent way throughout the assimilation period,

and to determine its dominant characteristics in a realistic North Atlantic model.

The paper is organized as follows. In Section 2, the model configuration adopted for the assimilation experiment is described. The main characteristics of the data sets available for assimilation during 1992 and 1993 are reviewed in Section 3, and an evaluation of the free model with respect to these data is discussed. The specific aspects of the SEEK assimilation system implemented for this study are described in Section 4. Section 5 is dedicated to the analysis of one major assimilation experiment of real SST and SSH satellite data. Finally, concluding remarks are given in Section 6.

2. The model configuration

The assimilation experiments examined in this paper have been carried out with a primitive equation model of the North Atlantic basin implemented and validated by the French CLIPPER project (Treguier et al., 1999).

The numerical code adopted to simulate the ocean circulation is OPA 8.1, a z -coordinate, primitive equation model developed at LODYC (Madec et al., 1998) that uses the hydrostatic approximation and the rigid-lid formulation. Vertical mixing of momentum, temperature and salinity is computed according to the TKE closure model developed by Blanke and Delcluse (1993), with enhanced turbulent viscosity in case of convective situations.

The model domain covers the North Atlantic basin from 20°S to 70°N and from 98.5°W to 20°E , with a horizontal resolution of $1/3 \times 1/3^{\circ} \cos(\text{latitude})$ (Fig. 1). The vertical discretization is achieved on 42 geopotential levels, with a grid spacing that increases from 12 m at the surface to 200 m below 1500 m depth. The bathymetry is derived from a smoothing of the bottom topography prepared by Smith and Sandwell (1997). The Southern boundary at 20°S , the Northern boundary at 70°N and the Gibraltar Strait are closed but the model solution is relaxed toward climatology within buffer zones defined off Portugal, in the Norwegian Sea and along the Southern boundary to simulate the supply of Mediterranean Water and the exchange with the Arctic and South Atlantic basins. The choice of an eddy permitting resolution

resulted from a balance between the need to represent the mesoscale turbulence fairly well and the computer resources needed to assimilate the variability signal observed by the satellites.

The thermodynamic variables in the model are initialized using a Northern hemisphere winter season state from the climatology compiled by Reynaud et al. (1998). The atmospheric forcing fields of heat, freshwater and momentum are derived from the reanalyses of the ECMWF 6-h forecasts of the 1979–1993 period. In addition, the model surface temperature is relaxed toward weekly Reynolds SST data in order to maintain interactivity between the ocean model and the atmosphere (Barnier et al., 1995). However, the relaxation term is removed during the assimilation experiments in order to avoid spurious competition with the assimilation of SST data. This also allows an easier diagnostic of the actual impact of the assimilation.

The model was spun up for 8 years, starting from rest in 1985. Several tests were performed to determine the most appropriate spin-up length, taking into account the existence of a slow model drift and the need to compute realistic modes of variability for initialization of the assimilation experiments. We realized that, by comparison with the Reynaud seasonal climatology, too long numerical integrations tend to deteriorate the three-dimensional distribution of temperature and salinity; on the other hand too short integrations leave some dynamical features unadjusted. A spin-up of 8 years was eventually found as a good trade-off between dynamical consistency and fit to climatology.

The integration was pursued until December 1993 to produce a reference run available for further comparison with the assimilation experiment. The barotropic stream function averaged between 1989 and 1993 is illustrated in Fig. 1. Some well-known elements of the North Atlantic circulation are easily recognized, such as the subtropical gyre, the intensification of western boundary currents along the American coast, and the subpolar gyre in the Labrador Sea.

A more detailed examination of the currents, however, reveals a number of unexpected features: the NAC is not well defined; the intensity of the subpolar gyre is too strong; the Gulf Stream shows a spurious recirculation southward with a permanent eddy at 35°N , and a bifurcation of the flow near the Hateras

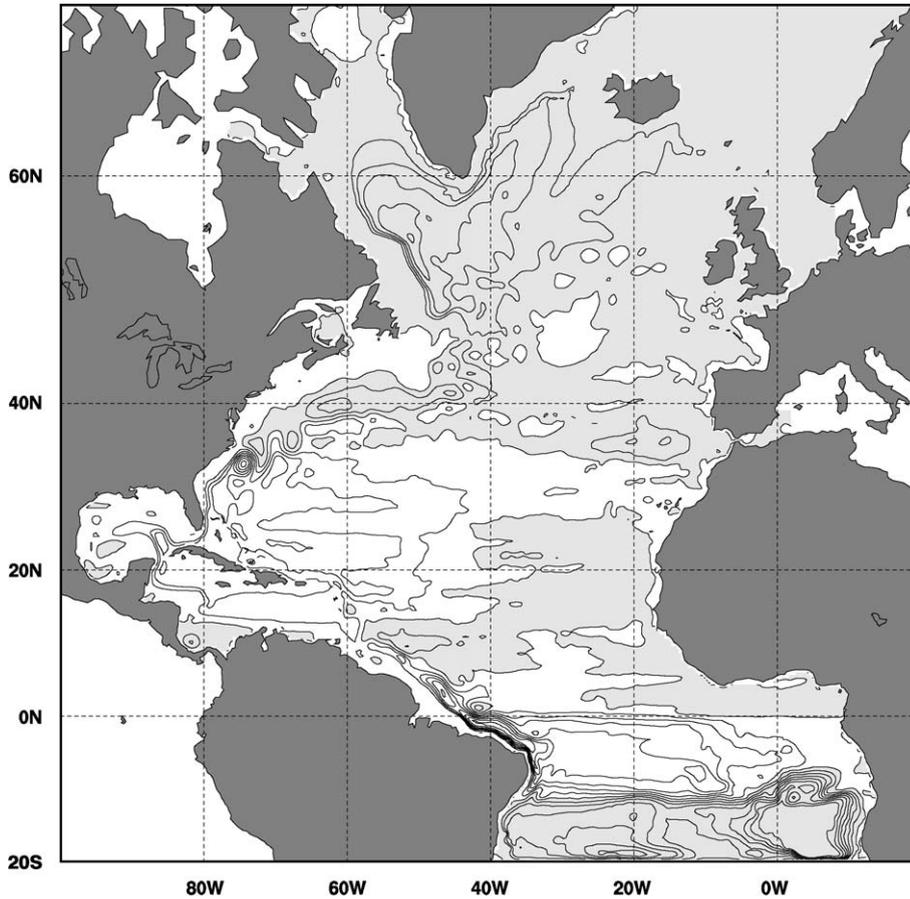


Fig. 1. Barotropic stream function of the NATL3 reference run averaged between 1989 and 1993 (CI=10Sv).

Cape advects warm surface water northward on the continental shelf (see also Fig. 8).

The insufficient resolution of the numerical model probably bears part of the responsibility in those unexpected features, though higher resolution configurations still exhibit similar deficiencies (Treguier et al., 1999). One can thus expect improvements in the representation of the currents by assimilating satellite data to constrain the flow field.

3. Sea-surface temperature and altimetric data sets

The time window chosen for the assimilation experiments extends from October 1992 to December 1993, i.e., a period during which both ECMWF

reanalysis, SST and SLA data products are available simultaneously.

The SST data consist of composite AVHRR observations, gathered and processed within the NASA Pathfinder project. In addition, these products have been quality controlled, compared with Reynolds analyses and gridded on maps at $1/4^\circ$ resolution every 10 days. In principle, AVHRR observations cover the entire model domain, but the presence of clouds, to some extent, restricts the practical availability of SST data at high latitudes and during the winter season. An accuracy of 0.5°C on the gridded SST products has been assumed, which corresponds to the sum of errors in acquisition procedures, processing algorithms and space–time interpolation.

Before the assimilation experiments, SST data have been used to get a prior evaluation of the reference run during 1993. Fig. 2 illustrates the bias between the model SST and the satellite observations mapped on the model grid. In spite of the flux correction applied on the surface temperature, a number of strong anomalies can be pointed out: the subpolar gyre in the model is too warm of ~ 2 to 4 $^{\circ}\text{C}$ in average near the surface; a large misfit also occurs near the African coast, reflecting a weakness in the representation of the African Upwelling off Senegal; negative values of -2 to -4 $^{\circ}\text{C}$ are observed in Southern sector of the Gulf Stream, while an excess of 4 to 7 $^{\circ}\text{C}$ in the model SST develops along the American coast north of 40°N . The latter problem is partly linked to a systematic error in the Gulf Stream pathway in the model.

The altimetric data sets considered for the assimilation consist of 10-day maps of SSH at $1/4^{\circ}$ resolution, obtained as the sum of a time-invariant dynamic topography, and SLA data from the AVISO project. The mean SSH added to SLA maps is derived from inverse modelling (at 1° resolution) of the Atlantic circulation (Le Grand, 1998), using the Reynaud climatology as a dynamical constraint. The gridded SLA at $1/4^{\circ}$ resolution are calculated by combining Topex/Poseidon and ERS altimeter data, with an interpolation between tracks according to the method described by Le Traon et al. (1998). The accuracy of the SLA products can be reduced to ~ 3 cm RMS in average because of the very efficient correction of orbit errors on along-track data. However, due to the lack of accurate tidal corrections, the SLA data will not be assimilated in several coastal

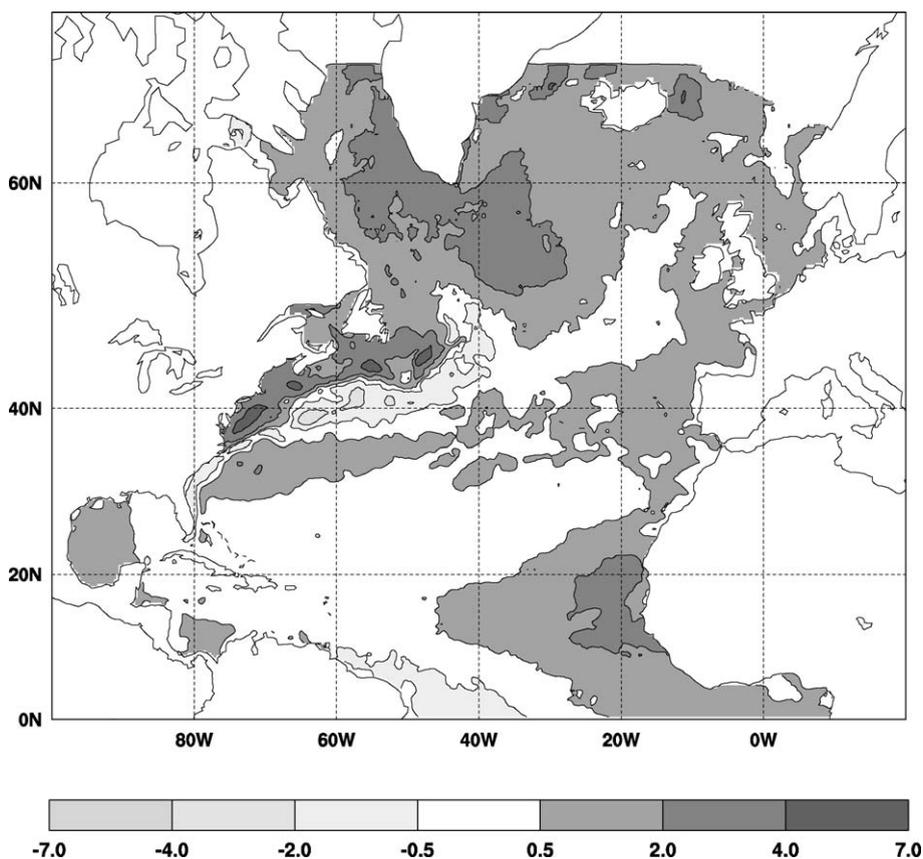


Fig. 2. Bias ($^{\circ}\text{C}$) between the simulated model SST and the SST observations during 1993. Dark (light) grey values indicate higher (lower) temperature in the model than in the observation.

zones like the Georges Bank or the Great Banks of Newfoundland. A bulk error of 5 cm RMS on the total SSH data has been prescribed in the assimilation system, taking into account the cumulated effect of measurements, inverse estimates, and mapping procedures.

Using a similar procedure as for SST, the bias between the SSH of the reference simulation and the data is diagnosed in order to evaluate systematic model errors in terms of surface topography (Fig. 3). The amplitude of the bias is maximum in the mid-latitude regions. A dipole of negative (north) and positive (south) sea-level misfit is present on each side of the Gulf Stream, reflecting the lack of kinetic energy in the jet by comparison with the mean currents of the inverse solution. The positive misfit centered at (50°N; 30°W) denotes the weakness of

the Gulf Stream extension eastward, and its consequences on the North Atlantic drift. In the Labrador Sea, a negative pattern, exceeding -20 cm, confirms the excess of cyclonic circulation in the subpolar Gyre which was already pointed out in the mean Barotropic Stream Function (see Fig. 1). Finally, the positive anomalies extending zonally located at 35°N in the eastern part of the basin reflect the absence of Azores Current in the reference run.

As an example of synoptic data, Fig. 4 (top) represents the SSH field in the Gulf Stream region that will be used in the assimilation experiment on October 21, 1992. This picture illustrates the crucial role of the mean surface topography added to the observed SLA to derive the absolute SSH data, resulting in a well-defined frontal structure along the Gulf Stream. A comparison with the model simulation

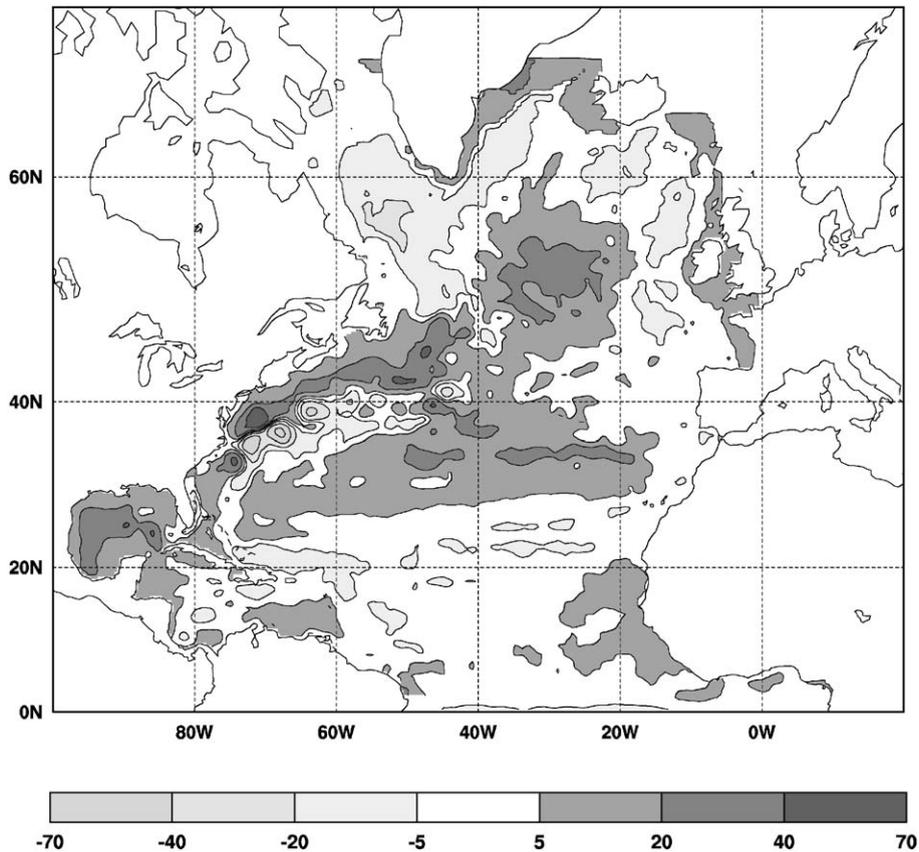


Fig. 3. Bias (cm) between the simulated model SSH and the SSH observations (constructed as a sum of a mean SSH and a sea level anomaly) during 1993. The spatial mean has been set to zero. Dark (light) grey values indicate higher (lower) sea level in the model than in the observation.

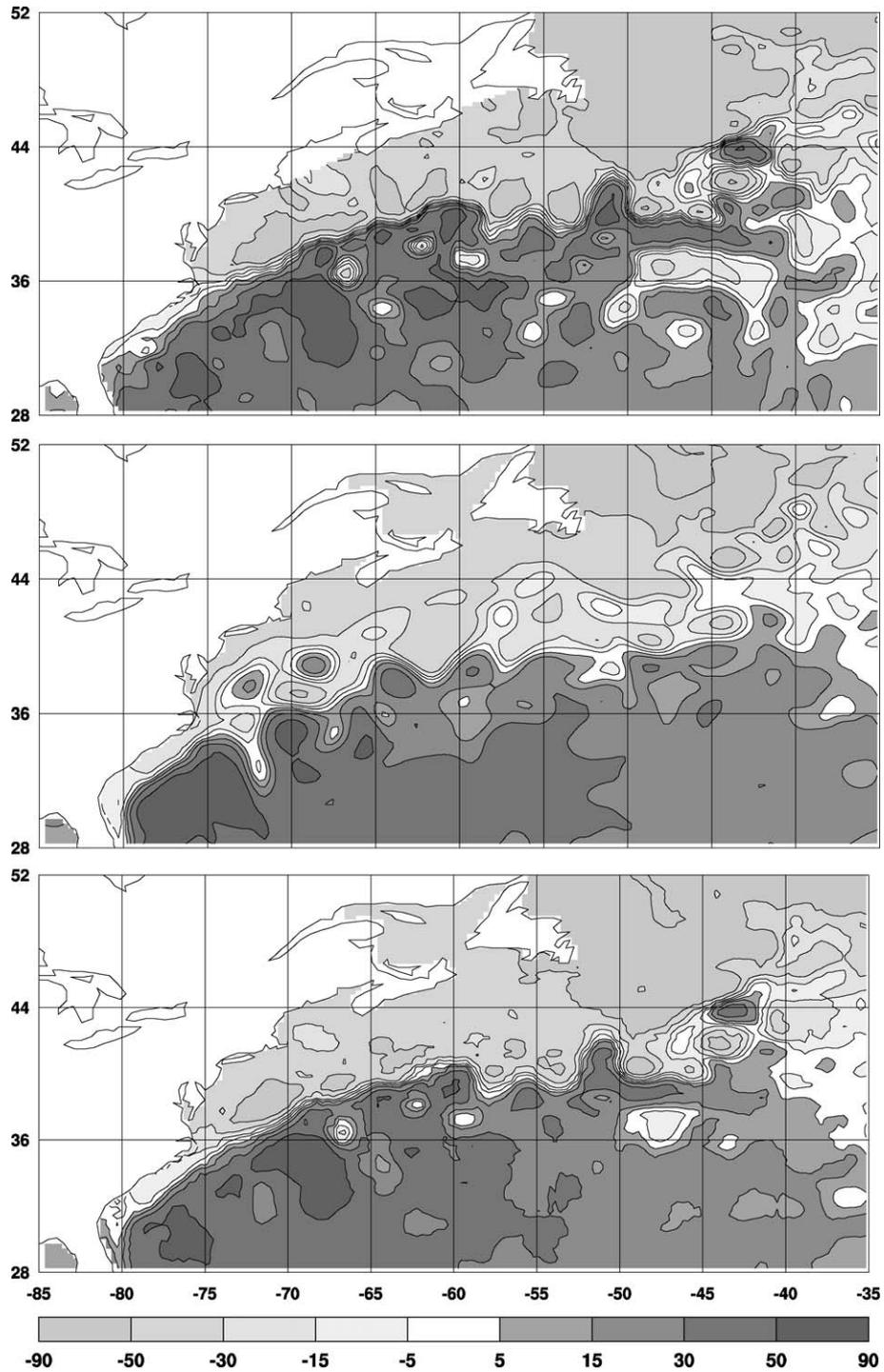


Fig. 4. Sea surface height (cm) for October 21, 1992 in the Gulf Stream region. Observation (top figure), free model simulation corresponding to initial first guess of the experiment (middle figure) and data assimilation analysis (bottom figure).

for the same day (Fig. 4, middle) reveals the lack of intensity in the surface current, and the wrong positioning of the meanders and eddies. Additional diagnostics (not shown here) indicate that, in general, the simulations suffer from a deficit in SST and SSH variability which are explained by the lack of meso-scale activity in the model.

The observation error that we specify in the assimilation experiments discussed hereafter is not only related to the accuracy of the gridded SST and SSH data, but also to the ability of the model to represent the observed signal. In order to evaluate this representativeness error, we have examined the RMS difference between the original data on a $1/4^\circ$ map and their projection onto the $1/3^\circ$ model grid. The standard deviation of the signal lost during the upscaling process varies between 0.1 and 0.6 $^\circ\text{C}$ for the SST and between 0.5 and 5 cm for the SSH. During the assimilation experiment, this signal will be considered as an additional component of the observation error, added to the 0.5 $^\circ\text{C}$ error on the gridded SST products, or to the 5 cm error on the gridded SSH products.

4. The assimilation method

The assimilation scheme implemented in these experiments is sequential, implying that only observations from the past can influence the current estimate of the oceanic state. The backbone of the assimilation algorithm is derived from the Kalman filter theory (Gelb, 1974). The model trajectory is corrected intermittently through a sequence of assimilation cycles, taking into account the confidence in the model prediction and the accuracy of the observed quantities. The assimilation cycles must be long enough to accumulate a sufficient amount of observations and correct the model forecast accordingly. A 10-day cycle has been prescribed in the present experiments, which is the time period needed to achieve a global coverage of the SSH and SST from satellites and generate the gridded products described in Section 3.

4.1. The state vector and the estimation vector

The sequential correction due to the Kalman filter analysis provides a new estimate of the state vector

which contains all the variables needed to restart the model. As the observed data only relate to a few variables of the model state, the assimilation scheme has to be multivariate, i.e., the whole state vector must be modified in a consistent manner in addition to the observed quantities themselves. Given the nature of the data available for assimilation in this study, a critical issue will be to correct the subsurface ocean properties from SST and SSH data only.

In the context of the statistical estimation theory, this can be achieved as long as the multivariate error statistics prescribed for the forecast state and the observations are reliable and robust. In practical assimilation problems, it is impossible to perfectly specify the error covariance between all state variables. The optimality of the Kalman filter is therefore restricted to the subset of the state vector with reliable statistical properties, which will be defined as the estimation vector. In order to update the rest of the state vector, simple physical constraints such as the geostrophic balance can be used to complete the statistical correction. This adjustment should minimize the occurrence of spurious perturbations associated to unbalanced increments during the next model initialization.

Formally, one needs to specify a partition of the model state vector ψ in order to distinguish the estimation vector x from the rest χ , so that:

$$\psi = [x|\chi]. \quad (1)$$

Strictly speaking, the state vector of the OPA model described in Section 2 is made of the two-dimensional barotropic stream function (BSF), and the three-dimensional arrays of the temperature (T), salinity (S), turbulent kinetic energy (TKE) and horizontal velocity components (U , V). In order to simplify the expression of the observation operator needed to compute the misfit between the model state and the data, the two-dimensional SSH fields is included as a part of the state vector in spite of the diagnostic nature of this variable in the rigid-lid approximation.

In order to define the estimation space, we proceed in two steps: first, we eliminate all state variables deeper than 1500 m because of the difficulty encountered in preliminary experiments to specify reliable and robust error covariances between these variables and the surface. Second, for the same reason, we exclude the velocity, the associated barotropic stream

function, and the turbulent kinetic energy from the estimation vector: the chaotic nature of the turbulent kinetic energy makes it difficult to calculate statistically significant covariances between TKE and the other state variables. The estimation vector is thus restricted to the sea surface elevation, the temperature and the salinity in the upper 1500 m.

Consistently with this partition, each assimilation cycle includes an analysis step to correct the estimation vector statistically using all pertinent observations of the system, an adjustment step to re-initialize the model state vector in a dynamically consistent manner, and a forecast operation to predict the ocean trajectory up to the next analysis time. These three steps are described hereafter.

4.2. The analysis step

The statistical method used to update the estimation vector using the satellite observations is derived from the Singular Evolutive Extended Kalman (SEEK) filter, which is a reduced-order assimilation scheme described in several earlier publications (e.g., Pham et al., 1998; Verron et al., 1999; Brasseur et al., 1999).

In spite of a reduced dimension of the estimation space compared to the model state space, a full Kalman filter still requires computational resources exceeding presently available computers, and a further reduction of the size of the estimation problem is needed. The SEEK filter has been developed to this aim, based on a reduced rank representation of the error covariance matrix associated to the estimation state.

4.2.1. The reduced order Kalman gain

The reduced-rank approximation of the background error covariance matrix at time t_i can be written as:

$$\mathbf{P}_i^f = \mathbf{S}_i^f \mathbf{S}_i^{fT} \quad (2)$$

where \mathbf{S}_i^f is the reduction operator (of dimension $n \times r$) related to the r modes $\{\mathbf{S}_{i,k}^f\}$ defining the error subspace. Using conventional notations (Ide et al., 1997), the analysis step of the SEEK filter takes the form

$$x_i^a = x_i^f + \mathbf{K}_i (y_i - \mathbf{H}x_i^f) \quad (3)$$

in which x_i^f is the forecast estimation vector (of dimension n) obtained by model integration up to time t_i ; x_i^a is the estimation vector after the analysis step, and y_i is the vector of observed quantities of dimension p (in our case study, gridded maps of SST and SSH). The gain matrix \mathbf{K} can be expressed in terms of the simplification operator \mathbf{S} and the observation operator \mathbf{H} (dimension $p \times n$) relating the observations to the prediction:

$$\mathbf{K}_i = \mathbf{S}_i^f [\mathbf{I} + (\mathbf{H}\mathbf{S}_i^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}_i^f]^{-1} (\mathbf{H}\mathbf{S}_i^f)^T \mathbf{R}^{-1} \quad (4)$$

where \mathbf{R} is the observation error covariance matrix. Eq. (4) shows that the size of the inversion problem is determined by the error subspace dimension (as long as the observation error is parameterized with a diagonal matrix), while the original Kalman gain requires an inversion in the observation space. As the number of observations is usually much larger than the rank of the error subspace used in practice, the inversion step of the SEEK algorithm is getting much cheaper than the corresponding computation of the original Kalman gain.

Finally, the scheme evaluates the error covariance of the analysis and updates the error subspace accordingly:

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i \mathbf{H}) \mathbf{P}_i^f = \mathbf{S}_i^a \mathbf{S}_i^{aT} \quad (5)$$

$$\mathbf{S}_i^a = \mathbf{S}_i^f [\mathbf{I} + (\mathbf{H}\mathbf{S}_i^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}_i^f]^{-1/2}. \quad (6)$$

4.2.2. The local approximation

The estimation of small correlations associated with remote observations is a well-known difficulty of reduced-order Kalman and ensemble methods (Houtekamer and Mitchell, 1998). A further simplification of the SEEK analysis scheme is thus introduced (Testut, 2000; Brankart et al., 2003), enforcing to zero the error covariances between distant variables. This algorithmic improvement will prevent the spurious influence of, say, equatorial data at high latitudes through large-scale signatures in the EOFs.

In practice, this is implemented by assuming that distant observations have negligible influence on the analysis. The global system is split into subsystems for which a traditional analysis is computed: only data points located within a specified region around the subsystem will contribute to the Kalman gain.

In practice, each subsystem includes 2×2 horizontal grid points ($\times 43$ vertical levels), and the associated regions of influence extend over 14×14 grid points, setting an upper bound to the correlation scales of about 200 km. Note that we did not observe any discontinuity of the gain between adjacent subsystems because the size of the subdomains was taken large enough to overlap each other in a way that two neighboring subdomains use almost the same local data set.

The local gain is an approximation, but it makes sense since only data points located in the “neighborhood” of a model grid point should have a significant impact on the analysis for that grid point. Besides, the regular distribution of the gridded observations (at $1/4^\circ$) on the model domain always provides at least a few data points within each region of influence (if there were no data available inside the region of influence, no correction would be applied). Further, we have observed that this also improves the analysis because the dimension of the error subspace, relative to the number of state variables in a particular subsystem, increases and therefore spans a larger part of the estimation space.

4.2.3. *The error subspace initialization*

In most earlier implementations of SEEK, the dominant EOFs of the system’s variability were used to initialize the background error covariance (e.g., Verron et al., 1999). The initialization of the error subspace could also be expressed in terms of singular vectors, breeding vectors or differences between slightly different model forecasts. A procedure along these lines was actually tested with some success by considering the divergence between two model forecasts obtained with and without SST relaxation (Testut, 2000).

In the present work, however, the computation of EOFs was found easy and robust enough to estimate relevant error statistics at initial time. We thus performed an EOF analysis of multivariate model dumps sampled every 10 days from a prior simulation of the 1990–1993 period. In each subsystem defined above, the local leading modes of the spectrum were retained out of a series of 154 realizations, preserving in this way the local dominant features of SSH and SST covariance simulated by the model (Testut, 2000; Penduff et al., 2003). In practice, local EOFs are

computed on each subdomain individually with only the subset of multivariate model variables available on the subdomain grid. By doing so, we observed that the model variability can be represented with only a few local EOFs (we used four local modes here), while a much larger number of EOFs covering the basin scale would be required to achieve an equivalent level of explained variance. The effective rank of the error covariance matrix is then considerably increased. The error subspace spanned by these modes specifies how to spread the information from the observed quantities to the whole estimation vector (e.g., it specifies how SST and SSH data are correlated to the thermohaline properties in the upper 1500 m of the water column).

4.3. *The adjustment step*

The analysis step described above updates the estimation vector only. Before starting a new forecast, the remaining part of the state vector χ should in principle be adjusted to the statistical update defined in the estimation space. This includes the temperature and salinity variables under 1500 m, the turbulent kinetic energy, the horizontal velocity components and the barotropic stream function.

Without the explicit use of in situ observations of the deep ocean properties, one can hardly expect an effective correction of the T/S field below 1500 m. In practice, our assimilation scheme will preserve the evolution of the deep water mass properties as they are computed by the model itself. A moderate smoothing of the correction is simply performed to avoid an abrupt transition between the corrected variables above 1500 m depth, and those uncorrected below.

Concerning the turbulent kinetic energy (TKE), we re-diagnose its distribution in space from the other analysed state variables, in order to restore an approximate balance in the production/destruction terms of the closure scheme (Blanke and Delecluse, 1993). Conversely, there is no explicit update of the velocity field from the previous forecast, but a dynamical adjustment to the thermohaline update is achieved by the model within a few time steps (about 1 day after restart). New solutions are currently examined to improve the dynamical balance of the state vector before restart by including the velocity components in the estimation vector or by enforcing geostrophic equilibrium.

A few additional details of the restart procedure concern the following items:

- the time-integration in OPA is based on a leap-frog scheme, requiring two successive time-steps to restart a model run after the analysis step; as only one analysed state can be calculated every assimilation cycle, a second “restart” state is generated by performing a Euler time step first before the next model forecast;
- due to the nonlinearities of the model physics and to the existence of a hydrostatic constraint in the model formulation, the hydrostatic balance of the analysed state is not guaranteed; a check of the water column stability is thus performed after the analysis, and an adjustment based on enhanced vertical diffusion coefficients takes place if needed;
- in buffer zones, the correction has been switched off to avoid conflicting interplay between the assimilation updates and the Newtonian relaxation terms.

4.4. The forecast step

The analysis and adjustment steps at time t_i supply a new state vector

$$\psi_i^a = [x_i^a | \lambda_i^a] \quad (7)$$

which is used as initial conditions for a new model forecast up to time t_{i+1} :

$$\psi_{i+1}^f = \mathbf{M}[\psi_i^a]. \quad (8)$$

In addition to the model state, a property of sequential filters belonging to the Kalman family is the propagation of the error covariance from one analysis step to the next using the model dynamics and a statistical description of the model error. This error propagation is needed to perpetuate the optimality of the estimation process with time. Introducing the low rank approximation into the error covariance equation and considering the linear tangent model \mathbf{M}' , one gets:

$$\mathbf{P}_{i+1}^f = (\mathbf{M}'\mathbf{S}_i^a)(\mathbf{M}'\mathbf{S}_i^a)^T + \mathbf{Q}_i. \quad (9)$$

Two major hurdles arise to explicitly resolve this equation in the assimilation algorithm. First, the

evolution of the error covariance with the model equations remains computationally expensive even with the reduced-rank approximation, requiring r model integrations in addition to the central forecast (Eq. 8). Second, one has to specify the explicit structure of the systematic error \mathbf{Q}_i ; however, a detailed knowledge of the nature of the model error is rarely available, and the impact of poorly specified statistics may be dramatic on the estimation process.

Instead of this formulation, we investigate in this paper a shortcut to Eq. (9), assuming that the forecast error covariance matrix can be written as:

$$\mathbf{P}_{i+1}^f = \mathbf{D}_i^{1/2} \mathbf{P}_i^a \mathbf{D}_i^{1/2} \quad (10)$$

where $\mathbf{D}_i^{1/2}$ is a diagonal amplification matrix. The role of this simple parameterization is to simulate the cumulated effects of the model error and the dynamical growth of the pre-existing error modes, with the interesting property to preserve the rank of the error covariance matrix. Despite its simplicity, such idealization does not imply that the model is assumed perfect.

In the general case, the diagonal elements of \mathbf{D}_i should be tuned individually to achieve the same error

Table 1

Statistical diagnostics averaged over 1993 for the adaptive assimilation: RMS misfit between estimated field and observation, error standard deviation as produced by the SEEK filter

	SSH (cm)	SST (°C)
<i>NATL3 Basin</i>		
Free run/observation misfit	15.1	1.41
Forecast/observation misfit	9.8	1.14
Analysis/observation misfit	4.5	0.74
Forecast error	6.6	1.09
Analysis error	3.0	0.59
Observation error	5.0	0.51
<i>Gulf Stream Region</i>		
Free run/observation misfit	24.7	1.92
Forecast/observation misfit	17.1	1.40
Analysis/observation misfit	6.6	0.79
Forecast error	11.1	1.68
Analysis error	3.7	0.58
Observation error	5.2	0.52

variance propagation as what could be obtained from Eq. (9). In this first attempt, we have adopted a further simplification by assuming for this matrix a block-diagonal structure which corresponds to the partitioning of the system into the subsystems introduced for the analysis step. A unique amplification factor is thus associated to each subsystem, the value of which is prescribed a priori or evaluated using the adaptive scheme described hereafter.

In Brasseur et al. (1999), it was shown that an adaptive mechanism can efficiently update the error subspace of the SEEK filter using the information left in the innovation vector after each analysis step (i.e., the residual innovation vector). A similar idea, also used by Dee (1995), has been implemented in the present study to determine the amplification matrix \mathbf{D}_i using the same source of information.

The baseline of the adaptive algorithm is to enforce the consistency between the error variances predicted by the filter and the ‘observed’ variance captured by the innovation vector. To achieve this goal, we consider the classical Eq. (11) from the optimal linear estimation

$$\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T = \mathbf{E}[(\mathbf{y} - \mathbf{H}\mathbf{x}^f)(\mathbf{y} - \mathbf{H}\mathbf{x}^f)^T] \quad (11)$$

in which we introduce Eq. (10). \mathbf{E} denotes the mathematical expectation operator. Taking the diagonal part and re-ordering, we obtain

$$\text{diag}(\mathbf{H}\mathbf{D}_i^{1/2}\mathbf{P}_i^a\mathbf{D}_i^{1/2}\mathbf{H}^T) = \mathbf{D}_{v_i} - \text{diag}(\mathbf{R}) \quad (12)$$

where \mathbf{D}_{v_i} is a diagonal matrix containing the innovation variance v_i . \mathbf{D}_{v_i} is estimated from the sequence of the last innovation vectors weighted with an exponential decreasing towards the past. In practice, the esti-

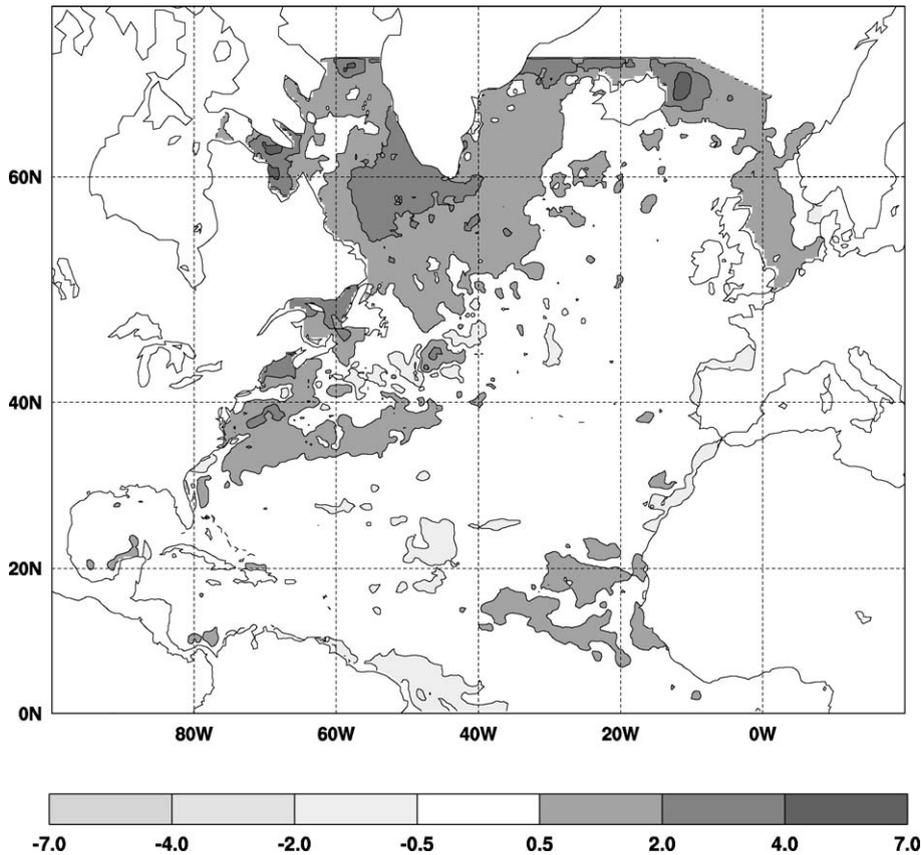


Fig. 5. Bias ($^{\circ}\text{C}$) between the model SST from the assimilation experiment (10-day forecasts) and the SST observations during 1993. Dark (light) grey values indicate higher (lower) temperature in the model than in the observation.

mation of v_i is done sequentially at each step by evaluating:

$$v_i = (1 - \theta)v_{i-1} + \theta v_* \quad (13)$$

where v_* is the square of the current innovation (Testut, 2000; Brankart et al., 2003). A value of 0.2 is prescribed for θ , which corresponds to an e-folding time of 50 days for 10-day assimilation cycles. Eq. (12) expressed that the background error variance should in average be equal to the innovation variance minus the observation error variance. Then, a procedure is set up, which determines adaptively the \mathbf{D}_i diagonal matrix (i.e., an amplification factor on each subdomain) in order to approximately verify the balance expressed by this equation. Of course, in Eq. (12), we have more equations than unknowns. So we need to use a criterion to find the “best” solution. Assuming that \mathbf{D}_H^f is the left

hand side of Eq. (12) and \mathbf{D}_{est} is the right hand side, we adjust the amplification factor to minimize:

$$\text{tr}(\mathbf{D}_H^{f-1} \mathbf{D}_{\text{est}} + \mathbf{D}_H^f \mathbf{D}_{\text{est}}^{-1}) \quad (14)$$

on each subsystem. This formulation is minimum if \mathbf{D}_H^f is equal to \mathbf{D}_{est} , and increase if strong ratio exist between them.

Despite its simplicity, this method is useful to control the evolution of the error variance during the assimilation sequence and preserve the statistical consistence of the scheme.

5. The assimilation experiments

In this section, we discuss the results of assimilation experiments using the data sets described in

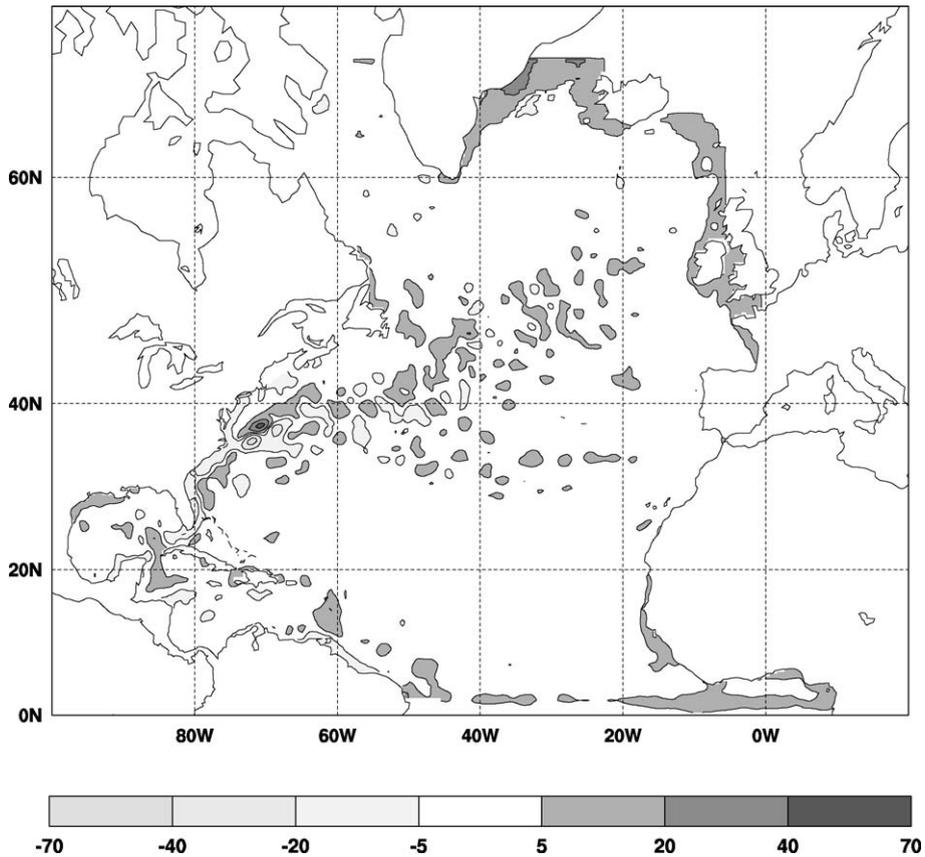


Fig. 6. Bias (cm) between the SSH from the assimilation experiment (10-day forecasts) and the SSH observations during 1993. The spatial mean has been set to zero. Dark (light) grey values indicate higher (lower) sea level in the model than in the observation.

Section 3 during the period from October 1992 to December 1993. The experiments have been implemented with the help of the SESAM software (Testut et al., 2001), which is a modular package developed to manage the various stages of the assimilation chain in a flexible manner.

The assimilation system needs a few cycles to adjust the initial error statistics (Testut, 2000), and the first 3 months in 1992 must be regarded as the “spin-up” of the experiment, while the complete annual cycle of

1993 can be considered as the adequate time interval for computing the assimilation diagnostics. In addition to assessing the global solution, we will focus some of these diagnostics on the Gulf Stream region because of its critical role on the dynamics of the whole North Atlantic basin.

The objective assessment of the system is not trivial because the data with a sufficient space–time coverage are already used by the assimilation process, while only a few independent data are available for

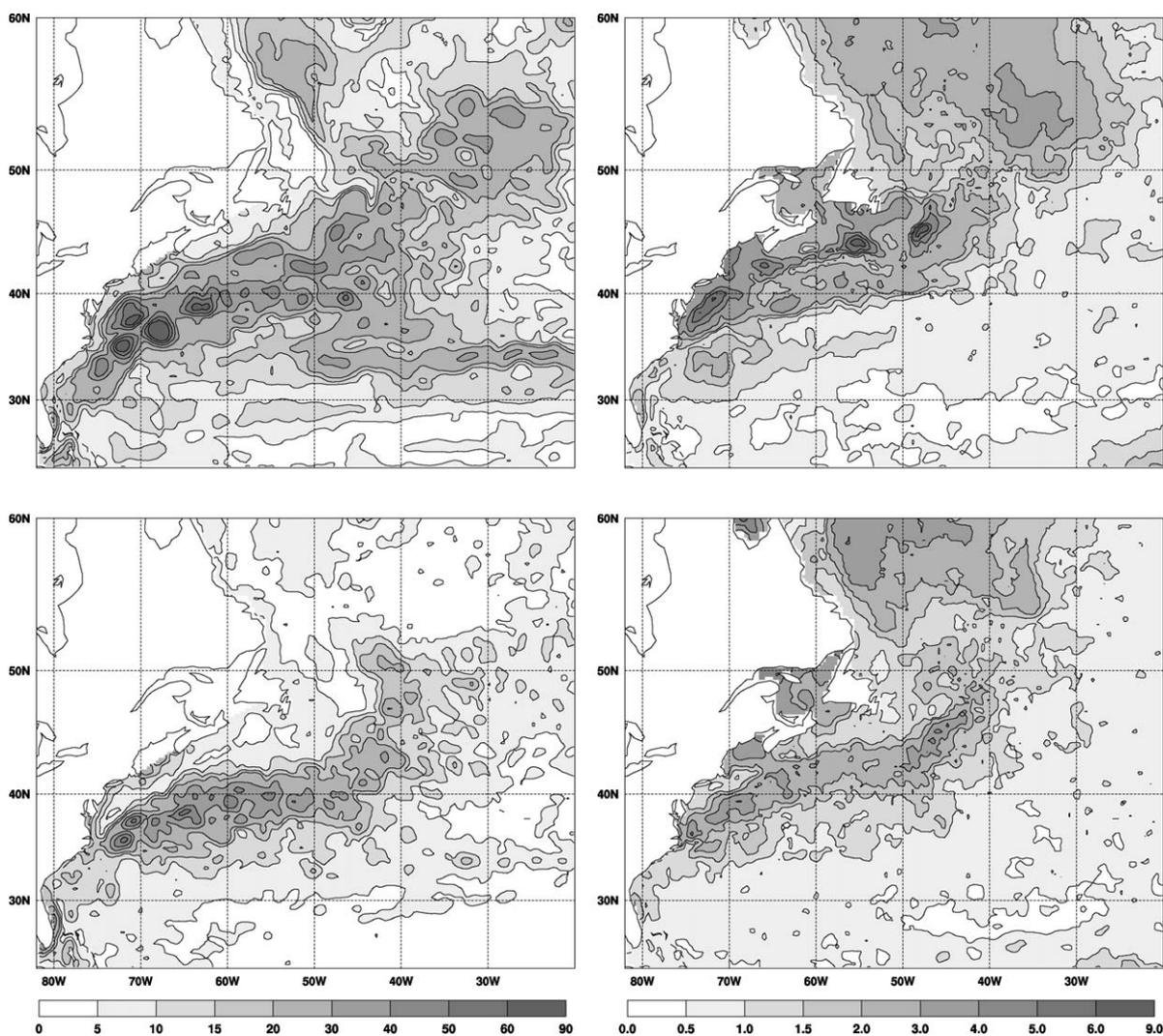


Fig. 7. RMS misfit to observation for sea surface height in cm (left column) and sea surface temperature in °C (right column) in the Gulf Stream region during 1993. Free model simulation (top figures) and 10-day forecast (bottom figures).

verification. The methodology to evaluate the assimilation experiments will therefore rely on three different metrics:

- (i) the computation of RMS misfits between the SST and SSH data fields and their equivalent estimates from the assimilation sequence (in spite of the fact that these statistics cannot be considered as an objective measure of the system's performance);
- (ii) the assessment of unobserved quantities (such as large-scale currents) by comparison with our prior knowledge of the ocean circulation;

(iii) and, the validation with fully independent data sets (e.g., in situ hydrographic profiles).

5.1. Comparison to the satellite data

As a first illustration, the SSH analysis on October 21, 1992 (Fig. 4, bottom) allows a detailed inspection of how one specific analysis modifies the surface topography in the Gulf Stream region. The frontal system associated to the Gulf Stream has been corrected in a positive way with respect to the first guess shown in Fig. 4 (middle). The mean position of the jet

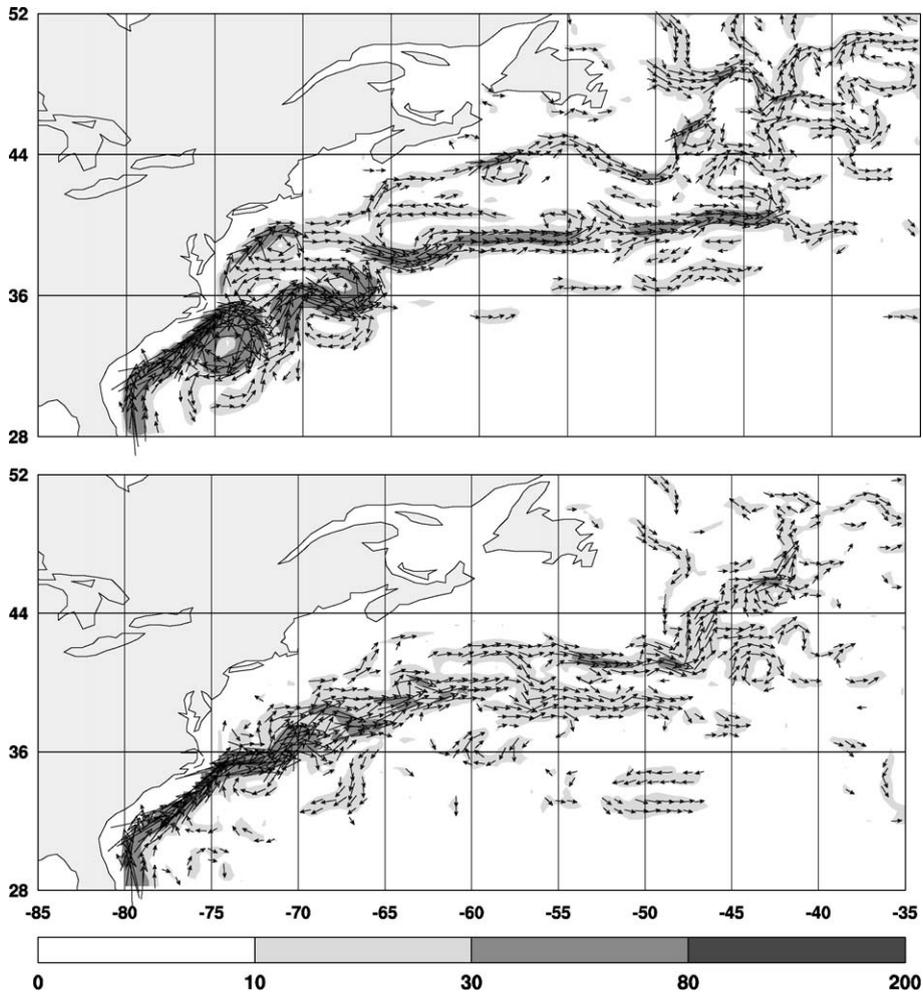


Fig. 8. Horizontal velocity (cm s^{-1}) in the Gulf Stream region (at 50 m depth). Free run (top) and 10-day forecast (bottom) are averaged over 1993.

is more consistent with the SSH map in Fig. 4 (top) after assimilation, and the magnitude of its meridional slope is more realistic. The observed mesoscale activity identified by the presence of eddies along the jet is also better represented and located in the analysis.

An evaluation of the averaged behaviour of the assimilation system during 1993 is given by Table 1, which represent the RMS misfits between the satellite data and the model estimates (in cm for SSH and °C for SST) in the free run, the 10-day

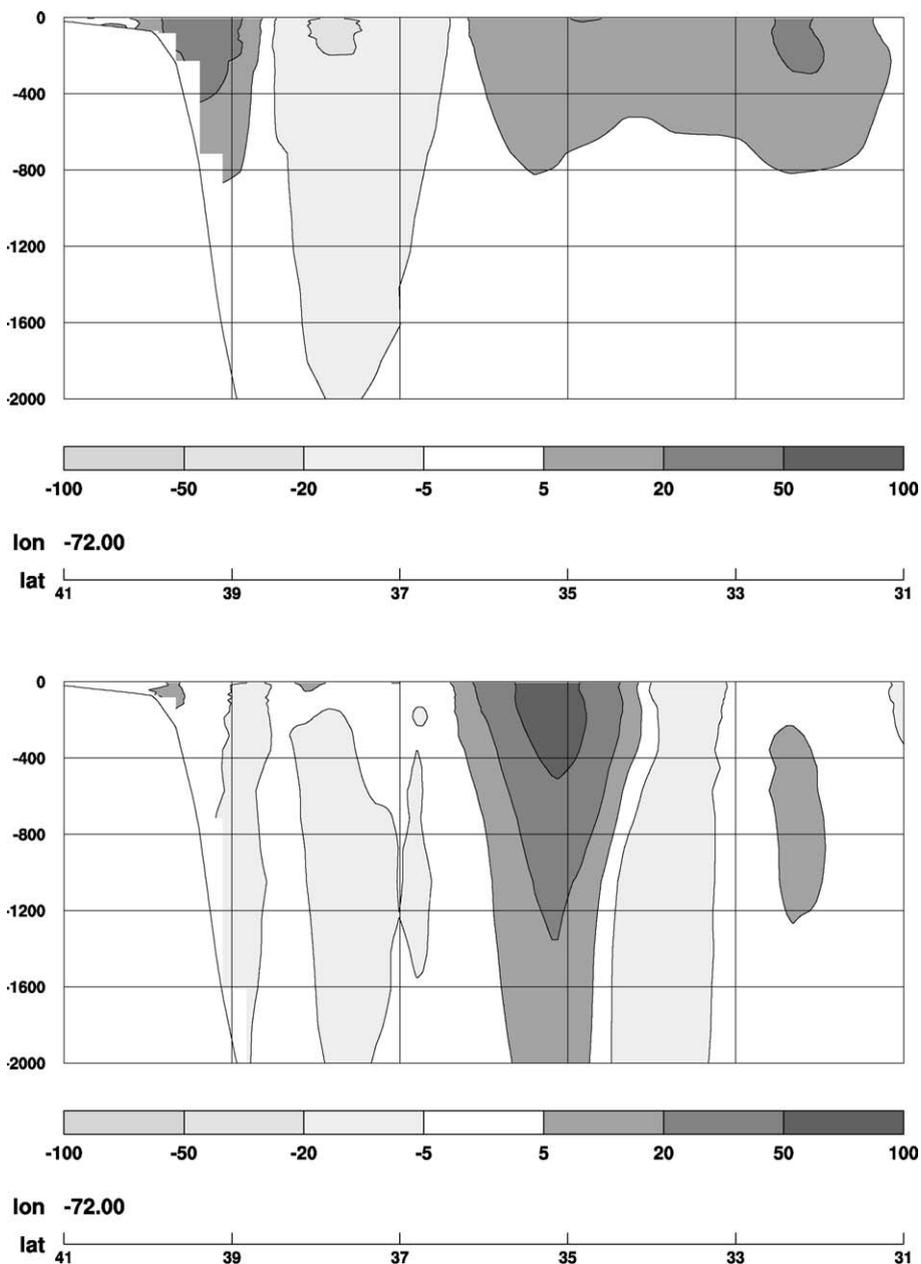


Fig. 9. Meridional section in the Gulf Stream region (at 72°W) for the velocity (cm s^{-1}) until 2000 m depth. Free run (top) and 10-day forecast (bottom) are averaged over 1993.

forecasts and the analysed states of the assimilation sequence.

In the North Atlantic basin, a systematic reduction of the RMS misfit is observed on the SSH, dropping from 15 cm in the free run to 4.5 cm in the analyses, i.e., slightly lower than the standard deviation of the observation error (5 cm). Regarding SST, the RMS misfit drops from 1.41 °C in the free run to 0.74 °C in the analysis. Concerning the 10-day forecasts, the RMS misfits with the data are larger (9.8 cm and 1.14 °C, respectively, for SSH and SST), but they remain well below the corresponding figures of the free run. The model is thus able to properly “ingest” the observed information at the analysis time, and to propagate this information dynamically up to the next assimilation cycle.

In addition to the model-data misfits, Table 1 provides the average levels of forecast and analysis errors predicted by the filter in the assimilation experiment. Those errors are systematically smaller than the corresponding misfits to the data, suggesting that some noise in the observations has been filtered out by the scheme. Similar statistics have also been

computed in the Gulf Stream area, showing the same tendencies but with generally higher figures as a result of a more intense oceanic variability in that region.

In optimal linear estimation, the error statistics should verify (Dee, 1995):

$$\mathbf{R} - \mathbf{H}\mathbf{P}^a\mathbf{H}^T = \mathbf{E}[(\mathbf{y} - \mathbf{H}\mathbf{x}^a)(\mathbf{y} - \mathbf{H}\mathbf{x}^a)^T] \quad (15)$$

indicating that the analysis/observation misfits should be smaller than the observation errors. Table 1 shows that these conditions are not satisfied everywhere for SST and for SSH. The lack of consistency between the estimated errors and the innovations can be explained by inadequate values of the observation error variances (including representativeness errors) which, in the Gulf Stream region, are probably underestimated. The estimation of the error statistics could therefore still be improved, for instance by means of a more complex adaptive algorithm.

In order to identify the nature of the signal which has effectively been modified by the assimilation, we show in Figs. 5 and 6 the bias between the data and the 10-day forecasts of all assimilation cycles in

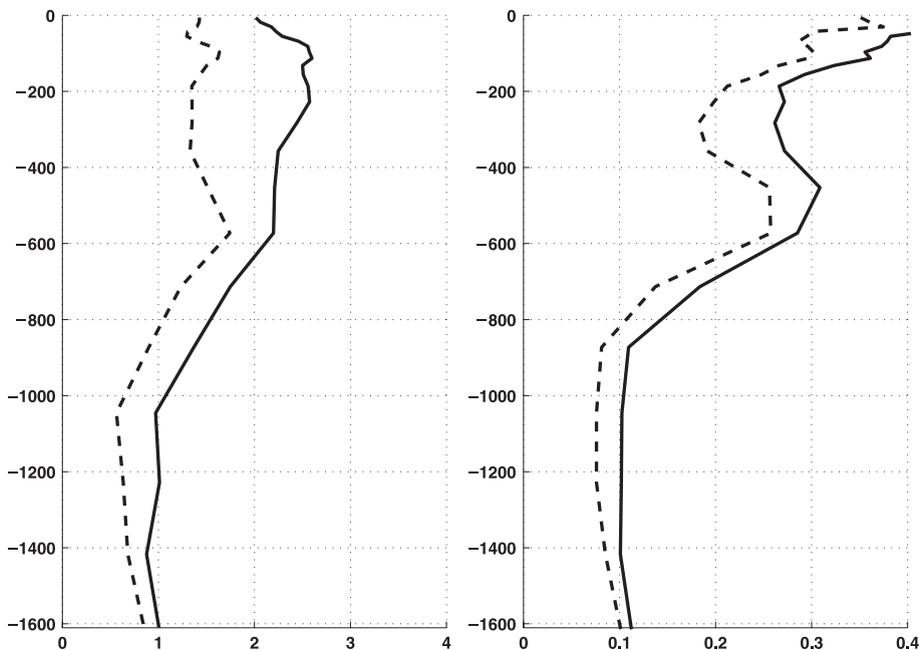


Fig. 10. RMS misfit to Reynaud climatology for temperature in °C (left column) and salinity in psu (right column) in the Gulf Stream region down to 1600 m depth averaged over 1993. The solid curve indicates the free run misfit while the dashed curve shows the 10-day forecast misfit.

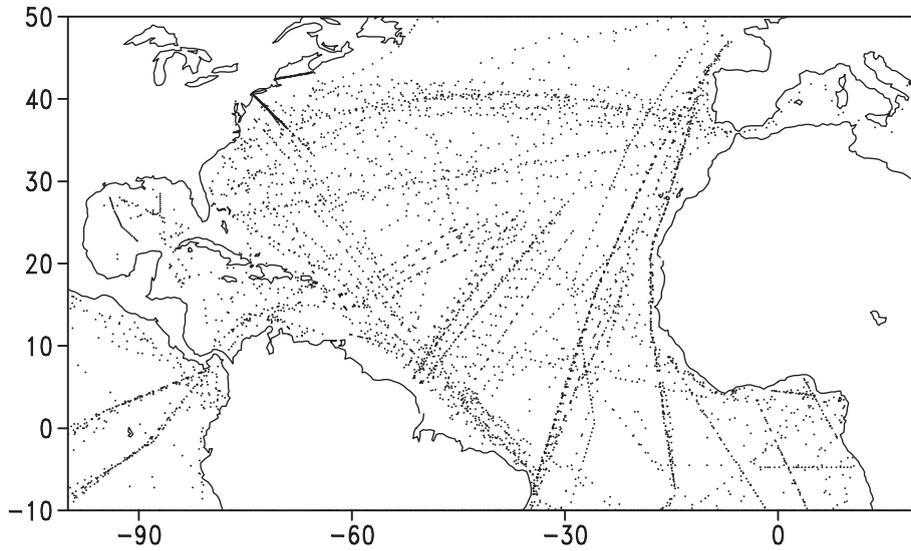


Fig. 11. Horizontal distributions of XBT profiles, available during 1993 (source SISMER).

1993. A significant reduction of the bias can be observed for both SSH and SST by comparing with the same quantity computed from the free run (Figs. 2 and 3).

In the Gulf Stream, particularly, the assimilation leads to an improved agreement with the data. Colder surface waters are now present north of 40°N, as a result of a better positioning of the jet, which separates

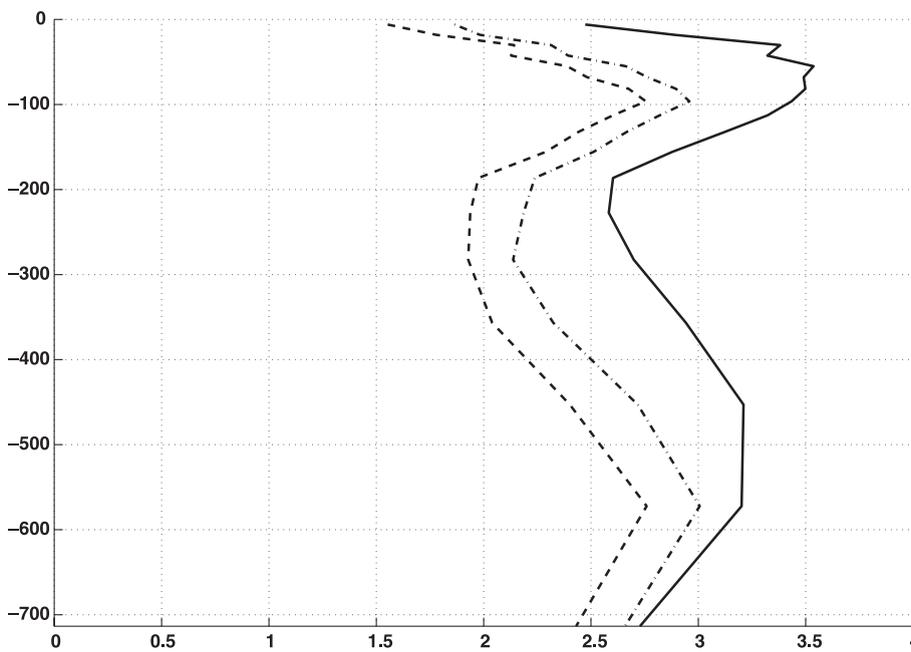


Fig. 12. Temperature RMS misfit (°C) with respect to XBT profiles in the Gulf Stream region down to 700 m depth averaged over 1993. The figure presents the free run (solid curve), the 10-day forecast (dashed and dotted curve) and the analysis state (dashed curve).

cold waters of subpolar origin and warm waters from the subtropical gyre. The new slope across the jet is apparently strengthened, and thus more consistent with what is expected in reality. The extension of the Gulf Stream has been modified too, and we observe a reduction of the bias in the Eastern North Atlantic accordingly. In the central North Atlantic, a significant bias was present at 35°N in Fig. 3 because of a too weak Azores Current in the free run. This feature has completely been removed by the assimilation (Fig. 6), and this can be partly attributed to the presence of this feature in the mean surface topography added to the SLA data.

In the low latitude region, the 10-day forecast bias is generally small, and the signature of a too weak upwelling along the African coast is less visible. The high latitudes are characterised by some improvement in the SSH representation. However, the signature of an important SST bias remains in the Labrador Sea,

and more generally in the subpolar gyre. This can be explained by the error in the forcing fields which dominates in that region, and significantly affects the forecast, which is performed without SST relaxation in the assimilation runs.

The assimilation of satellite data is also useful to improve the variability in the regions where the horizontal grid resolution is not sufficient to properly simulate the mesoscale turbulence. This is illustrated by Fig. 7, showing the distribution of RMS difference between the model runs (reference and assimilation) and the data. As expected, the maximum of the misfit amplitude is associated to the Gulf Stream path and its extension toward the Eastern North Atlantic. However, the misfits are significantly reduced in the assimilation run, suggesting a better consistency with the variability of the surface properties themselves. One can also notice that the well-marked SSH signature in the reference run characterising the absence

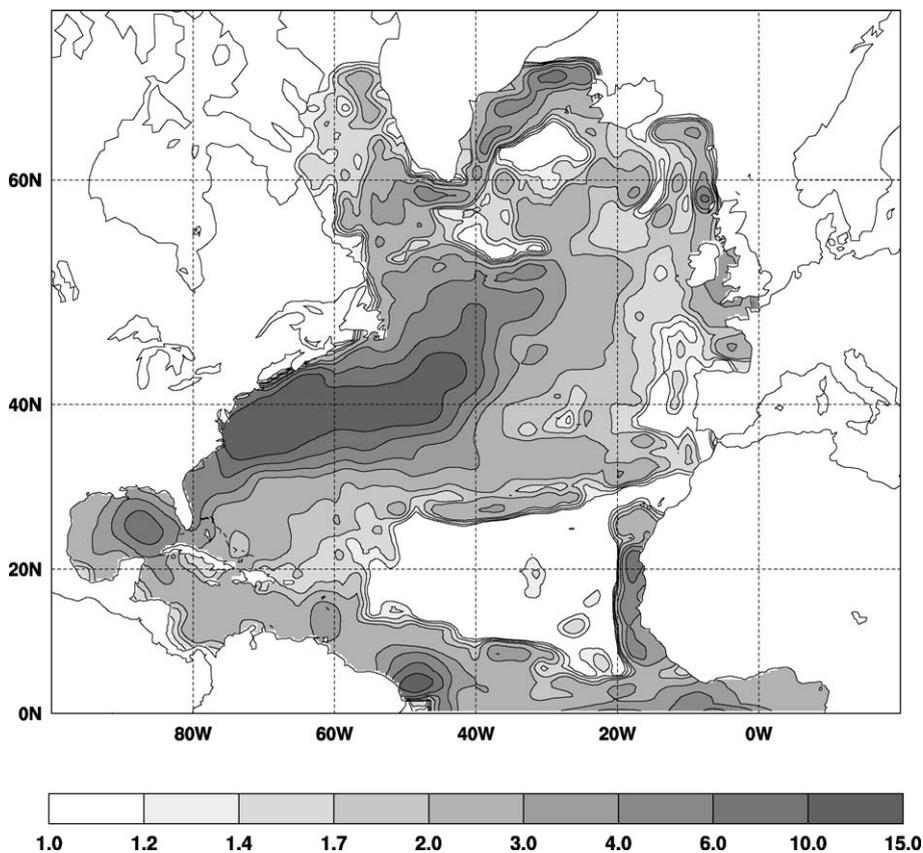


Fig. 13. Distribution of the amplification factor μ in the adaptive assimilation experiment averaged over 1993.

of zonal extension of the Azores current at 35°N , has been smoothed out completely in the assimilation.

These statistical results suggest that the assimilation system has the capability to correct the major model failures diagnosed from the observation of the surface variables. In addition, these corrections are sufficiently robust to persist during a 10-day forecast and impact the following analysis.

5.2. The gulf stream circulation

Another critical issue is to verify the correct extrapolation of the assimilated information onto unobserved variables, such as large-scale surface currents. The velocity field from the 10-day forecasts is an interesting quantity to diagnose if the correction of the state vector is sufficiently robust to permit the adjustment of the dynamics. Note that the new current structure does not directly result from the analysis

itself because in this experiment the velocity variables (U, V) are not included in the estimation space.

A focus on the Gulf Stream region (Fig. 8) illustrates the positive impact of SST and SSH data on horizontal currents at 50 m depth and their associated transports. The assimilation is able to improve the surface velocity and modify its direction efficiently. By comparison with the free simulation, the Northern stream along the American coast and the permanent eddy near Cape Hatteras have been removed in the assimilation run. One can also notice a better organized North Atlantic Current at $45^{\circ}\text{N}, 45^{\circ}\text{W}$. In spite of these improvements, however, the westward extension of the Gulf Stream is still too “viscous” with respect to what is expected in reality.

A vertical section through the Gulf Stream at 72°W (Fig. 9) demonstrates that the assimilation of surface data consistently modifies the three-dimensional structure of the flow. The zonal velocity has been

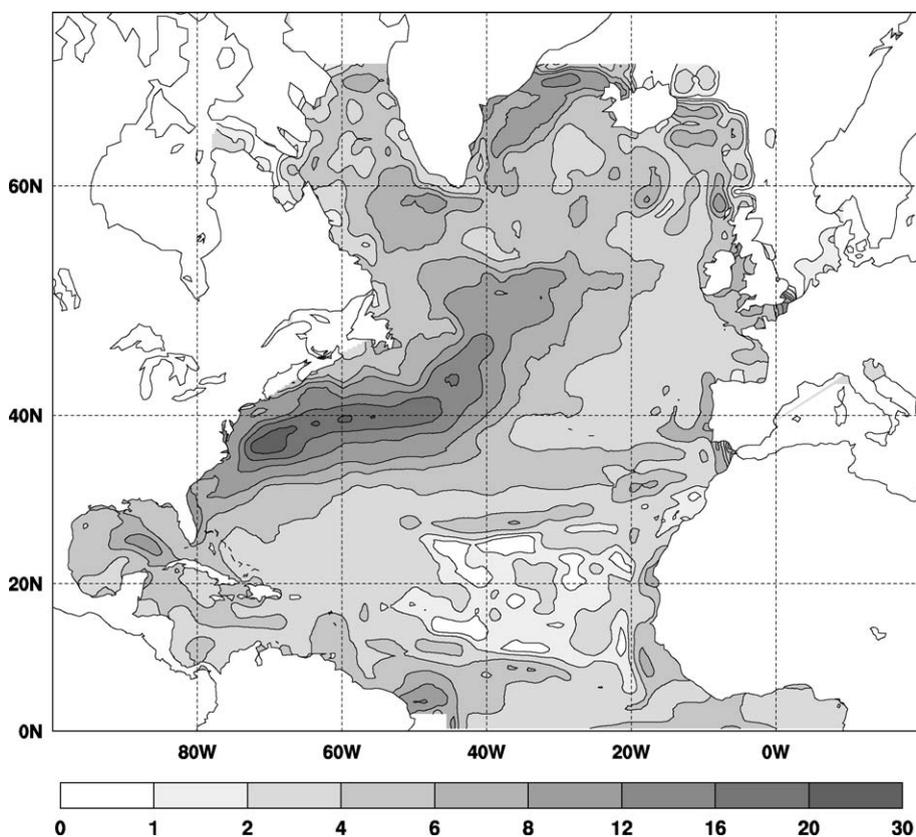


Fig. 14. Distribution of the 10-day forecast error in the adaptive assimilation experiment for the sea surface height (cm) averaged over 1993.

intensified between the surface and 2000 m depth, showing a well identified jet located at a more realistic latitude. The maximum surface flow of the Gulf Stream occurs now at $35\text{--}36^\circ\text{N}$ and, at more than 80 cm s^{-1} , in agreement with expected values.

5.3. Validation with hydrographic data

The only way to objectively validate the impact of assimilation is to compare with independent data, i.e., data which have not been used at any stage of the estimation process.

For this purpose, the model tracer fields (T , S) of the assimilation and the reference runs have been compared to the Reynaud climatology (Fig. 10). The climatological equivalent of the temperature and salinity fields have been computed by averaging the experiment in space and time, in order to compare

the same features as those described by the Reynaud climatology. The plots represent RMS misfits on the vertical between the surface and 1600 m depth in the Gulf Stream region. This comparison is useful to examine how the assimilation propagates the information from the surface to the ocean's interior and also from observed to unobserved variables of the estimation space (e.g., salinity).

By comparison with the free run, the climatology of the 10-day forecasts has been systematically improved with respect to both temperature and salinity distributions. This improvement is fairly consistent throughout the whole water column where the analysis correction is applied (i.e., the upper 1500 m depth). The mechanism responsible for the modification of the thermohaline properties is at first order related to the vertical structure of the multivariate error modes linking the many variables of the estimation space.

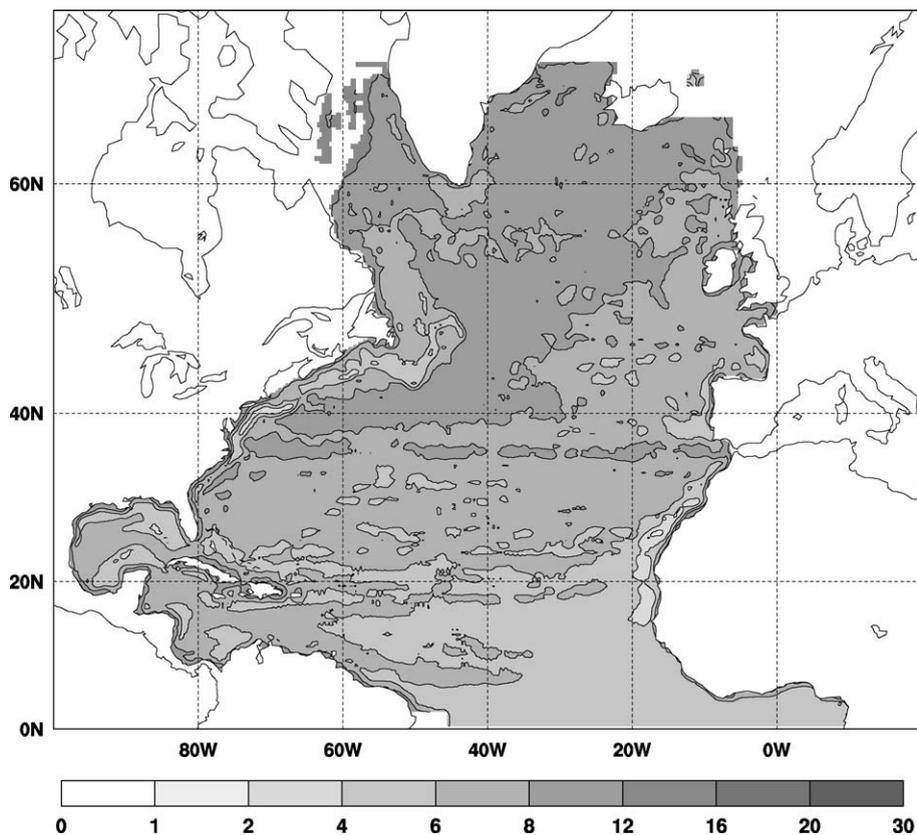


Fig. 15. Distribution of the forecast error for the sea surface height (cm) averaged over 1993 in the assimilation experiment using a constant amplification factor of $\mu = 5$.

A final assessment of the assimilation performances has been produced using an ensemble of 2500 XBT profiles collected in the North Atlantic during the period of the experiment (Fig. 11). These data have been extracted from a large historical data base gathered by the SISMER oceanographic data center (<http://www.ifremer.fr/sismer/>). We show in Fig. 12 the RMS misfits between the XBT temperature profiles and their model counterparts interpolated at the same times and locations during 1993.

Again, it is fortunate to notice the positive impact of the assimilation on the thermal field between the surface and 700 m depth. The analysis profile is better than the free run by almost 1 °C near the surface, and 0.7 °C down to 500 m depth. The profile calculated from the 10-day forecasts is slightly worse, remaining however systematically better than the free run. The reduction of the misfit results from a smaller bias and

a better representation of the variability in the assimilation.

The vertical structure of the misfits exhibit a local maximum at 100 m depth, which may be symptomatic of the difficulty to simulate the mixed layer depth correctly. This is a common feature of both the free run and the assimilation experiment (see also Brankart et al., 2003; Penduff et al., 2003). It is worth remembering here that the relaxation on SST is active on the free run and inactive in the assimilation experiment, indicating that the decrease of the misfits just below the surface is the consequence of different causes.

A minimum RMS misfit is observed between 200 and 300 m, i.e., at a depth not directly affected by the surface fluxes, but still under the influence of the mesoscale activity. The comparison with the XBT data demonstrates a positive impact of the altimetric data, which play the dominant role on the

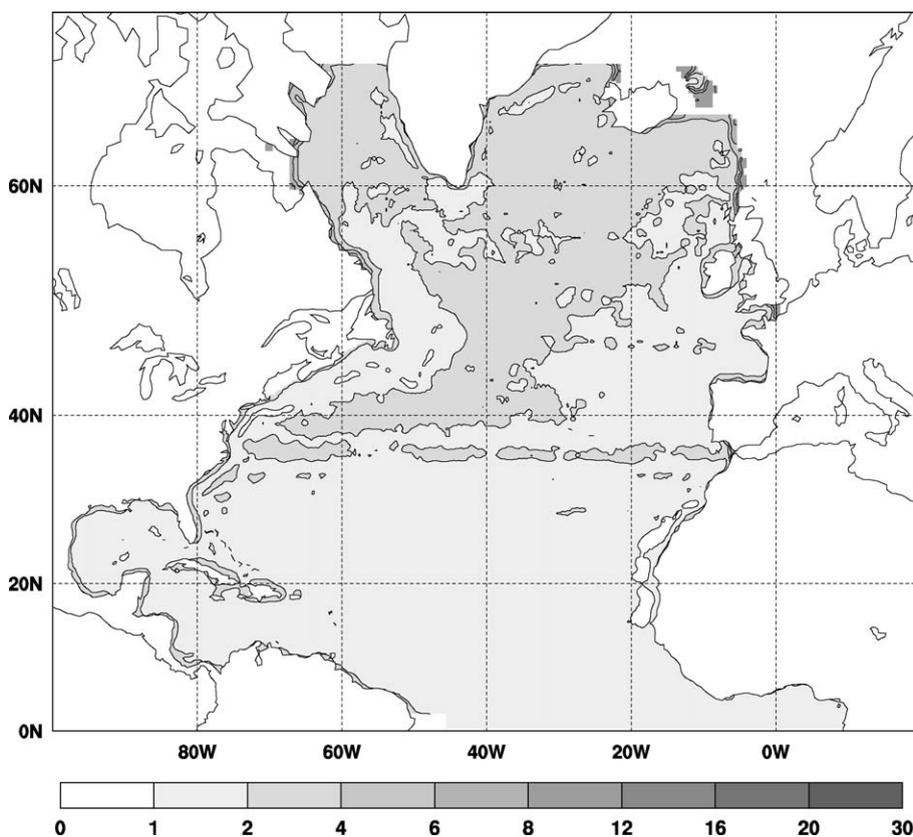


Fig. 16. Distribution of the forecast error for the sea surface height (cm) averaged over 1993 in the assimilation experiment using a constant amplification factor of $\mu = 1.25$.

representation of eddies in the assimilation experiment.

5.4. Forecast error statistics

The possibility to diagnose error statistics as a by-product of assimilation algorithms is one of the several motivations to develop advanced methodologies. Error estimates are useful per se to assign confidence levels to the oceanic field estimates, but also to verify the consistency of the prior statistical hypothesis needed by the methods.

As explained in Section 4, the parameterization of the forecast error is based on the idea of adaptive tuning of the forecast error. Informations contained in the innovation vector are used every assimilation cycle to compute a geographic distribution of the amplification factor according to Eq. (12). The results averaged during 1993 are illustrated by Fig. 13. The maximum amplification of the forecast error at a 10-day range takes place in the region of high mesoscale variability extending on both sides of the Gulf Stream path. This is a manifestation of the limited skill that one can expect from eddy permitting models to predict the synoptic evolution of mesoscale feature. By contrast, the amplification factor is smaller in the regions where the growth rates of the forecast error are representative of a slower dynamics, such as in the tropical regions.

The associated distribution of the forecast error on SSH is shown in Fig. 14. As expected, the maximum of the forecast error is found along the Gulf Stream path between Cape Hatteras and 40°N, where the forecast error standard deviation exceeds 20 cm in some places. Local maxima can also be detected along the North Atlantic Current extension and the Azores Current at 35°N. This picture of the forecast error looks fairly realistic, and can be considered as relevant of the first guess error statistics in asymptotic regime.

Two additional assimilation experiments have been conducted with prescribed amplification factors homogeneous in space, to test the sensitivity of the scheme to the adaptive parameterization and to assess the impact of these choices on the forecast error patterns. Figs. 15 and 16 show the forecast error distribution on SSH obtained with $\mathbf{D} = \mu \mathbf{I}$ for, respectively, $\mu = 5$ and $\mu = 1.25$. By comparison with the

adaptive experiment, the averaged amplification factor corresponding to Fig. 13 was $\mu = 2.7$. The main effect of taking a fixed amplification factor is to flatten the forecast error over the basin and to disconnect its distribution from the dynamical regimes. At first glance, these picture are less relevant of the model error, which is believed to be higher in the regions of strong mesoscale activity than elsewhere. In addition, the general performance of the assimilation scheme evaluated with respect to independent data or prior knowledge are significantly worse (Testut, 2000), and deserve less attention.

6. Conclusion

Hindcast experiments assimilating sea-surface temperature and sea-surface height data during 1993 have been successfully conducted with an eddy permitting circulation model of the North Atlantic basin. The assimilation method is an adaptation of a reduced-order Kalman filter (SEEK filter) based on a local parameterization of the background error covariance and a mechanism to extract all pertinent informations from the innovation vector adaptively.

This study has demonstrated the positive impact of satellite data to reconstruct the variability of the upper ocean circulation in the North Atlantic. In general, a reduction of the RMS misfit with respect to the assimilated data has been obtained, reflecting both a positive impact of the assimilation on the model bias and a significant improvement of the model in terms of variability. In addition, the analysis procedure is shown to be efficient to propagate the information from the surface SSH and SST data towards unassimilated variables in the interior of the ocean. With respect to our prior knowledge of the circulation, the pattern of the mean currents between the surface and 2000 m depth has been corrected positively in many areas like in the Gulf Stream region.

The validation of the hindcast results with independent data (the Reynaud climatology and XBT profiles collected during 1993) objectively demonstrates that the combined use of the two data sets allows the thermohaline properties of the upper ocean to be improved almost everywhere between the surface and 700 m depth. The strongest improvements are located in regions where the mesoscale activity

dominates the variability signal, like in the Gulf Stream extension.

In order to make the assimilation system stable and efficient during long integration periods, a simple adaptive scheme has been implemented to enforce the internal consistency between the forecast errors diagnosed by the filter and the statistics of the innovation sequence. This adaptive mechanism has been shown of critical importance to make these hindcast experiments successful. However, a single factor has been used to parameterize the amplification of the error in a given subsystem, irrespective of the model variable. A possible improvement could be to discriminate the amplification factor according to the model variables, as one can expect different error growth rates for the different physical quantities. Ensemble forecasts could be performed to examine this issue in more details.

In spite of these encouraging results, a number of implementation issues still form the subject of ongoing developments. The first limitation of the assimilation system used in this study concerns the numerical resolution of the model in the regions of strong mesoscale activity: an horizontal grid size of 1/3 at mid-latitude only permits the existence of mesoscale features in the solution, but such a resolution is still insufficient to resolve the underlying dynamics explicitly. The poor performances of the model predictions can be diagnosed from the forecast error distributions which exhibit quite high values along the Gulf Stream extension, and also from the too weak persistence score of the predictions at medium range (not discussed in the present paper). It will therefore be essential to reduce the model error as much as possible, for instance by increasing the horizontal resolution and by using the best possible atmospheric forcings.

Another current limitation is due to the use of gridded SSH and SST products instead of the original satellite measurements collected along tracks. The technical difficulties to remove this limitation have been addressed, and the benefit that can be drawn from the assimilation of original data without prior gridding has been evaluated in the context of academic models of the mesoscale ocean circulation. Other minor updates concern the implementation of a dynamical constraint in the adjustment operator to produce geostrophically balanced analysis states and

thereby, reduce the drift of the model forecast after re-initialization.

Further investigations have been undertaken recently, which will focus on the complementarity between the data sets used in this study and in situ measurements from hydrographic profiles, drifting buoys or surface salinity fields from climatologies. A more sophisticated observing system is expected to provide a better control of a number of the mixed-layer properties, which are not sufficiently constrained by satellite observations only.

Acknowledgements

This research was supported by Centre National d'Etudes Spatiales under contract CNES No. 794/987142. This work took place within the framework of the French MERCATOR project and the TOPAZ project of the European Communities (Grant EVK3-2000-00601). We would like to thank Dr. Eric Dombrowsky (CLS) for preparing the SSH and SST mapped observations, and Dr. Pascal Legrand (LPO/IFREMER) for providing us a mean SSH. We also thank Dr. Catherine Maillard (SISMER/IFREMER) for providing the XBT database, and Dr. Laurent Debreu for his help in dealing with the profile data. The calculations were conducted with the support of IDRIS/CNRS and the Centre Grenoblois de Calcul Vectoriel (CGCV-CEA).

References

- Ballabrera, J., Brasseur, P., Verron, J., 2001. Dynamical evolution of the error statistics with the SEEK filter to assimilate altimetric data in eddy-resolving ocean models. *Q. J. R. Meteorol. Soc.* 127, 233–253.
- Barnier, B., Siefridt, L., Marchesiello, P., 1995. Thermal forcing for a global ocean circulation model using a three-year climatology of ECMWF analyses. *J. Mar. Syst.* 6, 363–380.
- Bell, M.J., Forbes, R.M., Hines, A., 2000. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Syst.* 25, 1–22.
- Bennett, A.F., 1992. *Inverse Methods in Physical Oceanography* Cambridge Univ. Press, New York. 347 pp.
- Bennett, A.F., Chua, B.S., Harrison, D.E., McPhaden, M.J., 1998. Generalized inversion of tropical atmosphere-ocean (TAO) data and a coupled model of the Tropical Pacific. *J. Climate* 11, 1768–1792.
- Blanke, B., Delecluse, P., 1993. Variability of the tropical Atlantic

- ocean simulated by a general circulation model with two different mixed layer physics. *J. Phys. Oceanogr.* 23, 1363–1388.
- Brankart, J.M., Testut, C.E., Brasseur, P., Verron, J., 2003. Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: application to a 1993–96 hindcast of the North Atlantic Ocean circulation. *J. Geophys. Res.* (in press).
- Brasseur, P., Ballabrera, J., Verron, J., 1999. Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. *J. Mar. Syst.* 22 (4), 269–294.
- Brusdal, K., Brankart, J.M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P.J., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble based assimilation methods with a layered OCGM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* (this issue).
- Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* 126, 1719–1724.
- Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C., Busalacchi, A.J., 1996. Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.* 101 (C10), 22599–22617.
- Chua, B.S., Bennett, A.F., 2001. An inverse ocean modeling system. *Ocean Model.* 3, 137–165.
- Courtier, P., 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* 123B, 2449–2461.
- Dee, D., 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.* 123, 1128–1145.
- De Mey, P., Benkiran, M., 2001. A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. In: Pinardi, N., Woods, J.D. (Eds.), *Ocean Forecasting, Conceptual Basis and Applications*. Springer-Verlag, Heidelberg. 472 pp.
- De Mey, P., Awaji, T., Bell, M.J., Bennett, A.F., Brasseur, P., Evensen, G., Haines, K., Fukumori, I., Smedstad, O.M., Stammer, D., Weaver, A., 2002. Approaches to data assimilation within GODAE. *Proceeding of the En route to GODAE Symposium*, Biarritz.
- Durand, F., Gourdeau, L., Delcroix, T., Verron, J., 2002. Assimilation of sea surface salinity in a tropical OGCM: a twin experiment approach. *J. Geophys. Res.* 107 (C12), 8010.
- Evensen, G., 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99 (C5), 10143–10162.
- Evensen, G., van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamic. *Mon. Weather Rev.* 124, 1852–1862.
- Fukumori, I., 1995. Assimilation of TOPEX sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific Ocean. *J. Geophys. Res.* 100 (C12), 25027–25039.
- Fukumori, I., Malanotte-Rizzoli, P., 1995. An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.* 100 (C4), 6777–6793.
- Gelb, A., 1974. *Applied Optimal Estimation*. M.I.T. Press, Cambridge, MA. 374 pp.
- Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* 126 (3), 796–811.
- Ide, K., Bennett, A.F., Courtier, P., Ghil, M., Lorenc, A.C., 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn.* 75, 181–189.
- Lee, T., Marotzke, J., 1998. Seasonal cycles of meridional overturning and heat transport of the Indian ocean. *J. Phys. Oceanogr.* 28, 923–943.
- Le Grand, P., 1998. An inverse modelling estimate of the geoid height in the North Atlantic. *CERSAT Publ.* 9, 1–2.
- Le Traon, P.Y., Nadal, F., Ducet, N., 1998. An improved mapping method of multi-satellite altimeter data. *J. Atmos. Ocean. Technol.* 25, 522–534.
- Luong, B., Blum, J., Verron, J., 1998. A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems* 14, 979–997.
- Madec, G., Delecluse, P., Imbard, M., Levy, C., 1998. OPA 8.1 general circulation model reference manual. *Notes de l'IPSL* no. 11, Université P. et M. Curie, Paris. 91 pp.
- Penduff, T., Brasseur, P., Testut, C.E., Barnier, B., Verron, J., 2003. A four-year eddy-permitting assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean. *J. Mar. Res.* (in press).
- Pham, D.T., Verron, J., Roubaud, M.C., 1998. Singular evolutive extended Kalman Filter with EOF initialization for data assimilation in oceanography. *J. Mar. Syst.* 16 (3–4), 323–340.
- Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F., Simmons, A., 2000. The ECMWF operational implementation of four-dimensional variational assimilation: I. Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* 126, 1143–1170.
- Reynaud, T., Legrand, P., Mercier, H., Barnier, B., 1998. A new analysis of hydrographic data in the Atlantic and its application to an inverse modelling study. *Int. WOCE Newsl.* 32, 29–31.
- Smedstad, O.M., Hurlburt, H.E., Metzger, E.J., Rhodes, R.C., Shriver, J.F., Wallcraft, A.J., Kara, A.B., 2003. An operational Eddy-resolving 1/16° global ocean nowcast/forecast system. *J. Mar. Syst.* (this issue).
- Smith, W.H.F., Sandwell, D.T., 1997. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277, 1956–1962.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C.N., Marshall, J., 2003. Volume, heat and freshwater transports of the global ocean circulation during 1993–2000, estimated from a general circulation model constrained by WOCE data. *J. Geophys. Res.* 108 (C1), 3007.
- Testut, C.E., 2000. Assimilation de données satellitales avec un filtre de Kalman de rang réduit dans un modèle aux équations primitives de l'océan Atlantique. PhD thesis, Université Joseph Fourier, Grenoble, France.
- Testut, C.E., Brankart, J.M., Parent, L., 2001. An integrated system of sequential assimilation modules. *SESAM2.0 reference manual*, MEOM-LEGI, technical report.
- Treguier, A.M., et al., 1999. The CLIPPER project: high resolution modelling of the Atlantic. *Int. WOCE Newsl.* 36, 3–5.

- Verron, J., Gourdeau, L., Pham, D.T., Murtugudde, R., Busalacchi, A., 1999. An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical Pacific: method and validation. *J. Geophys. Res.* 104, 5441–5458.
- Voorrips, A.C., Heemink, A.W., Komen, G.J., 1999. Wave data assimilation with the Kalman filter. *J. Mar. Syst.* 19 (4), 267–291.
- Weaver, A.T., Vialard, J., Anderson, D.L.T., Delecluse, P., 2002. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. ECMWF Tech. Memo 365 (74 pp.).
- Wenzel, M., Schröter, J., Olbers, D., 2001. The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. *Prog. Oceanogr.* 48, 73–119.