Ocean Dynamics (2008) 58:1–17
DOI 10.1007/s10236-007-0128-z

Estimation of model errors generated by atmospheric forcings for ocean data assimilation: experiments in a regional model of the Bay of Biscay

Gregoire Broquet · Pierre Brasseur · David Rozier · Jean-Michel Brankart · Jacques Verron

Received: 7 June 2007 / Accepted: 18 July 2007 / Published online: 21 November 2007

Abstract The characterization of model errors is an essential step for effective data assimilation into open-ocean and shelf-seas models. In this paper, we propose an experimental protocol to properly estimate the error statistics generated by imperfect atmospheric forcings in a regional model of the Bay of Biscay, nested in a basin-scale North Atlantic configuration. The model used is the Hybrid Coordinate Ocean Model (HYCOM), and the experimental protocol involves Monte Carlo (or ensemble) simulations. The spatial structure of the model error is analyzed using the representer technique, which allows us to anticipate the subsequent impact in data assimilation systems. The results show that the error is essentially anisotropic and inhomogeneous, affecting mainly the model layers close to the surface. Even when the forcings errors are centered around zero, a divergence is observed between the central forecast and the mean forecast of the Monte Carlo simulations as a result of nonlinearities. The 3D structure of the representers characterizes the capacity of different types of measurement (sea level, sea surface temperature, surface velocities, subsurface temperature, and salinity) to control the circulation. Finally, data assimilation experiments demonstrate the superiority of the proposed methodology for the implementation of reduced-order Kalman filters.

Keywords Regional data assimilation · Model error · Atmospheric forcings · Bay of Biscay · Representers

Responsible editor: Jean-Marie Beckers

G. Broquet (✉) · P. Brasseur · D. Rozier ·
J.-M. Brankart · J. Verron
Laboratoire des Ecoulements Géophysiques et Industriels,
Université de Grenoble, BP 53,
Grenoble, 38041, France
e-mail: gbroquet@pmc.ucsc.edu

Present Address:
G. Broquet
Grégoire Broquet, Ocean Sciences Department,
University of California Santa Cruz, CA 95064, USA

1 Introduction

It is now well recognized that the correct specification of model error statistics is required to improve the performance of data assimilation systems in regional ocean and coastal models (Bennett 2002; Kurapov et al. 2005). The present study seeks to gain insights into the structure of the modeling error in coastal zones as a necessary prerequisite to implementing data assimilation techniques. The problem is studied using the Hybrid Coordinate Ocean Model (HYCOM) for a regional configuration of the Bay of Biscay with a 1/15° resolution, nested in a 1/3° resolution model of the North Atlantic. This configuration has three particular characteristics of coastal models that raise a number of issues in the context of the usual data assimilation schemes: high resolution, the existence of open boundaries, and the presence of shallow water over a large part of the domain (Celtic and Armorican shelves).

The actual model error associated with this particular configuration depends on both intrinsic parameterizations (e.g., resolution, subgrid scale parameterizations, turbulent closure scheme) and external data required to run the model (atmospheric forcings, initial conditions, open boundary conditions, bathymetry). Therefore, the characterization of the

model error structure requires prior hypothesis about the error sources.

The dynamics of the Bay of Biscay regional model is strongly constrained by atmospheric forcings and open boundary conditions. Actually, in a coastal region, specific atmospheric data sets would be required, as the standard products from the reanalysis of large-scale atmospheric data are generally unsuitable for reproducing the features of the regional circulation (Chelton et al. 2004; Koracin et al. 2004). A major assimilation issue therefore concerns the control of the error related to the poor determination of the atmospheric forcings and to the use of the necessarily approximate conditions at the open boundaries. The hypothesis that atmospheric forcing errors constitute the main source of error in ocean models has often been adopted (Miller and Cane 1989; Kurapov et al. 2005), but the related studies have generally focused on the role of wind stress error. In this article, the response of the model to errors in both wind and thermal atmospheric parameters is studied.

The aim of studying the response of the ocean model to errors in the atmospheric parameters is to characterize the corresponding model error with a view to parameterize this error later in a sequential assimilation scheme derived from the Kalman filter, such as the singular evolutive extended Kalman (SEEK) filter (Pham et al. 1998). A second objective is to identify those observations that are potentially useful for correcting this error. The study will focus in particular on sea surface height and temperature (from satellite observations) and on sea surface velocities (for instance from land-based high-frequency radars), which are among the most important data sets potentially available in coastal regions. Monte Carlo (or ensemble) simulations will be used to estimate statistics of the model error. This methodology is inspired by studies on forecast errors such as Auclair et al. (2003) and by assimilation schemes such as the ensemble Kalman filter (EnKF, Evensen 1994), which also conduct Monte Carlo simulations to forecast error statistics. Model error covariance will be studied by using the representer approach (Bennett 1992; Echevin et al. 2000). Twin data assimilation experiments using the SEEK filter are performed to evaluate the impact of this new parameterization of the model error. The proposed methodology, which is applied here to characterize the model error structure associated with uncertain atmospheric forcings, can be easily generalized to study the effects of other error sources.

After a description of the model used for the ensemble experiments (Section 2) and of the theoretical context of the Kalman filters underpinning these experiments (Section 3), the experimental protocol used will be presented (Section 4), followed by the main results dealing with the estimation of the model error due to uncertainties in the atmospheric parameters (Sections 5 and 6). Finally, data assimilation experiments comparing the use of this estimation to a more traditional parameterization are described (Section 7).

2 Model and configuration

The ocean circulation model used is version 2.1.3 of the hybrid coordinate model HYCOM (Bleck 2002; Chassignet et al. 2003). The model comprises a regional configuration of the Bay of Biscay, noted hereafter as BB15, nested in a configuration of the North Atlantic Ocean, noted hereafter as NA3 (cf. Fig. 1). This 1/3° configuration of the North Atlantic basin covers the zone 98°W-36.4°E and 27.95°S-70.03°N. It was developed within the context of the TOPAZ project and has already been used for assimilation experiments using the SEEK system (Brankart et al. 2003; Birol et al. 2004).

The regional configuration of the Bay of Biscay has a 1/15° resolution and covers the zone 14.80°W–1.04°W and 43.21°N–50.82°N, which includes the abyssal plain of the Bay of Biscay, the Armorican shelf, and the southern part of the Celtic shelf. It has four open boundaries, two meridional and two zonal: to the east in the English Channel, to the north, to the west, and to the south in the open Atlantic Ocean. In both configurations, subgrid-scale mixing is parameterized with a biharmonic dissipation operator, while vertical mixing is modeled by the K-profile parameterization (KPP) mixing scheme (Large et al. 1994). Tidal effects are not taken into account in both configurations.

The hybrid coordinate model can be viewed as a stack of layers that smoothly transform from isopycnal layers in the open stratified ocean to z coordinate layers in the mixed layer and unstratified areas, and to σ coordinate layers in shallow coastal regions. Thus, the model state vector comprises (1) three-dimensional fields describing each layer of the model: salinity (S), temperature (T), interface pressure (p), and zonal (u') and meridional (v') baroclinic velocities, and (2) two-dimensional fields describing the barotropic mode: the bottom pressure anomaly (p_b) and the zonal (u_b) and meridional (v_b) barotropic velocities.

Nesting of the models in a one-way off-line mode (without feedback of BB15 on NA3) is made possible with HYCOM by adjusting the open boundary conditions in BB15. These conditions are imposed with

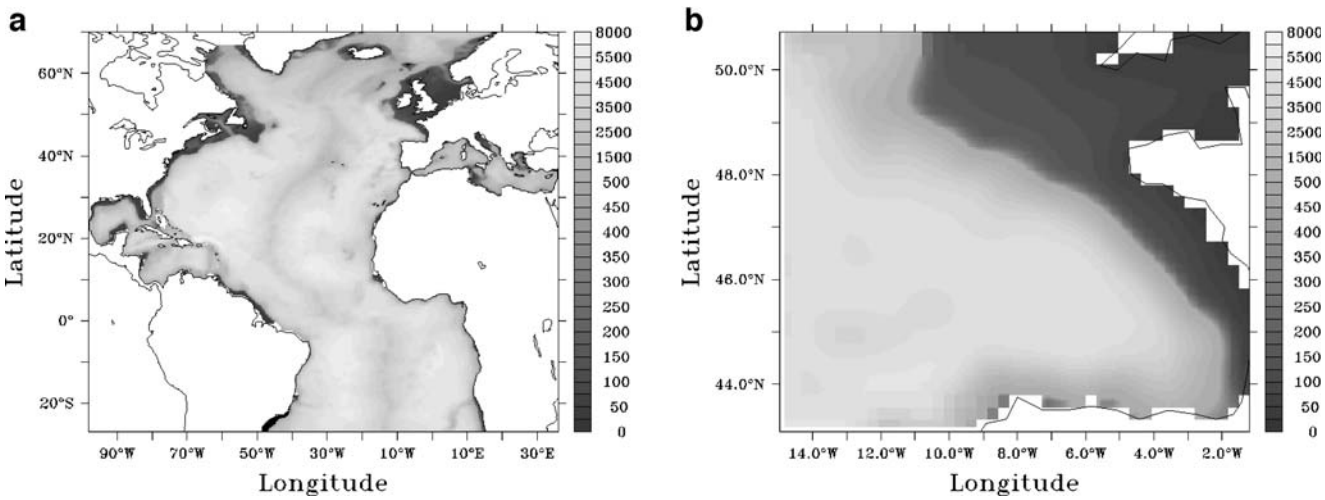

Fig. 1 Bathymetries (in meters) of nested configurations: **a** NA3 **b** BB15

the help of data that can be interpolated from the output of NA3 using standard schemes: the barotropic variables, (u_b, v_b) and p_b, are calculated at the open boundaries using the method of characteristics (Oliger and Sundström 1978; Browning and Kreiss 1982); a Newtonian relaxation (Davies 1976) towards the NA3 fields is applied to the field of S, T, p, and (u', v') in the vicinity of the open boundaries. The relaxation zones extend over 10 grid points from the open boundaries to the interior of the domain.

The bathymetry of NA3 is interpolated from ETOPO5 data (cf. Fig. 1). NA3 has 26 vertical discretization levels. The bathymetry of BB15 is simply interpolated from that of NA3, except in the relaxation zones where it reproduces exactly the bathymetry of NA3. The application of the conditions at the open boundaries is therefore perfectly coherent with the use of the hybrid coordinate. BB15 also has 26 hybrid layers in the vertical so that relaxations can be applied using the same layer discretization as that used in the NA3 configuration.

For the surface boundary conditions (the atmospheric forcings) of the model, Bulk formulations from Kara et al. (2000) are used to compute the net heat flux, whereas the momentum flux is directly available from wind stress data sets. There is no freshwater flux. The model is forced here with atmospheric parameters from the European Centre for Medium-Range Weather Forecasts (ECMWF) (with data every 6 h for the years 1979–2000), which have already been used for simulations on NA3 in the context of the TOPAZ project (Birol et al. 2004). There are seven atmospheric parameters at each point of the atmosphere–ocean interface: scalar wind w; long-wave radiation flux q_{lw}; short-wave radiation flux q_{shw}; air temperature T_a; air humidity E_v; and the two components of wind stress, τ_x and τ_y.

The simulation SIM-NA3 was carried out with NA3 to produce the initial and open-boundary conditions for BB15. It was initialized using the LEVITUS climatology and includes a spin-up phase of 10 years, then an integration phase with interannual forcings (obtained from ECMWF data) from 2 January 1985 to 1 January 1997. The simulation SIM-BB15 was then carried out with BB15 for the period from 1 January 1993 to 1 June 1995 with interannual forcings (as for SIM-NA3, this period of simulation was quite arbitrarily chosen with the only constraint to be covered by the available ECMWF data set). This was performed using the state of SIM-NA3 interpolated on the BB15 grid on 1 January 1993 as the initial condition and, for open-boundary conditions, data produced daily by SIM-NA3, also interpolated on the BB15 grid. Sea surface salinity and temperature (SSS and SST) in SIM-NA3 and SIM-BB15 were relaxed towards LEVITUS climatological data. This relaxation was not used when running Monte Carlo simulations.

3 Error statistics for assimilation studies

In this section, the model error associated to the surface boundary conditions is described in mathematical terms, focusing on its contribution to a sequential assimilation scheme.

3.1 The model error

The notations adopted here are those recommended by Ide et al. (1997). The state vector is noted $\mathbf{x} \in R^m$. In

a Kalman filter, the model is assumed to be imperfect, and the model error is taken into account at the filter forecast stage (between time t_i and time t_{i+1}) to estimate the forecast error covariance matrix $\mathbf{P}^f_{i+1} \in R^{m\times m}$ as follows:

$$\mathbf{P}^f_{i+1} = \mathbf{M}\mathbf{P}^a_i\mathbf{M}^T + \mathbf{Q} \tag{1}$$

The model error covariance matrix $\mathbf{Q} \in R^{m\times m}$ is added to the covariance matrix on the initial condition $\mathbf{P}^a_i$ (resulting from the previous analysis at time t_i) propagated by the model $\mathbf{M} \in R^{m\times m}$ (or its linear tangent if the model is nonlinear). $\mathbf{Q}$ corresponds to the model error accumulated during the assimilation cycle $[t_i, t_{i+1}]$.

The forecast error covariance matrix $\mathbf{P}^f_{i+1}$ is then used in the analysis stage to calculate the optimal gain:

$$\mathbf{K}_{i+1} = \mathbf{P}^f_{i+1}\mathbf{H}^T\left[\mathbf{H}\mathbf{P}^f_{i+1}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{2}$$

where $\mathbf{H} \in R^{p\times m}$ is the observation operator and $\mathbf{R} \in R^{p\times p}$ the observation error covariance matrix. This gain enables observations $\mathbf{y}^o_{i+1} \in R^p$ to be taken into account in an optimal manner when calculating the analyzed state $\mathbf{x}^a_{i+1}$ from the forecast state $\mathbf{x}^f_{i+1}$:

$$\mathbf{x}^a_{i+1} = \mathbf{x}^f_{i+1} + \mathbf{K}_{i+1}\left[\mathbf{y}^o_{i+1} - \mathbf{H}\mathbf{x}^f_{i+1}\right] \tag{3}$$

A good parameterization of the model error in data assimilation schemes based on the Kalman filter is critical in guaranteeing the optimality of the filter and in ensuring that observations are correctly taken into account. This is even more important in a reduced-order Kalman filter such as the SEEK filter (Pham et al. 1998; Brasseur and Verron 2006), where $\mathbf{P}^f$ is represented in the form $\mathbf{P}^f = \mathbf{S}^f\mathbf{S}^{fT}$, which is a decomposition along the few dominant directions of the forecast error. It is therefore essential to identify the dominant directions of error growth.

The main sources of model error can be identified as uncertainties in distinct parts of the model parameterization. For example, in the present configuration, they are related to uncertainties in atmospheric forcings, parameterization of the subgrid-scale diffusion, bathymetry, boundary conditions, parameterization of the turbulence closure scheme, etc. Under the assumption that these distinct sources generate independent errors that combine in the model error, the model error can be characterized by focusing on the isolated effects of the main sources. This focus is made possible under the assumption that a particular source of error is the only source of model error.

The error arising from uncertainties about the atmospheric parameters constitutes an important part of the model error, especially in the case of regional configurations for which regional atmospheric forcings with appropriate scales are not available. Estimation of $\mathbf{Q}$ in the present study is based on the hypothesis that uncertainties about the atmospheric parameters are the only source of model error (which is an hypothesis used by Miller and Cane 1989). It is based on Monte Carlo simulations, as described in Section 4. Estimating $\mathbf{Q}$ prior to the assimilation experiments is justified by the fact that, in the Kalman filter theory, it is assumed that the model error is independent of forecast errors, analysis errors, observation errors, etc., generated during the assimilation experiments.

To evaluate the impact of estimating $\mathbf{Q}$ in such a manner, it is assumed that the analyzed error $\mathbf{P}^a$ is negligible in a reduced-order Kalman filter, so that Eq. 1 is approximated by

$$\mathbf{P}^f_{i+1} = \mathbf{Q} \tag{4}$$

which corresponds to an assimilation scheme similar to that used by Oke et al. (2002) or Kurapov et al. (2005).

The protocol proposed here to estimate the model error due to uncertainties in the atmospheric parameters could be extended to estimate the model error due to uncertainties in the other main types of model parameterization. While this study focus on the effect of only one source of model error, it still provides a way of characterizing a more realistic model error through the combination of such different error estimations, under the hypothesis that they are independent.

3.2 Definition of representers

The interpretation of the model response to the error in the atmospheric parameters, expressed by its covariance matrix $\mathbf{Q}$, will be facilitated here by the use of representers as proposed by Echevin et al. (2000). Note that the definition of the representers used in this paper is not the same as the standard definition of Bennett (1992), where representers are a function of space and time. Here, we use the definition of Echevin et al. (2000), and our representers can be considered as Bennett's representer snapshots. Taking Eqs. 2 and 3, the correction made by a Kalman filter at an analysis stage may be written as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{x}^c \tag{5}$$

where $\mathbf{x}^c$ is a linear combination of the columns of the matrix $\mathbf{P}^f\mathbf{H}^T$. In the special case where, during this analysis phase, only one observation $y^o \in R$ is assimilated (the dimension of $\mathbf{H}$ is $1 \times m$, so that $\mathbf{R}$ and

$\mathbf{HP}^f\mathbf{H}^T$ are scalar), $\mathbf{x}^c$ is thus proportional to $\mathbf{P}^f\mathbf{H}^T$. Thus, taking Eqs. 2 and 3 again gives us:

$$\mathbf{x}^c = \lambda \mathbf{r} \quad \text{where} \quad \lambda = \frac{y^o - \mathbf{H}\mathbf{x}^f}{1 + \frac{\mathbf{R}}{\mathbf{HP}^f\mathbf{H}^T}} \tag{6}$$

and

$$\mathbf{r} = (\mathbf{P}^f\mathbf{H}^T)/(\mathbf{HP}^f\mathbf{H}^T) \tag{7}$$

$\mathbf{r} \in R^m$. In particular, if $y^o - \mathbf{H}\mathbf{x}^f = 1$ and $\mathbf{R} = 0$, $\mathbf{x}^c = \mathbf{r}$. $\mathbf{r}$ is called, for the sake of convenience, the representer associated with the observation y^o, and it thus gives the exact correction made with a Kalman filter from a unitary innovation on y^o, with zero observation error. In the general case of an assimilation of an ensemble of scalar observations, the correction carried out is a linear combination of the representers associated with these observations.

If the representer is associated with a state variable x_i of the system (assumed directly observed), then the jth component of $\mathbf{r}$ is $r_j = \mathbf{P}^f_{i,j}/\mathbf{P}^f_{i,i}$. The structure of the representer is thus close to that of the correlations between the error in x_i and the errors in the other system variables x_j. Note that, following Eq. 7, the representer component r_j is nondimensional if x_i and x_j have the same physical unit; if not, the physical unit of the component r_j is the unit of x_j divided by the unit of x_i. For instance, if a representer associated with an innovation in sea surface elevation is computed, the unit of a velocity component of the representer is s^{-1}.

The representers, in addition to providing information comparable to the correlation calculations for the exploration of the error covariance matrices, thus also make it possible to visualize the exact influence of the observations on the model through the use of data assimilation. In our study, adopting hypothesis Eq. 4, Eq. 7 becomes

$$\mathbf{r} = (\mathbf{QH}^T)/(\mathbf{HQH}^T) \tag{8}$$

The spatial structure of the representers of the model error will thus enable exploration of the corrections, which are calculated in response to the perturbations in the atmospheric forcings.

In the more general case, however, Eqs. 1 and 7 show that, in situations where the forecast error is a combination of the term $\mathbf{MP}^a\mathbf{M}^T$ and of several types of independent model errors $\mathbf{Q}_k$, the representers will be a linear combination of the representers, calculated under the assumption that other errors are negligible, for $\mathbf{MP}^a\mathbf{M}^T$ and each $\mathbf{Q}_k$.

4 Monte Carlo simulations

In the Monte Carlo method used here, the estimation of model error statistics is based on an ensemble of simulations resulting from different sets of atmospheric parameters that consistently sample the probability distribution of the error in these parameters. Note that, in this study, the sample of the probability distribution function is not chosen randomly, as is usually the case in Monte Carlo simulations. It is generated using an ad hoc approach similar to the method used in Oke et al. (2002) and Kurapov et al. (2005).

In addition, it is assumed that changes in the model error due to uncertainties about atmospheric forcings are relatively limited at the scale of a season, so that it is possible to estimate $\mathbf{Q}$ as constant over a given season.

An essential aspect of the calculation of the model error covariance matrix is the determination of the minimum size of the ensemble required to obtain a correct representation of the probability distribution. A series of model error covariance matrices is thus estimated from ensembles of increasing sizes, although the best estimate of $\mathbf{Q}$ is assumed to be calculated from the entire ensemble of Monte Carlo simulations that have been computed.

4.1 Generation of the ensemble of atmospheric parameter sets

It is assumed that the probability distribution of error on the atmospheric parameters for a given period (August–October in our case) is consistently sampled by an ensemble of real sets of atmospheric parameters obtained from many different years, or from the same year but staggered over time with intervals larger than synoptic temporal scales. The probability distribution of errors is studied for the period from August to October because the error statistics can be assumed relatively constant during the corresponding period.

As ECMWF data are only available for a limited number of years, it was necessary to use not only the interannual variability of the atmospheric parameters but also their weekly variability to have a sufficiently large ensemble of parameter sets. As illustrated in Fig. 2, the ensemble was thus constructed by taking eight samples of 31 days of sets of atmospheric parameters in each of the 22 years of ECMWF data (from 1979 to 2000). For each of the 22 years, the samples cover periods staggered at 7-day intervals, from August 1 onwards. These parameters, therefore, cover the period from August 1 to October 20.

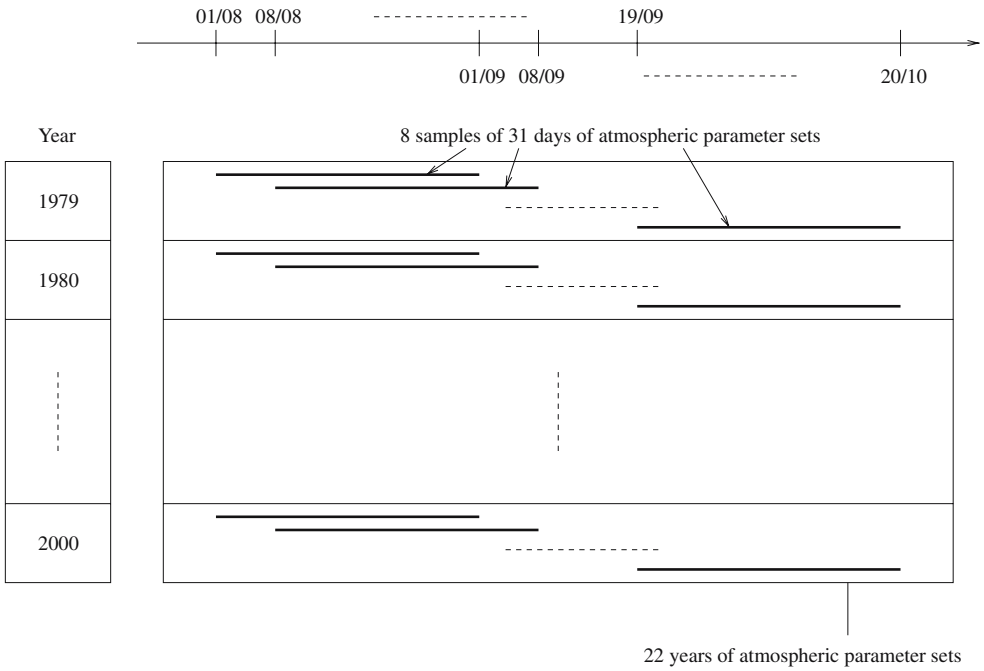

Fig. 2 Schematic representation of the ensemble of 176 atmospheric data sets extracted from the ECMWF forcing fields

4.2 Generation of the ensemble of model trajectories

Using the ensemble of atmospheric parameter sets described above, eight subsets of 22 simulations of 31 days on BB15 were generated, as shown in Fig. 3. Each of these eight subsets is related to one of the 31-day periods of the year previously defined as staggered at 7-day intervals. In each subset, all the simulations have the same initial state and the same data for open-boundary conditions as SIM-BB15 at the corresponding

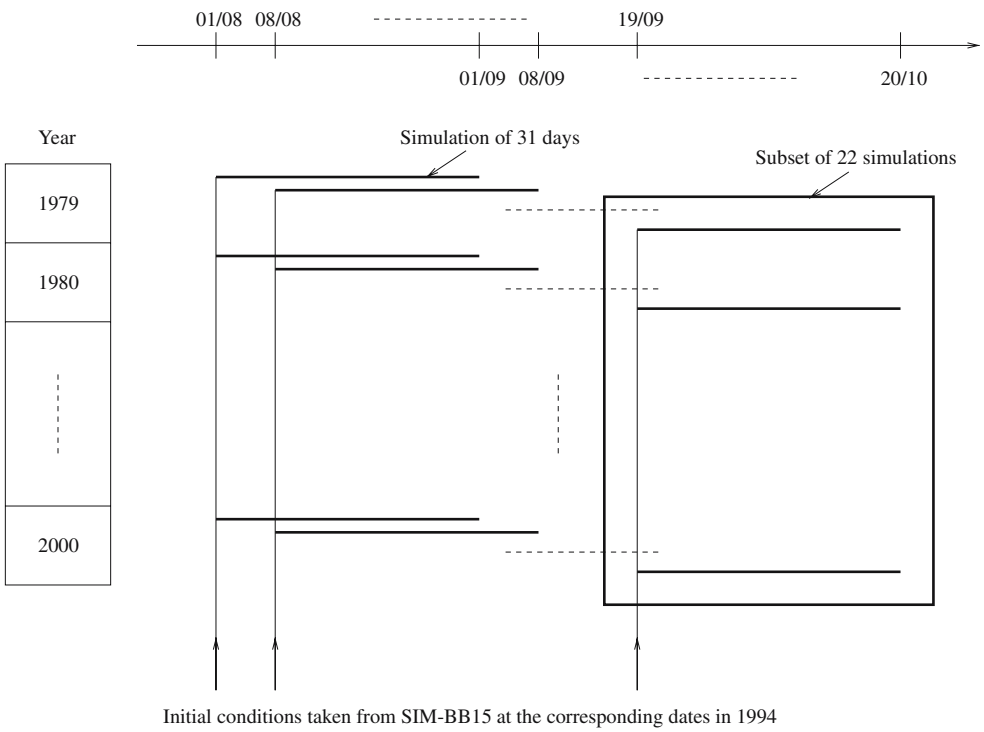

Fig. 3 Schematic representation of the ensemble of 176 simulations of 31 days conducted during 1994

dates in 1994. This use of several subsets has been designed to run ensemble experiments with consistent initial conditions and forcing fields. The dispersion on each subset of 22 simulations comes from the use of 22 different years of ECMWF data. The dispersion of simulations between the different subsets is representative of the weekly variability in the ECMWF data, if we exclude the dispersion due to the difference in the initial conditions and in the data for open-boundary conditions.

4.3 Ensemble covariance and mean

To study the error covariance arising from the ensemble model trajectories described earlier, anomalies are constructed for each simulation in relation to a reference linked to the subset to which it belongs. It is assumed that the influence due to the initial conditions and the data for the open boundary conditions is independent of the variations between the ECMWF forcings used in the simulations of the same subset, so that this influence is identical for all the simulations of the same subset. By calculating anomalies, the influence due to the initial conditions and the data for the open boundary conditions particular to each subset is thus eliminated with the help of a suitable reference simulation.

These considerations led us to calculate two different reference simulations for each subset, as illustrated in Fig. 4:

- The central simulation, which is the simulation obtained from the sets of mean parameters for the 31 days considered (the mean being calculated for the 22 years of ECMWF data), referred to hereafter as $S(\overline{f})$. The calculation of the sets of mean parameters is used to calculate the anomalies of the other sets of forcing parameters in relation to them and, thus, to define anomalies in the forcing parameters.
- The mean simulation (the mean of the 22 simulations in the subset), referred to hereafter as $\overline{S(f)}$

Calculations of model anomalies will only be carried out in relation to the $S(\overline{f})$, which, unlike the $\overline{S(f)}$, are solutions of the model. The first reason is that HYCOM requires numerous constraints on the state variables (linked in particular to the hybrid vertical coordinate), to ensure that the simulation outputs are physically consistent, and these constraints are not applied to $\overline{S(f)}$. Furthermore, the central forecasts are more consistent with the evolution of the best estimate obtained from a deterministic filter like the SEEK. However, our covariance estimations give similar results if the differences between the $S(\overline{f})$ and the $\overline{S(f)}$ do not depend on the subsets of simulations. Section 5 shows that there are actually differences between central and mean simulations, but these differences are systematic and mainly similar in all subsets of simulations, so that the choice of the reference simulation has a weak impact on the covariance estimations. The computation of the covariances using $\overline{S(f)}$ instead of $S(\overline{f})$ has also been carried out, giving very similar results, which are not presented here. $\overline{S(f)}$ will only be used here for direct comparisons with the $S(\overline{f})$.

As a result of the procedure described above, an ensemble of 176 sets of anomalies in the atmospheric parameters (for 31 days with data every 6 h) centered around 0 is thus available to compute an ensemble of 176 trajectories of anomalies to the $S(\overline{f})$ in the corresponding model outputs written as:

$$\{\delta\mathbf{x}_k(t)\ ,\ t \in [0, 31d]\}_{k\in[1,176]} \tag{9}$$

It was thus possible to calculate, for a fixed time $t \in [0, 31d]$ and a fixed ensemble size $n \in [2, 176]$, the estimation of the covariances on the ensembles of model anomalies $\{\delta\mathbf{x}_k(t)\}_{k\in[1,n]}$ given by $\frac{1}{n-1}\sum_{k=1}^{n}(\delta\mathbf{x}_k - \overline{\delta\mathbf{x}})(\delta\mathbf{x}_k - \overline{\delta\mathbf{x}})^T(t)$ where $\overline{\delta\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n}\delta\mathbf{x}_k$.

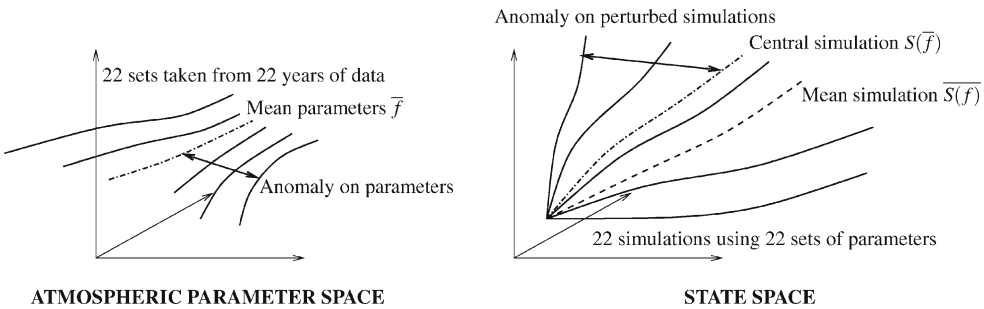

Fig. 4 Schematic representation of the ensemble mean and associated spread of the forcings and model simulations

The model error covariance matrix computed from the ensemble of size n is given by

$$\mathbf{Q}^n = \frac{1}{n-1}\sum_{k=1}^{n}(\delta\mathbf{x}_k - \overline{\delta\mathbf{x}})(\delta\mathbf{x}_k - \overline{\delta\mathbf{x}})^T(t_{KF}) \qquad (10)$$

This is the covariance matrix of the ensemble of model anomalies after model integration time t_{KF} corresponding to the sequential assimilation cycle. The value of t_{KF} is normally chosen to fit the length of the assimilation window. That is why it will depend on the system parameterized with such model error covariance matrices. However, t_{KF} is fixed here at $t_{KF} = 15d$ when studying the model error structures, and at $t_{KF} = 5d$ for the data assimilation experiment described in Section 7 (the model error is better characterized at $t_{KF} = 15d$; on the other hand, $t_{KF} = 5d$ is more representative of traditional assimilation cycles of the system that will be used for data assimilation). Representers associated with $\mathbf{Q}^n$ matrices are noted as $\mathbf{r}^n$. The best estimate of the model error covariance is assumed to be given here by $\mathbf{Q}^{176}$.

5 Divergence between the central and mean forecasts

It may be observed that the ensemble of 176 sets of anomalies on the atmospheric parameters has a quasi-Gaussian distribution if the anomalies on scalar wind w are excluded. As an example, the anomaly distribution for air temperature at point (11.7°W, 45.4°N) on the 16th day of simulation is given in Fig. 5.

Despite this characteristic, differences appear between $S(\overline{f})$ and $\overline{S(f)}$, which is not surprising given the significant nonlinearity effects in the model. These differences steadily increase and are the most significant on temperature. Figure 6 illustrates the fact that these differences are related to a general trend observed in all the simulations and not to the behavior of a few

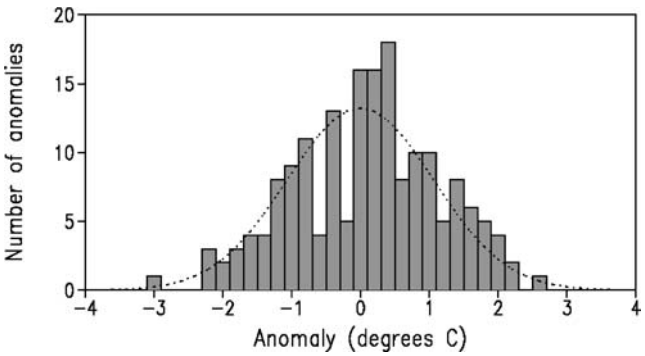

Fig. 5 Distribution of the 176 anomalies at $t = 16d$ for T_a at point (11.7°W, 45.4°N)

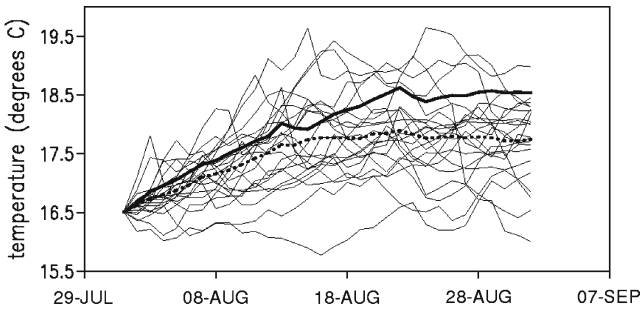

Fig. 6 SST at point (11.7°W, 45.4°N) on first subset of 22 simulations, with $\overline{S(f)}$ (*bold dashed line*) and $S(\overline{f})$ (*bold line*)

atypical simulations. It shows the evolution of SST at point (11.7°W, 45.4°N) for the different simulations of the first subset and for the associated simulations $S(\overline{f})$ and $\overline{S(f)}$. The order of magnitude of the divergence between $S(\overline{f})$ and $\overline{S(f)}$ is not negligible in relation to the standard deviation of the dispersion.

The general tendency for a mean deviation from $S(\overline{f})$ to appear is also shown in terms of spatial structure. Figure 7 shows a vertical cross-section of temperature differences (at latitude 45.4°N) between $S(\overline{f})$ and $\overline{S(f)}$, the mixed layer depth in $S(\overline{f})$, and the mean of the mixed layer depth in the 22 simulations after 15 and 30 days in the first subset (this latter mean defines the mixed layer depth in $\overline{S(f)}$). Figure 7 also shows that the mean deviation from $S(\overline{f})$ is characterized by a relative horizontal homogeneity (at a given geopotential depth) of temperature differences. Simulation $S(\overline{f})$ gives a temperature higher than that given by $\overline{S(f)}$ down to a depth of approximately 10 m below the mixed layer. At greater depths, however, and down to depths of 80 to 100 m, simulation $S(\overline{f})$ gives a lower temperature than that given by $\overline{S(f)}$. Finally, the divergence is virtually nonexistent below a depth of 100 m. This warming over time of the mixed layer of $S(\overline{f})$ in relation to $\overline{S(f)}$ occurs at the same time as its cooling in relation to $\overline{S(f)}$ at greater depths. These phenomena occur during the natural thickening of the mixed layer at the end of the summer.

The combination of two processes, a nonlinear rectification and the ill-posed estimation of the average of a vector variable (the wind stress, whose amplitude has been significantly reduced), may be responsible for this divergence between the central and mean forecasts. The difference in mixed-layer depths could be related to the determination of τ_x and τ_y in the sets of mean parameters: as τ_x and τ_y vary considerably $||\overline{\tau}|| << \overline{||\tau||}$. A weak wind stress is therefore applied to $S(\overline{f})$. This causes weaker mixing under the ocean surface in $S(\overline{f})$. As the mixed layer is shallower in $S(\overline{f})$, heat from the surface nonradiative fluxes is less propagated

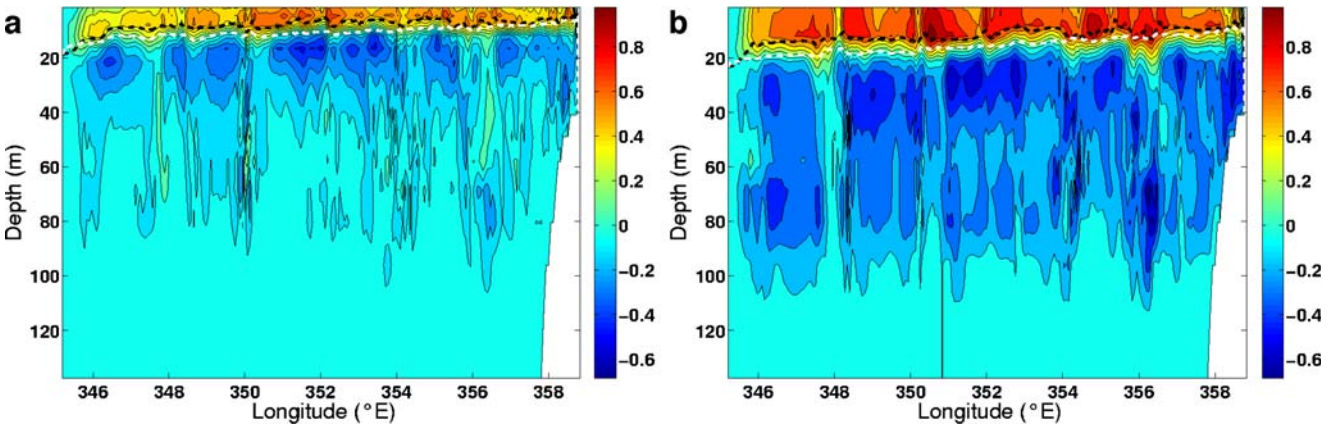

Fig. 7 Vertical section at latitude 45.4°N of temperature differences (in °*C*) between $S(\overline{f})$ and $\overline{S(f)}$ after **a** 15 days and **b** 30 days in the first subset of simulations. The *black dashed line* indicates the mixed layer depth in $S(\overline{f})$ and the *white dashed line* indicates the mean of the mixed layer depths on the 22 simulations of the first subset

to the great depths. This explains the warming of the upper water layers and the cooling of the deeper layers for $S(\overline{f})$ in relation to $\overline{S(f)}$. The other nonlinear phenomena affecting the model probably only have a limited impact on the divergence between $S(\overline{f})$ and $\overline{S(f)}$. Thus, the averaged model parameters or variables in the definition of $S(\overline{f})$ and $\overline{S(f)}$ are shown to be less realistic.

6 Model error covariances

6.1 Convergence of representer calculations

Before analyzing model error covariance structures (with the best estimate of the model error covariance matrix $\mathbf{Q}^{176}$), it is important to ensure that the calculations of model error covariance matrices $\mathbf{Q}^n$ have converged in relation to size $n \leq 176$ of the ensembles of anomalies $\{\delta\mathbf{x}_k(t_{KF})\}_{k\in[1,n]}$, from which they were obtained. This will show that the ensemble $\{\delta\mathbf{x}_k(t_{KF})\}_{k\in[1,176]}$ is sufficiently large to express the covariance of the model error using the Monte Carlo method presented in Section 4.

As the convergence of the calculations for error covariance matrices is technically difficult to estimate, the demonstration of this convergence was restricted to a few cases of representers associated with the model state variables. Figure 8 shows the changes in relation to n of the RMS differences on the SST between the representers $\mathbf{r}^n$ and $\mathbf{r}^{n-10}$ of SST at different points. The figure shows that the representers $\mathbf{r}^n$ no longer really change in relation to n when $n \geq 130$, which implies that, with 176 model error members, it may be considered possible to calculate representative matrices of the model error covariance.

In the following, representers $\mathbf{r}^{176}$ will be described.

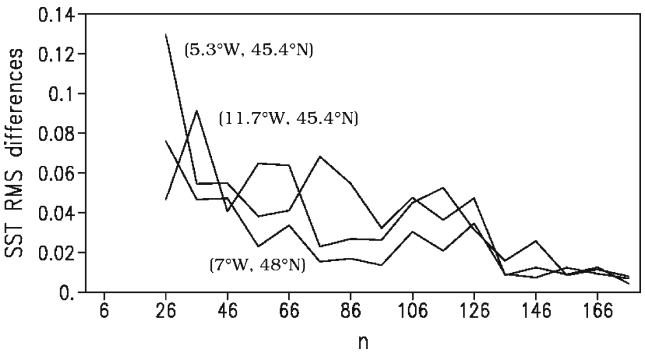

Fig. 8 Evolution according to n of RMS differences on the SST between the representers r^n and r^{n-10} associated with SST observations at different points in the model

6.2 Horizontal extension of the correction induced in the mixed layer by the representers of surface variables

The study of the representers associated with SST and sea surface height (SSH) data, which are the main observations potentially available for assimilation, shows that these observations have an extensive horizontal influence in the mixed layer. This is illustrated in Fig. 9, showing the SST value of the representers associated with the SST at point (11.7°W, 45.4°N) in the basin and at point (1.4°W, 45°N) near the coast of the Bay of Biscay. The corresponding correlations are also indicated in these figures. An innovation of 1 degree at these points would imply corrections of more than 0.4 degrees on the SST of the greater part of the domain. These figures show that the extent of the influence of

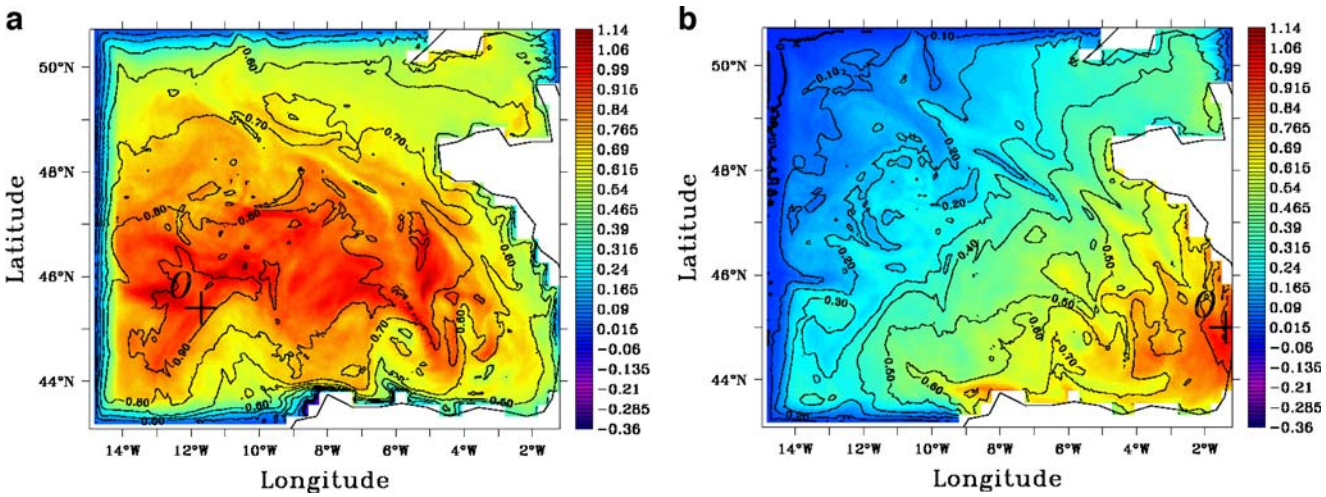

Fig. 9 SST fields (*color*) of representers r^{176} associated with the SST at point O and correlation fields (*isolines*) between SST and the SST at point O: **a** O=(11.7°W, 45.4°N), **b** O=(1.4°W, 45°N)

an observation is strongly related to its position and, in particular, to whether it is situated in the basin or near the continental slope. A key point is the absence of isotropy in the model error, which is characteristic of coastal environments (Echevin et al. 2000). Correlation fields between SST or SSH observations and the other variables of the system have structures that are very similar to those of the associated representers, with significant values over a large part of the surface. It confirms that the horizontal influence of SST and SSH observations, as indicated by the representers, covers an extensive area.

The influence of surface velocities on other velocities is more localized in the vicinity of these data. This can be seen in Fig. 10, which shows the surface velocity fields of the representers associated, respectively, with meridional and zonal surface velocities at point (11.7°W, 45.4°N) in the basin and point (7°W, 48°N) over the shelf break. The orientation of these fields is virtually uniform in the direction of the observed velocity. However, the bathymetric gradient through the shelf break and its associated model currents affect these fields. The surface velocities of the zonal velocity representer at point (7°W, 48°N) are clearly weakened over the main part of the shelf break. As the mean currents in this zone are intense due to the bathymetric slope, atmospheric perturbations would have a limited influence on them.

The influence of surface velocities on model temperature and salinity also varies considerably, depending on the location and orientation of these velocities. This marked contrast may be related to the origin of the

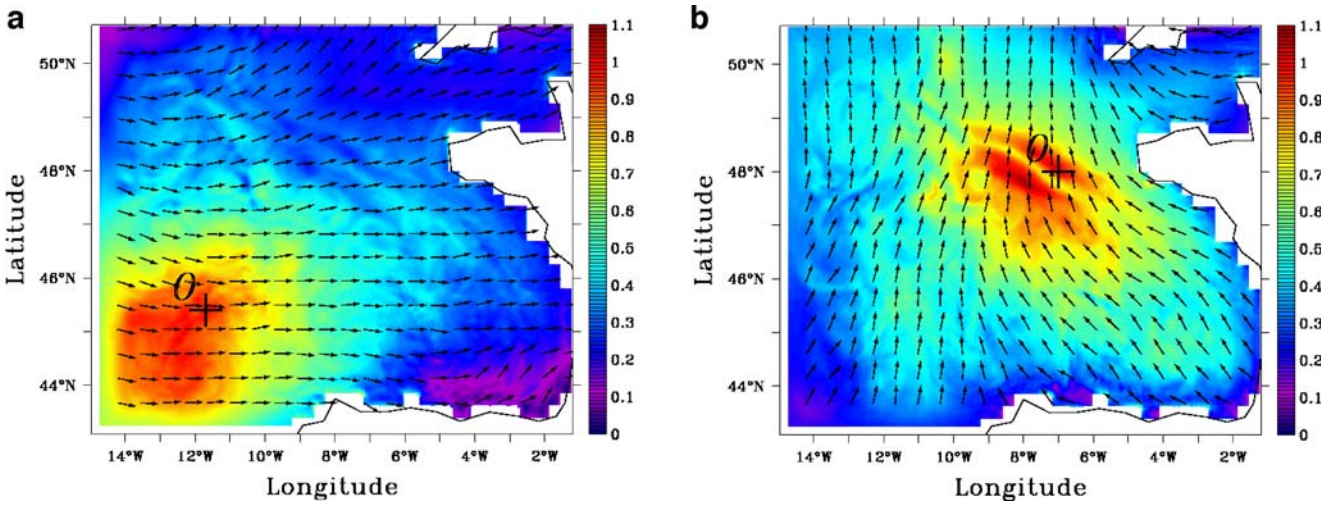

Fig. 10 Surface velocity fields – modules and directions (every 10^{th} vector is displayed) – of representers r^{176} associated with surface velocities at point O: **a** representer of zonal velocity, O=(11.7°W, 45.4°N); **b** representer of meridional velocity, O=(7°W, 48°N)

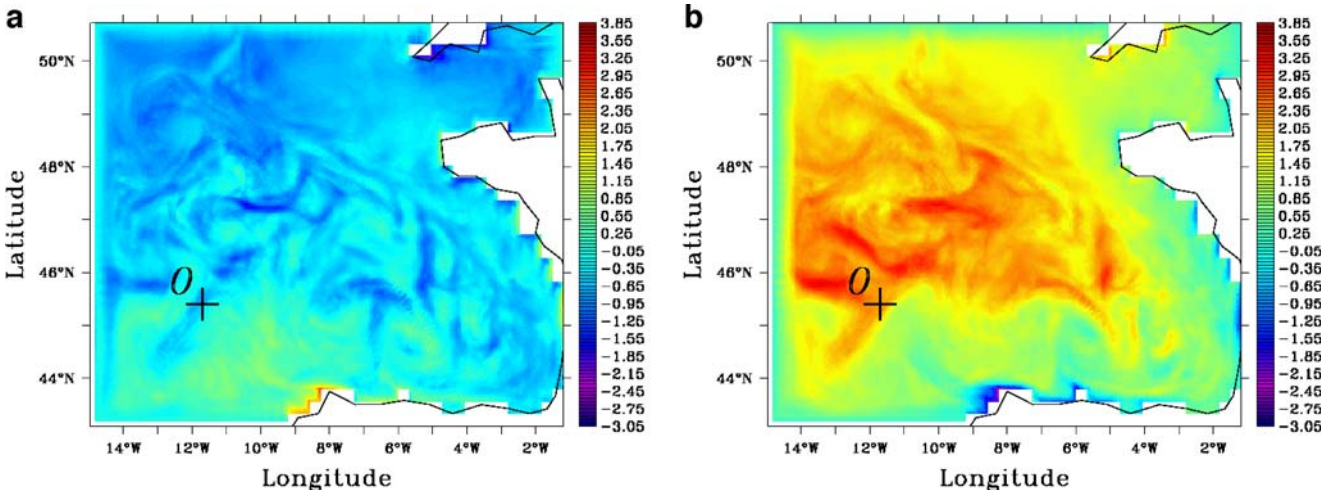

Fig. 11 SST fields (in $°C.s.m^{-1}$) of representers r^{176} associated with surface velocity at point O=(11.7°W, 45.4°N): **a** representer of zonal velocity; **b** representer of meridional velocity

variability in surface velocity, depending on whether this variability is determined mainly by adjustments to heat flux variations or by the perturbation from wind forcings. The SST of the representers associated with the observations of surface zonal and meridional velocities at point (11.7°W, 45.4°N) is given in Fig. 11a, b. These fields are characteristic of a general tendency observed for positive zonal surface velocities to induce negative temperature fields and for positive meridional surface velocities to induce positive temperature fields in the mixed layer through the calculation of their associated representers. It may be assumed that this general tendency is linked to the uniform inflow of cooler or warmer surface waters, respectively, from the north west or south west, generated by the current corrections in the mixed layer. The opposite tendency near the coasts can generally be attributed to upwelling or downwelling phenomena. The wind stress variations would thus play a predominant role in the response of the model to atmospheric perturbations.

6.3 Influence of representers at open boundaries

Near the open boundaries, the strong relaxation towards the data from SIM-NA3 explains the limited influence of the perturbation in atmospheric fluxes on salinity, temperature, or baroclinic velocities. Barotropic open boundary conditions, on the other hand, allow the SSH and the barotropic velocities to be significantly modified by these perturbations. This is why representers can induce corrections on barotropic variables but not on three-dimensional variables near the open boundaries, as can be seen in Figs. 9, 10, 11, and 12. However, in Fig. 12, which shows the SSH of the representer associated with the SSH at a point in the basin, it may be noted that associated corrections vanish near the southern and western boundaries. It may be assumed that the barotropic open boundary conditions have not sufficiently taken into account the internal model variations linked to the atmospheric perturbations. Another interpretation would be that the principal variations in SSH due to perturbations in atmospheric fluxes are linked to steric effects and, thus, to temperature variations. This would explain why such extensive structures characterize the SSH fields in representers.

6.4 Influence of representers of surface variables at depth

As most of the data to be assimilated are observations of surface properties, it is important to indicate how the

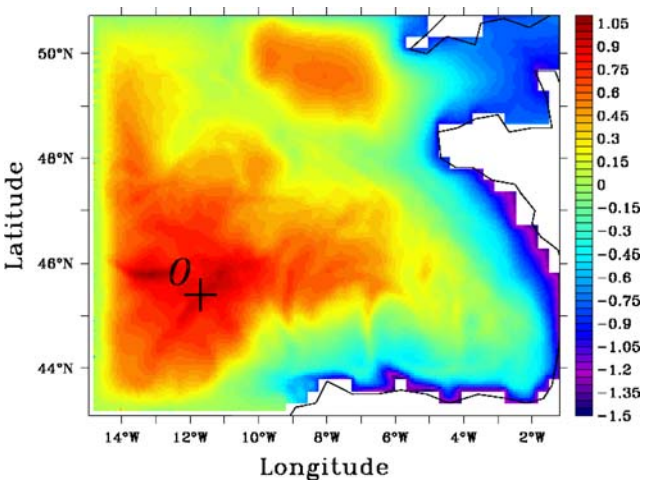

Fig. 12 SSH field of the representer r^{176} associated with SSH at point O=(11.7°W, 45.4°N)

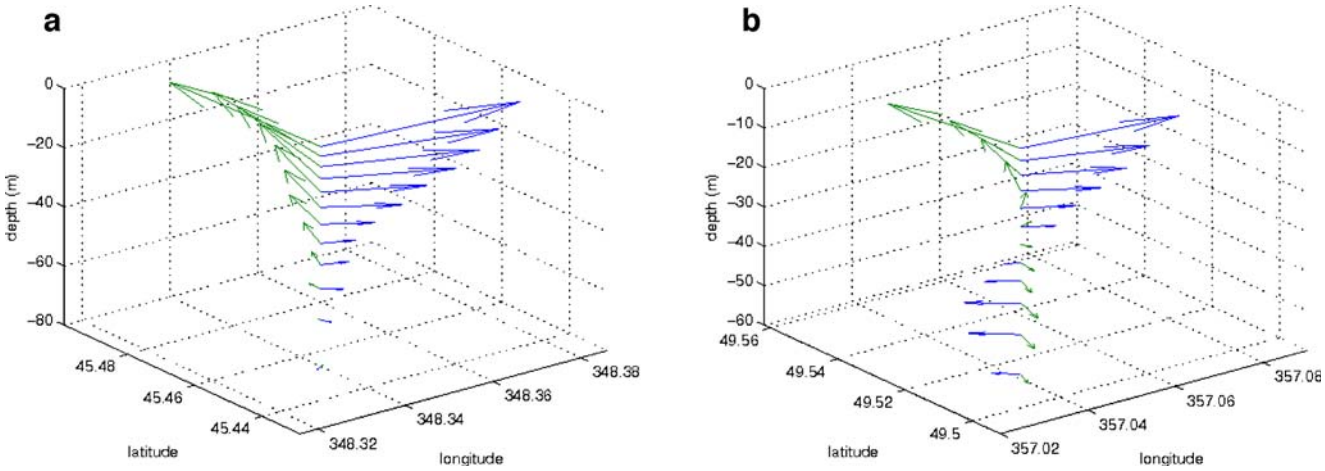

Fig. 13 Vertical profiles at point O of horizontal velocities of the representers r^{176} associated with the surface velocities at point O: **a** O=(11.67°W, 45.44°N); **b** O=(2.97°W, 49.5°N). *Blue vectors*: representer of the zonal velocity. *Green vectors*: representer of the meridional velocity. Vectors at the surface have a unit length

information available at the surface is inverted into subsurface corrections. In particular, it is necessary to determine which dynamical mechanisms are responsible for the main correlation patterns on which the inversion is based. These mechanisms are therefore studied by describing some of the main structures of representers of surface variables at depth. Representers indicate that the surface data (SST, SSH, SSS, or surface velocities) have a limited influence at depth so that the errors on surface fluxes do not affect the model at depths far below the mixed layer after a period of 15 days.

Figure 13 presents vertical profiles of velocity fields for the representers of surface zonal and meridional velocities at point (11.7°W, 45.4°N), in the deep part of the basin, and point (3°W, 49.5°N) in the English Channel, where the bottom depth is nearly 70 m. Velocities vanish at depths of more than 80 m in the basin. It has been systematically verified, even near the coasts, that the influence of surface velocities propagates downwards in what can be considered as a true Ekman spiral. However, the proximity of the bottom boundary and the coasts over the shelf is responsible for relatively strong counter-currents between depths of 30 and 60 m in the representers associated with surface velocities in the English Channel.

Figure 14 shows the representer relative to SSH at point (11.7°W, 45.4°N) on a vertical temperature section (at longitude 7°W). The positive correction of SSH at point (11.7°W, 45.4°N) induces, on the one hand, a positive correction of SSH a long way from the

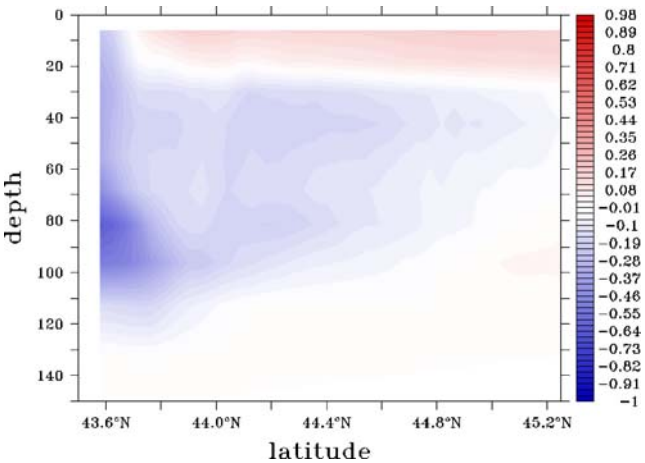

Fig. 14 Vertical cross-section at longitude 7°W of the temperature correction (in $°C$) induced by an innovation of 1 cm in SSH at (11.7°W, 45.4°N) according to the representer r^{176} associated with this observation

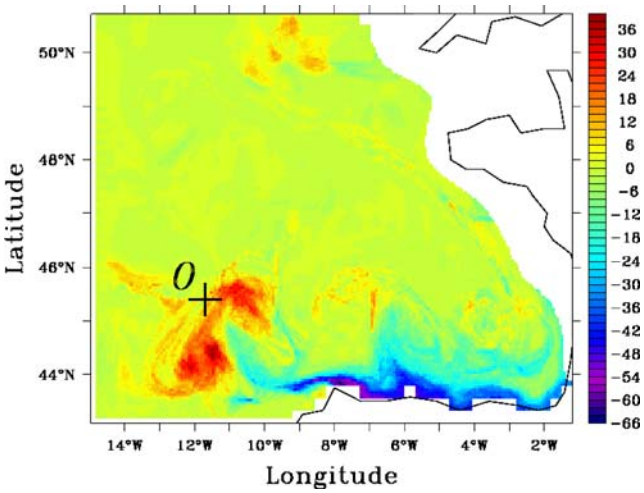

Fig. 15 Temperature field (in $°Cm^{-1}$) at 100-m depth of the representer r^{176} associated with SSH at point (11.7°W, 45.4°N)

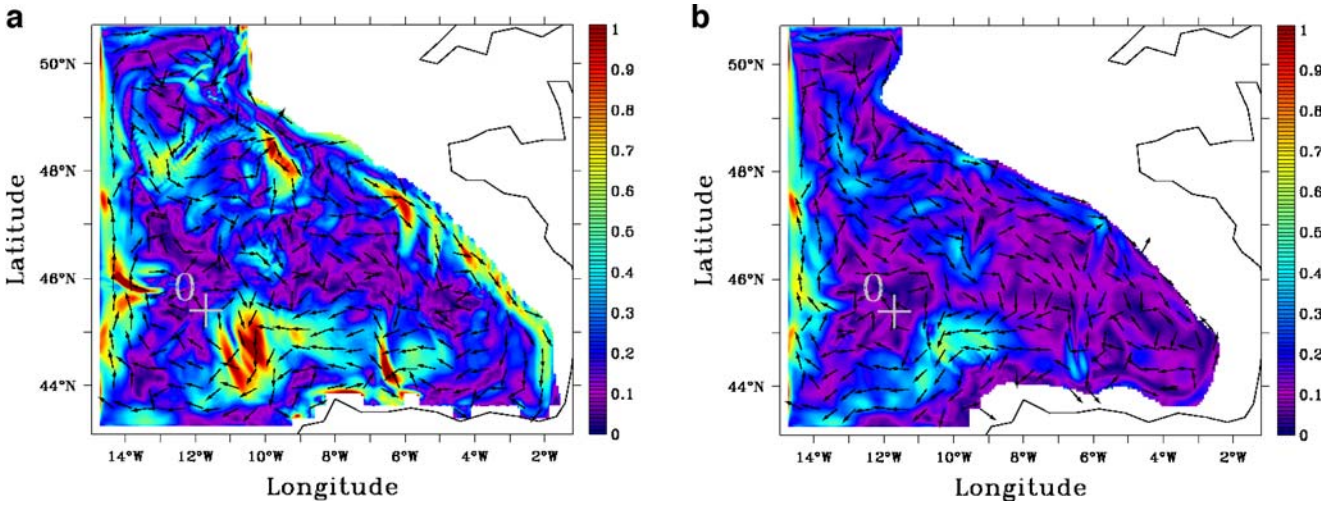

Fig. 16 Representer r^{176} associated with SSH at point (11.7°W, 45.4°N): module (in s^{-1}) and direction (every 8th vector is displayed) of total velocity at **a** 155 m and **b** 1,000 m

coasts in the basin and, on the other, a decrease in the surface height from offshore to the coasts, as indicated in Fig. 12. This induces an upwelling phenomenon in the corrections, with water masses rising beneath the mixed layer near the North of Spain. This can be seen in Fig. 14, where changes in temperature are observed down to a depth of 120 m. SSH exerts no influence on temperature, however, below 120 m.

In Fig. 15, which shows the temperature correction at 100-m depth induced by the SSH representer at (11.7°W, 45.4°N), it may be observed that the perturbation in atmospheric forcings has no influence on the main part of the domain beneath the mixed layer, but that there are horizontally restricted zones in deep layers where it appears to be significant. The direct effect of the descent of the isopycnals along the vertical of the increase observed in SSH is, in fact, slightly offset towards the East and the South. This is another consequence of the strong anisotropy of phenomena in coastal regions. The effect of upwelling near the coasts of Northern Spain is also observed in Fig. 15.

Figure 16a, b shows the impact on total velocities of this increase in SSH down to a depth of 1,000 m. The adjustment of velocities at 1,000 m is compatible with the data for the correction in SSH given in Fig. 12, where the main streamlines follow the main slopes in SSH. This is due to the decoupling of the processing of barotropic and baroclinic velocities in the HYCOM model and, thus, to the only correct adjustment of the barotropic velocities with SSH variations. Although the influence of SSH on baroclinic velocities can be observed at a depth of 155 m – where, for example, velocities are adjusted to the rise of isopycnals cen-

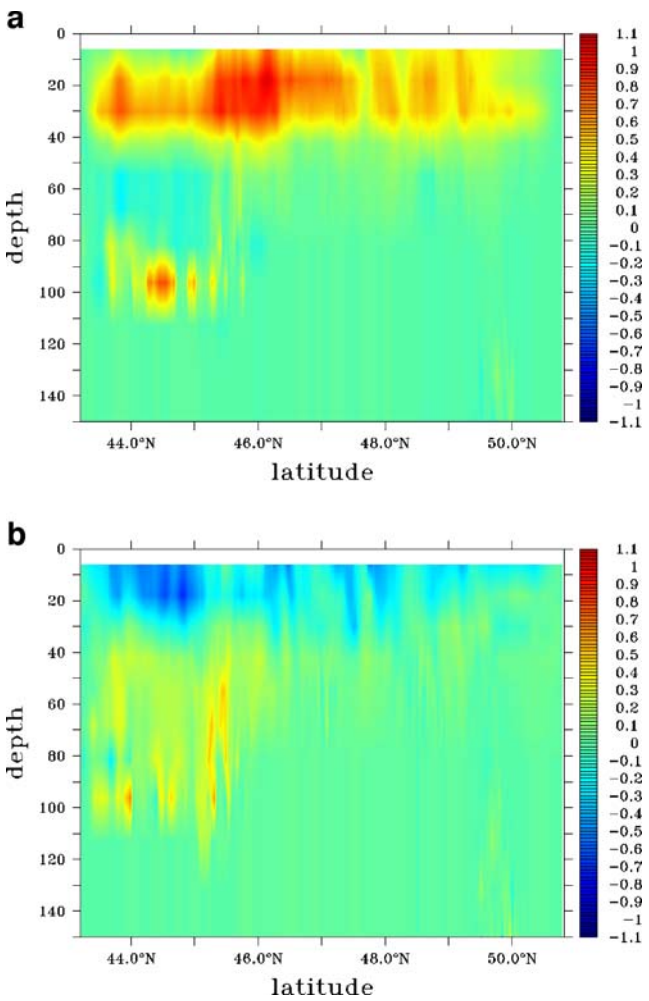

Fig. 17 Vertical cross-sections of temperature fields at longitude 11.7°W on representers r^{176} of depth temperature data at (11.7°W, 45.4°N): **a** representer of temperature at 30 m, **b** representer of temperature at 55 m

tered near (10.5°W, 44.5°N) as indicated on Fig. 15 – it appears almost nonexistent at 1,000 m.

6.5 Representers of variables at depth

The representers associated with data at depth reflect that the influence of the latter does not extend deeper than that of the surface observations. Figure 17a, b shows the temperature on the representers associated with the temperature at point (11.7°W, 45.4°N) at respective depths of 30 and 55 m. The temperature observation at 30 m does not influence the model at a depth of more than 50 m. At 55 m, the error covariance calculations no longer allow the observed temperature to be taken into account in a realistic way. The warming induced by a positive temperature correction at 55 m can be assumed to be linked to the increase in depth of the mixed layer, which is one of the main potential effects of wind stress perturbation (cf. Section 5), and thus induces a cooling of the mixed layer itself. This would appear to confirm that errors in wind stress data have the most critical influence on the structure of the model error due to atmospheric forcings. But on the representer, the warming of deep waters is not visible in the vicinity of the associated temperature observation.

7 Data assimilation experiment

In this section, a data assimilation experiment, performed with the new statistical estimation for **Q**, is described. It stresses the benefit of the ensemble statistics by comparison with a more traditional approach, based on the dominant EOFs of the model variability proposed by Pham et al. (1998). The underlying assumption of this traditional method is that the model is sufficiently good for its intrinsic variability to be statistically representative of real ocean variability. When adapting the SEEK filter to HYCOM, the hypothesis of a stationary error covariance matrix is assumed, as was done by Birol et al. (2004) in the North Atlantic configuration NA3. Both approaches provide the estimation of this stationary forecast error covariance matrix, with the assumption of Eq. 4.

To test the impact of the ensemble parameterization, twin data assimilation experiments have been performed with the SEEK filter, as follows: two simulations from the fourth subset of Monte Carlo simulations (initialized on August 22) have been integrated on 60 days. They only differ by their atmospheric forcings. The simulation using atmospheric parameters from 1994 is considered as the "true" ocean. Pseudo-

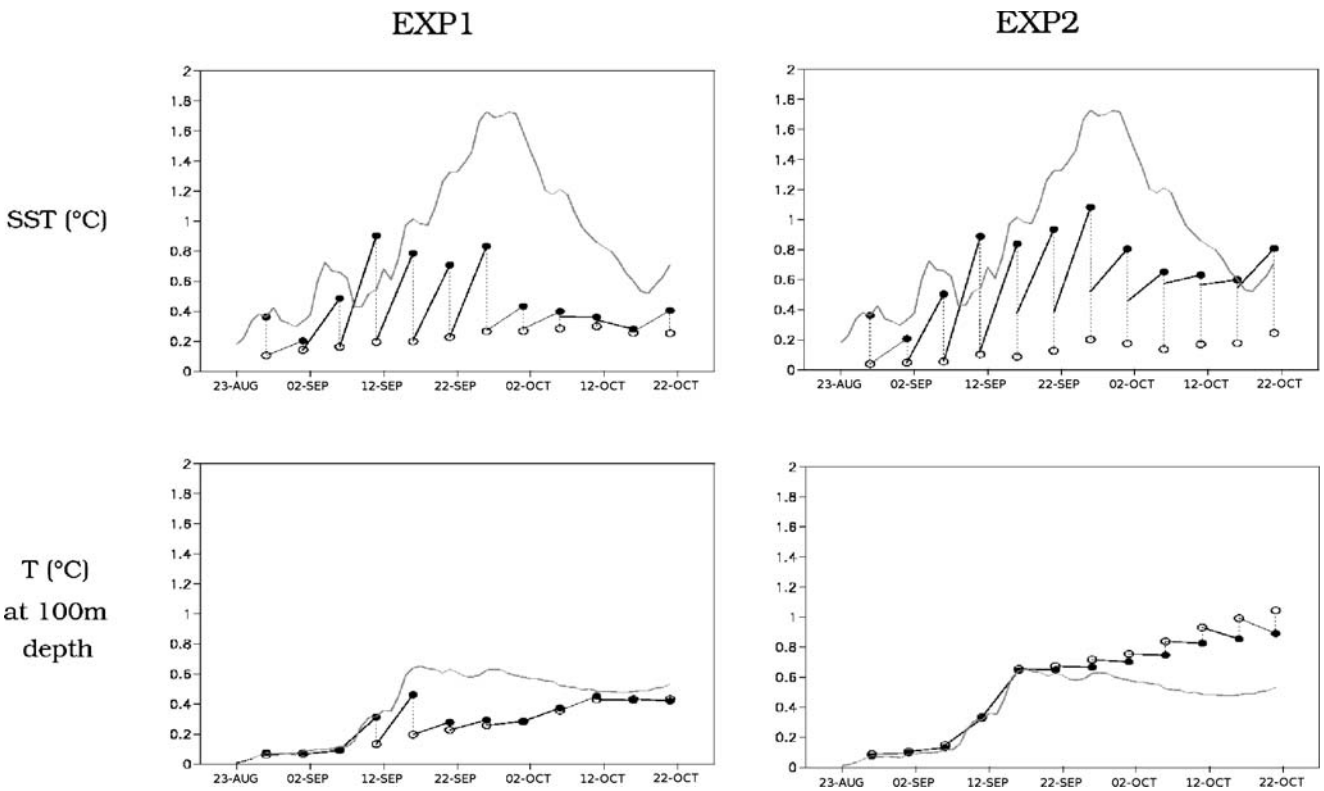

Fig. 18 RMS misfits to the "true" ocean on temperature for EXP1 and EXP2: on forecasts before analysis stage (*black bullets*) and on analysis before the application of adjustment operators (*circles*). *Solid lines* join initial conditions after analysis stages and adjustment operation to forecasts before next analysis stage. The free simulation misfits are given by the *continuous light curve*

observations of SST, covering the whole model domain, are derived from this simulation. The second simulation, using atmospheric parameters from 1985, is the "free simulation." Assimilation of SST is then applied to the free simulation to reduce its error from the "true" ocean. This assimilation of SST is realized with the SEEK filter based on Eq. 4, on assimilation cycles of length $t_{KF} = 5d$, and on the two different parameterizations of the model error described above. The model error covariance matrix estimated from ensemble of 176 Monte Carlo simulations is noted $\mathbf{Q}_{atmos}$. The one estimated with the traditional method of the SEEK filter, using the temporal variability of the free simulation, is noted $\mathbf{Q}_{variab}$. It is computed as follows:

$$\mathbf{Q}_{variab} = \frac{1}{60}\sum_{i=0}^{60}\left(\mathbf{x}_i^{free} - \overline{\mathbf{x}^{free}}\right)\left(\mathbf{x}_i^{free} - \overline{\mathbf{x}^{free}}\right)^T \qquad (11)$$

where $\{\mathbf{x}_i^{free}, i \in [0, 60]\}$ represent the 61 daily outputs of the free simulation during the data assimilation period, and where $\overline{\mathbf{x}^{free}} = \frac{1}{61}\sum_{i=0}^{60}\mathbf{x}_i^{free}$.

An order reduction is applied in the assimilation scheme, and only the 60 dominant modes of $\mathbf{Q}_{atmos}$ and $\mathbf{Q}_{variab}$ are considered. The simulation with data assimilation using $\mathbf{Q}_{atmos}$ is noted EXP1 and the one using $\mathbf{Q}_{variab}$ is noted EXP2.

The estimation vector is reduced to interface pressures, temperature, and salinity on each of the 26 layers of the model. The observation error covariance matrix $\mathbf{R}$ is assumed to be diagonal. As specific constraints (as the hydrostatic stability) have to be verified by model initial conditions, which is not guaranteed by the statistical estimation, adjustment operators are applied (Brankart et al. 2003) to restore them after an analysis stage.

The results of the assimilation experiments are illustrated in Figs. 18 and 19. Figure 18 shows the evolution of horizontally averaged RMS misfits on temperature of the simulations with respect to the "true" ocean. In both EXP1 and EXP2, the misfits on SST have been significantly reduced during the analysis phase. However, at depth, EXP1 provides more reliable results than EXP2, showing that the vertical structure of the error is better represented in $\mathbf{Q}_{atmos}$ than in $\mathbf{Q}_{variab}$. This is also apparent on Fig. 19, showing a vertical section of temperature misfits for the free simulation, EXP1 and EXP2. As explained above, the error due to atmospheric forcings concentrates within the first 100 m below the surface. In EXP2, inappropriate corrections are applied at depth because of inappropriate error structure in $\mathbf{Q}_{variab}$. Hence, a valid parameterization of

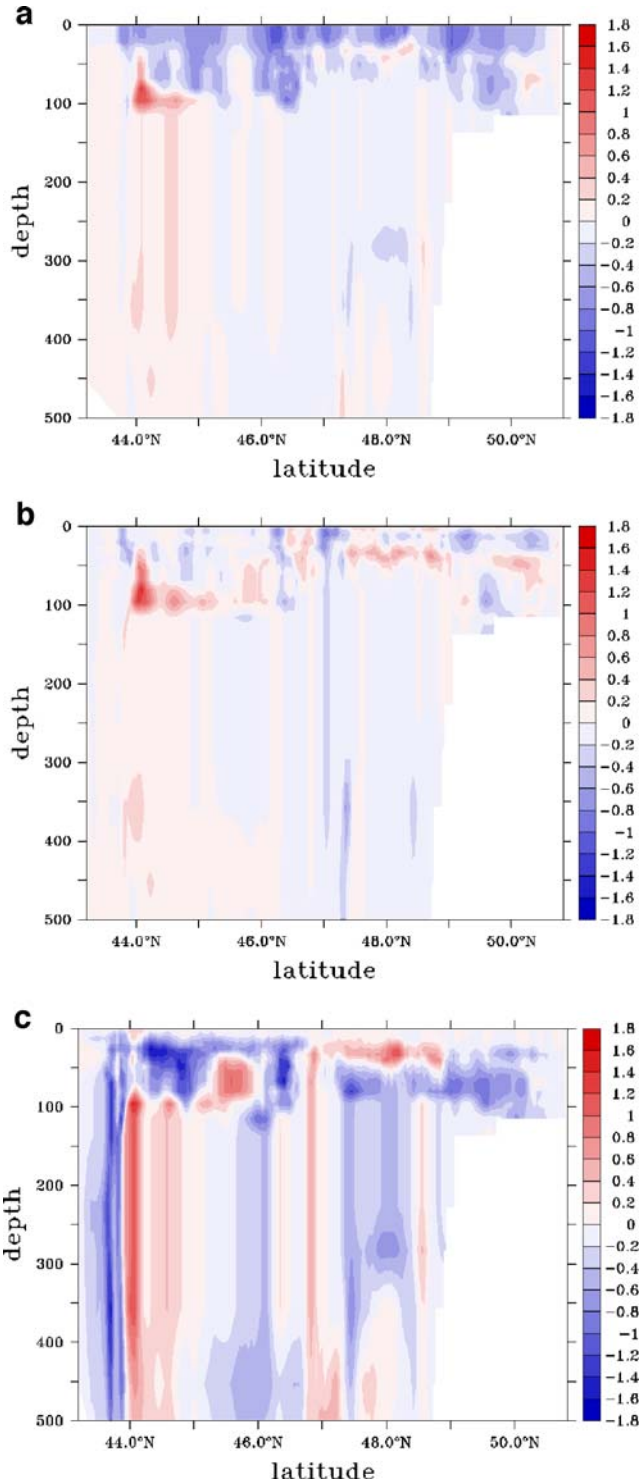

Fig. 19 Vertical section at longitude 10°W of misfits to the "true" ocean on temperature after 60 days for: **a** the free simulation, **b** EXP1 (after analysis), **c** EXP2 (after analysis)

$\mathbf{Q}$ is shown to be a key component of any assimilation system, strongly influencing the propagation of the information from the observed part of the state vector to the unobserved zone (here, even close to the surface).

Moreover, Fig. 18 points out that the modifications of the analysis states by the adjustment operators are much stronger in EXP2 than in EXP1 (the impact of adjustment operators being symbolized by gaps between circles and the beginning of the solid lines). This is symptomatic of the better physical consistency of the analysis states in EXP1. This is also part of the reason why the quality of the forecast is poorer in EXP2 than in EXP1, although the analyses of EXP2 are closer to the SST observations. All these facts converge to show that the parameterization of the forecast error in $\mathbf{Q}_{atmos}$ is more appropriate than that in $\mathbf{Q}_{variab}$.

8 Conclusion

A protocol based on Monte Carlo simulations has been designed to calculate a realistic model error induced by imperfect atmospheric forcings in a regional model of the Bay of Biscay, during the summer period. The results of the numerical experiments have been analyzed using the representer technique to assess the proposed method. The coastal model used here is shown to be significantly influenced by atmospheric parameter variations. The ensemble of simulations revealed the existence of a nonlinear rectification process linked to the estimation of mean wind stress data. This problem was responsible for the deviation of the central forecast from the mean of the model simulations. Such divergence in a central forecast – which, a priori, is considered as the most probable – is usual, particularly in high-resolution coastal models where nonlinearities in physical processes are intensified. This underlines the usefulness of ensemble methods.

The covariance of the error and the corrections induced through its use in data assimilation methods were described with the help of representer snapshots. These representers revealed that their major structures are coherent with the main potential responses of the model to variations in atmospheric forcings. This physical consistency indicates that the error estimation is realistic and should provide a good parameterization of reduced-order Kalman filters for data assimilation in such a model. This parameterization would be coherent with the hypothesis assumed here that open boundary conditions are well posed, given that the corrections induced by the parameterization do not modify the data that are strongly constrained by these conditions. The advantage of computing ensemble experiments is also underlined by the anisotropy and the inhomogeneity of the error structures that representers reveal in our model, mainly due to its coastal characteristics, in response to an error in the atmospheric parameters, which are much more isotropic and homogeneous. Such complex features prevent us, a priori, from realistically defining errors using analytical models as in optimal interpolation (OI) approaches. While such estimations of the model error should be conducted for other seasons, these main conclusions have no reason a priori to be linked to a specific behaviour of the system in summer.

Twin data assimilation experiments have been carried out to evaluate the impact of this new approach for estimating the model error. The results show a clear benefit of the assimilation scheme with respect to a more traditional approach, used for data assimilation in basin configurations. The covariances estimated from the classical method were inappropriate here and highlight the role of the model error parameterization for improving the assimilation scheme efficiency and increasing its reliability.

Systems derived from the Kalman filter and parameterized with the help of the model error estimation method developed here must now be adapted for realistic data assimilation experiments in coastal models. The question of determining the best estimate (Evensen 1994) is addressed by examining the divergence between the mean and the central forecasts of our Monte Carlo simulations. The limited availability of surface data does not, in principle, overly restrict the capacity of data assimilation systems parameterized with model errors related to atmospheric forcings in summer. This is because of the considerable horizontal extent of the influence of these data in the mixed layer, which is the essential zone to be controlled in coastal models. The error in the atmospheric forcings, however, does not seem to influence the very deep zones of the model at the considered time scales. Consequently, in situ measurements taken well below the region directly affected by the atmospheric forcings will not be useful to control the corresponding model error. Assimilation experiments are now needed to validate these adaptations of the assimilation system for controlling the parameters and state of the coastal model. This should be followed by the estimation of a more realistic model error through the characterization (using the same ensemble method) and the combination of errors due to uncertainties in other model parameterizations.

Acknowledgements This work was supported by the Délégation Générale pour l'Armement (DGA), by the Service Hydrographique et Océanographique de la Marine (SHOM), by the Centre National de Recherche Scientifique (CNRS), and by the MERSEA project of the European Commission. We would

like to thank Eric Chassignet (Rosenstiel School of Marine and Atmospheric Science) for providing the North Atlantic model configuration. The calculations were conducted with the support of the Institut du Développement et des Ressources en Informatique Scientifique/CNRS.

References

Auclair F, Marsaleix P, De Mey P (2003) Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions. Dyn Atmos Ocean 36:309–346

Bennett AF (1992) Inverse methods in physical oceanography. Cambridge University Press, Cambridge, pp 346

Bennett AF (2002) Inverse modeling of the ocean and atmosphere. Cambridge University Press, Cambridge, pp 234

Birol F, Brankart JM, Castruccio F, Brasseur P, Verron J (2004) Impact of ocean mean dynamic topography on satellite data assimilation. Mar Geod 27:59–78

Bleck R (2002) An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model 37: 55–88

Brankart JM, Testut CE, Brasseur P, Verron J (2003) Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: application to a 1993–1996 hindcast of the North Atlantic Ocean circulation. J Geophys Res 108:C33074, doi:10.1029/2001JC001198

Brasseur P, Verron J (2006) The SEEK filter method for data assimilation in oceanography. Ocean Dyn 56:650–661, doi:10.1007/s10236-006-0080-3

Browning GL, Kreiss HO (1982) Initialization of the shallow water equations with open boundaries by the bounded derivative method. Tellus 34:334–351

Chassignet EP, Smith LT, Halliwell GR, Bleck R (2003) North Atlantic simulation with the Hybrid Coordinate Ocean Model (HYCOM): impact of the vertical coordinate choice, reference density, and thermobaricity. J Phys Oceanogr 33:2504–2526

Chelton DB, Schlax MG, Freilich MH, Milliff RF (2004) Satellite measurements reveal persistent small-scale features in ocean winds. Science 303:978–983

Davies HC (1976) A lateral boundary formulation for multi-level prediction models. Q J R Meteorol Soc 102:405–418

Echevin V, De Mey P, Evensen G (2000) Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model. J Phys Oceanogr 30:2627–2635

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):10143–10162

Ide K, Courtier P, Ghil M, Lorenc AC (1997) Unified notation for data assimilation: operational, sequential and variational. J Meteorol Soc Jpn 75(1B):181–189

Kara AB, Rochford PA, Hurlburt HE (2000) Efficient and accurate bulk parameterizations of air-sea fluxes for use in general circulation models. J Atmos Ocean Technol 17: 1421–1438

Koracin D, Dorman CE, Dever EP (2004) Coastal perturbations of marine-layer winds, wind stress, and wind stress curl along California and Baja California in June 1999. J Phys Oceanogr 34:1152–1173

Kurapov AL, Allen JS, Egbert GD, Miller RN, Kosro PM, Levine M, Boyd T (2005) Distant effect of assimilation of moored currents into a model of coastal wind-driven circulation off Oregon. J Geophys Res 110:C02022, doi:10.1029/2003JC002195

Large WG, McWilliams JC, Doney SC (1994) Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev Geophys 32:363–403

Miller RN, Cane MA (1989) A Kalman filter analysis of sea level heights in the tropical Pacific. J Phys Oceanogr 19: 773–790

Oliger J, Sundström A (1978) Theoretical and practical aspects of some initial-boundary value problems in fluid dynamics. SIAM J Appl Math 35:419–446

Oke PR, Allen JS, Miller RN, Egbert GD, Kosro PM (2002) Assimilation of surface velocity data into a primitive equation coastal ocean model. J Geophys Res 107:C93122, doi:10.1029/2000JC000511

Pham DT, Verron J, Roubaud MC (1998) Singular evolutive extended Kalman filter with EOF initialization for data assimilation in oceanography. J Mar Syst 16:323–340