

# Multivariate assimilation in MERCATOR project: New statistical parameters from forecast error estimation

H. Etienne\*, M. Benkiran

*CLS 8-10 rue Hermès 31526 Ramonville Saint Agne, France*

Received 24 September 2004; accepted 21 November 2005

Available online 21 November 2006

## Abstract

The new operational prototype of Mercator (french Global Ocean Data Assimilation Experiment contribution) is composed of a North Atlantic primitive equation ocean model OPA (Ocean Parallel Algorithm between 20°S and 70°N, [Madec, G., P. Delecluse, M. Imbard and C. Lévy (1998). OPA8.1 ocean general circulation model reference manuel. Notes du pôle de modélisation IPSL. n°11: 91p]) and of a multivariate and multidata assimilation scheme [De Mey, P. and M. Benkiran (2002). “A multivariate reduced-order optimal interpolation method and its application in Mediterranean basin-scale circulation.” Ocean Forecasting : Conceptual basis and application, Pinardi, N., Springer Verlag.] This system has already given some significant improvements from previous Mercator configurations (M. Benkiran, personal communication). However some biases on ocean state still remain in the tropics where the reduced-order optimal interpolation scheme is suspected to be ill-parameted in the model forecast error. Indeed the guess error covariance matrix is decomposed into an error variance value and a spatio-temporal correlation function which are assumed to have some “good” properties (spatial homogeneity of the correlation function, constant ratio between signal and error variance). This study shows how we can use ensemble methods to validate these assumptions. We can see that the correlation function can reach negative values locally, mostly in regions of high variability contradictory with the homogeneous hypothesis. The reduced space used in the operational configuration is based on the signal seasonal Empirical Orthogonal Functions (EOFs). An empirical relationship between signal and error variance has been set and the correlation function is the same on every dimension of the reduced space. By projection of the estimated guess error variance onto the reduced space, we find a repartition of this quantity quite different to what was set in the system. The error statistics is found to be inhomogeneous compared to hypothesis made in the assimilation scheme. These two new parameters tested separately in the assimilation scheme gives significant improvements of the forecast and analysis results. This is particularly obvious in the tropics. But relationship between signal and error statistics (as assumed in the optimal interpolation) is found to be complex.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Multivariate assimilation; Optimal interpolation; Ensemble experiments; Error covariance matrix; Operational oceanography; Atlantic Ocean; Reduced space

## 1. Introduction

Mercator, the french contribution to GODAE (Global Ocean Data Assimilation Experiment), is now providing routine forecasts of the ocean state.

\* Corresponding author. Tel.: +33 561 393 829; fax: +33 561 393 899.

*E-mail addresses:* [helene.etienne@mercator-ocean.fr](mailto:helene.etienne@mercator-ocean.fr) (H. Etienne), [mounir.benkiran@mercator-ocean.fr](mailto:mounir.benkiran@mercator-ocean.fr) (M. Benkiran).

The final goal of this project is to supply high resolution global ocean model outputs. Since 2001, different operational systems have been developed towards the ultimate configuration. It has begun with a low resolution model of the Northern Atlantic (OPA 1/3 degree, Madec et al., 1998) combined with reduced order optimal interpolation scheme (De Mey and Benkiran, 2002) assimilating along track satellite Sea Level Anomaly (SLA). The second phase of the prototype development was based on an increase of the model resolution and an extension of the domain to Mediterranean Sea. The last stage was to implement a fully multivariate and multidata assimilation scheme, first in the low resolution, then in the high resolution operational system. The third prototype of Mercator is now available and provides 7- and 14-day forecast of the North Atlantic Ocean, assimilating satellite and in-situ data. The new assimilation scheme, described in Benkiran et al. shows how the use of in situ data improves both the assimilation statistics and the ocean state forecast. The Mercator configuration studied hereafter extends over the Atlantic Ocean from 20°S to 70°N. The horizontal resolution is 1/3 degree and there are 43 levels on the vertical. The reduced order optimal interpolation scheme used is SOFA (System for Ocean Forecasting and Analysis) developed by P. De Mey (De Mey and Benkiran, 2002). It is based on the Kalman filter (Kalman, 1960) equations using Optimal Interpolation (OI) formulation. The assimilation is sequential and provided every week. This scheme is a fully multivariate and multidata configuration. It computes  $\Psi$ ,  $T$ ,  $S$  analysis increments (where  $\Psi$  is the barotropic part of stream function,  $T$  is temperature and  $S$  is salinity) in a reduced space by combination of model state with satellite along track SLA,  $T$  and  $S$  in situ profiles and Sea Surface Temperature (SST) from Reynolds data base.

Nevertheless, this method is based on a number of statistical hypotheses related to forecast, system or observational errors. A special effort has to be made to solve the problem of the forecast error covariance matrix representation. This matrix, called  $\mathbf{B}$ , is usually needed in most of the assimilation techniques to compute the correction of the forecast state.

Many parameterizations have been developed to express the forecast error statistics. In linear system,  $\mathbf{B}$  is propagated by the model, as in the classical Kalman filter equations (Kalman, 1960). But for non-linear system, it leads to some numerical problems linked to closure of the time propagation of the density probability function of the error (Evensen, 1994). The optimal interpolation scheme is one solution to repre-

sent  $\mathbf{B}$ . It has been useful in many ocean configurations (Gavart and De Mey, 1997; Gavart et al., 1999; Maes et al., 1999). Therefore, we will first focus on the guess error covariance matrix as formulated in the multivariate scheme. This will draw attention to the two major hypotheses made on forecast error statistics and how they can be tested by ensemble methods. We propose a better parameterization of the guess error covariance matrix that can dramatically improve the model results. Ensemble experiments are used in meteorology to evaluate model predictability (Murphy, 1988; Mureau et al., 1993; Toth and Kalnay, 1993). These techniques are used in this paper to answer to a number of questions concerning the error statistics (Section 3). They allowed computation and test of a new correlation function and error variance on the assimilation and forecast results, as presented in Sections 4 and 5 of this document.

## 2. Forecast error in the multivariate assimilation scheme

A general overview of the assimilation scheme can be found in Benkiran et al. In the present paper, we only recall the parameterization of the forecast error covariance matrix used in the optimal interpolation scheme. The analysis  $X^a$  is obtained by correction of the forecast state  $X^f$  using misfits between  $X^f$  and observations  $y_o$ . Observation data sets are temperature and salinity profiles, SLA, daily Sea Surface Temperature (SST) from Reynolds at the analysis date. The correction is computed in a reduced space via the  $S$  operator. The reduced space is constructed with the ten first ( $\Psi$ ,  $T$ ,  $S$ ) EOFs. So  $S$  contains the ten first eigenvectors of the ( $\Psi$ ,  $T$ ,  $S$ ) covariance matrix, stored by decreasing eigenvalues. Each of these modes stands for a part of the signal variance, equal to the corresponding eigenvalue. Ten are enough here to explain ninety percent of the signal variance. Then, using the Kalman equations we can write:

$$X^a = X^f + K(y_o - HX^f) \quad (1)$$

$$K = S^{-1}K_r \text{ and } K_r = SK \quad (2)$$

$$K_r = B_r H_r^T (H_r B_r H_r^T + R)^{-1} \quad (3)$$

The low case  $r$  is for matrices in the reduced space,  $H$  for the observational operator with  $H_r = HS^T$ .  $K$  is the Kalman gain and  $\mathbf{R}$  is the observation error covariance matrix.  $\mathbf{B}_r = S^T \mathbf{B} S$  is the reduced guess error covariance

matrix, which is expressed in the optimal interpolation formulation by:

$$\mathbf{B}_r = D_r^{1/2} C D_r^{1/2} \quad (4)$$

where  $D_r$  is the guess error variance and  $C$  the correlation function. This last formulation means that we can decompose the guess error covariance into its variance part (the amplitude of the covariance) and its spatial and temporal structure  $C$  in every dimension of the signal subspace (EOFs). We will now discuss these two parameters.

### 2.1. The correlation function $C$

In each of the Mercator operational configuration,  $C$  can be described by some spatial and temporal correlation length scales. It's an exponential space and time decreasing function, with positive amplitude depending on the distance. The correlation is nearly equal to 0.5 for distances equal to the spatial correlation length  $R_x$  and  $R_y$  (respectively zonal and meridional radius), and reaches zero by positive value for larger distances.  $R_x$ ,  $R_y$  and also  $R_t$  (time correlation radius) are estimated from analysis increments of a previous simulation (in this simulation, analysis is performed using length scales from SLA altimetry to parameterized  $C$ ). Even if resulting radii are space-varying, the structure of  $C$  is kept unchanged over the whole domain. Notice that same radii are used on every dimension of the reduce space (on the ten modes).

First, considering the projection of the  $\mathbf{B}$  matrix on the signal EOFs (leading to  $\mathbf{B}_r$ ), we assume there that the guess error can be estimated on the same base as the signal. Now, with this  $C$  function, we have made the hypothesis that the spatial and temporal structures of the error are similar on every dimension of the signal reduced space. We know that most of the large scale can be found on mode 1 (barotropic mode), whereas the mesoscale patterns are typically represented by higher order modes. Moreover it has shown a strong correlation between error structures and dynamics (Moore, 1999; Kuragano and Kamachi, 2000). So to which extent can we make this assumption of a unique  $C$  function and set of correlation length scale over each reduce space dimension? This is the first question that we will try to solve in next sections.

### 2.2. The guess error variance $D$

In our context,  $D_r$  is strongly correlated to the definition of  $S$ . In fact,  $D_r$  depends both on the mode

and on the grid point. EOFs are computed seasonally. Four sets of empirical modes (one every 3 months) are then available to characterize the ( $\Psi$ ,  $T$ ,  $S$ ) ocean state seasonal variation. Let us call  $V(m)$  the signal variance explained by the mode  $m$  ( $m^{\text{th}}$  eigenvalue). The guess error variance  $D_r$  on each mode is defined as a fraction of the mode variance:

$$D_r^{1/2}(m) = 0.2V^{1/2}(m) \quad (5)$$

Value of 0.2 has been chosen empirically. In that case we still assume the connection between signal and error standard deviation (hereafter SD): we impose the same EOFs decomposition, spatial repartition and the same seasonal variation. We know that dynamics and error growth are strongly correlated (Etienne and Dombrowsky, 2003). Therefore, the second question to rise is, to what extent will the parameterization of the guess error variance of Eq. (5) be accurate enough to make the assimilation efficient?

### 2.3. Methodology

To estimate guess error covariance, we choose the ensemble methods, as in Etienne and Dombrowsky (2003). We have tried to estimate the initial condition error covariance (therefore the analysis error covariance while analysis of cycle  $i$  is the initial condition for cycle  $i+1$ , see Eq. (1)). Consider a given model  $M$  with initial state  $X^a(t_0)$  and associated initial error statistics, for example error covariance  $C_{ov}(t_0)$ . This statistic can be computed by comparison of  $X^a(t_0)$  with the true state  $X^t(t_0)$ :

$$\begin{aligned} C_{ov}(t_0) &= \overline{[X^a(t_0) - X^t(t_0)]^T [X^a(t_0) - X^t(t_0)]} \\ &= \varepsilon(t_0)^T \varepsilon(t_0) \end{aligned} \quad (6)$$

These matrix is then propagated by the model:

$$\begin{aligned} C_{ov}(t_n) &= \overline{[X^t(t_n) - X^a(t_n)]^T [X^t(t_n) - X^a(t_n)]} \\ &= \overline{[M(t_0, t_n)X^a(t_0) - X^t(t_n)]^T [M(t_0, t_n)X^a(t_0) - X^t(t_n)]} \end{aligned} \quad (7)$$

The overbare stands for the statistical mean. This last formulation shows that we can compute error statistics at every model time step. This supposes we can access the true state and this is obviously not realistic. First because this imply a huge quantity of observations (proportional to the model resolution and coverage) and second because the measurements are suppose to be perfect (no observational error). Ensemble methods are

an efficient way to estimate statistics. The true state  $X^t$  is replaced by the most probable state  $\bar{X}$ , given an initial error distribution. This means that we can find a cloud of  $N$  states  $X_p(t_0)$  around  $X^a(t_0)$  such as

$$X_p(t_0) = X^a(t_0) + \varepsilon_p(t_0), \quad p \in [1, N] \quad (8)$$

With  $E[\varepsilon(t_0)] = 0$  and  $\overline{[\varepsilon(t_0)^T \varepsilon(t_0)]} = C_{ov}(t_0)$ ,  
 $\varepsilon = \{\varepsilon_p\}$ ,  $p \in [1, N]$

Then, from Eq. (7), we can write:

$$C_{ov}(t_n) = \overline{[X^f(t_n) - \bar{X}^f(t_n)]^T [X^f(t_n) - \bar{X}^f(t_n)]} \quad (9)$$

In practical, we need information about the initial statistics we try to estimate, such as variance range, spatial and temporal covariance patterns. Evensen (1994) proposes a method to compute such an ensemble of perturbations  $\varepsilon_p$ , based on inverse Fourier transform (see appendix of Evensen, 1994).

Nevertheless we need a first guess for our initial ensemble statistics. One direct available estimation is the statistics on the assimilation increment from the univariate MERCATOR scheme, assimilating only SLA and using the same OPA model configuration. The

assimilation scheme is based on the lifting–lowering method (Cooper and Haines, 1996) to propagate the information on the full model state (convert SLA increment into T,S,U,V increments). Nonetheless, these statistics contain information about the guess error (distance between model state and observations), but also errors due to the assimilation scheme (possible bad estimated parameters). Using the method described in Evensen (1994), we compute an ensemble of perturbations similar to SLA increments. Each of them is propagated on the full model state to obtain  $\{\varepsilon_p\}$  as fully described in Etienne and Dombrowsky (2003). We found that 50 members were enough to correctly set the initial statistics written in Eq. (8).

Three experiments are thus conducted over different periods. Ensemble are integrated over 28 days, even if the assimilation window takes 7 days long to study error doubling time and to allow dissipation of possible noise introduced by the small scales perturbation. Periods chosen in the following are 1995 winter and summer and 1996 winter (hereafter W95, S95 and W96 respectively) as in Etienne and Dombrowsky (2003) to eventually find some similarities in the two configurations. These periods allow us to crudely point out the possible interannual and seasonal variation of the error growth. Fig. 1 shows the SLA SD (from SLA increments

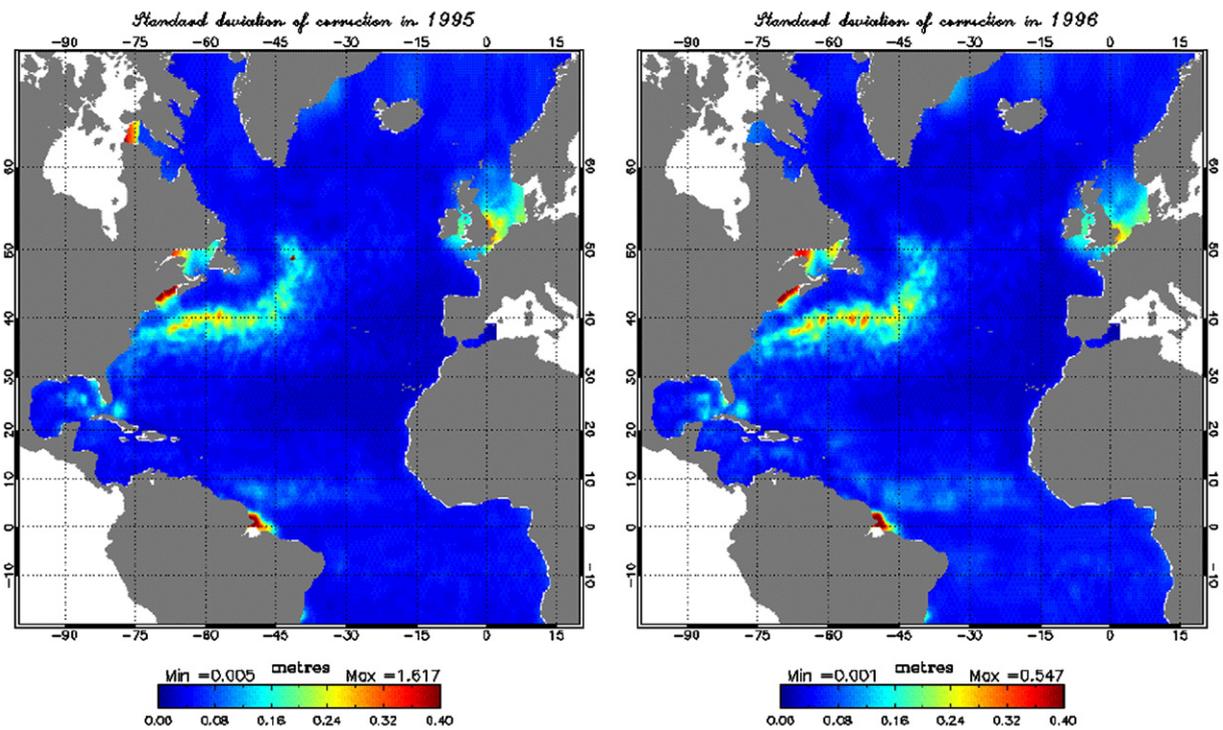


Fig. 1. SLA increment standard deviation (in m) from the univariate Mercator system over year 1995 (left plot) and year 1996 (right plot).

of the univariate scheme) used to estimate initial ensemble spread (ensemble SD) for year 1995 and 1996 corresponding to diagonal components of  $C_{ov}(t_0)$  (Eq. (9)).

### 3. Estimation of the error statistics

A result that could be expected from the 28-day ensemble integration is the strong correlation between error growth (spread or collapse of the ensemble) and the local dynamics. Even if the major patterns are similar, the three experiments lead to different 7 days error variances (time period of the assimilation window), depending on the period of integration. The example of the barotropic height is presented in Fig. 2. Whatever the experiments, we can identify strong error variance in:

- the southern part of the Gulf Stream and its extension;
- the East Greenland Current and the Labrador Sea;
- the tropical belt, except at the equator.

The equator is a particular case, because initially no error variance on barotropic height has been generated. This is mostly due to the lifting–lowering method we used to propagate the perturbation on the full state. The barotropic increment is set to zero at the equator. That is why there is no initial error variance on this variable and it is obvious that this error did not propagate from surrounding regions. The magnitude is interannually and seasonally dependant. The major changes occur in the tropics north of 5°N, with a more important barotropic height error variance in summer and especially in 1996. Depending on the model depth, the error growth characterizes the local dynamics. We obtain different error variance features for temperature, salinity and horizontal velocity. In the area of high variability (Gulf Stream, Gulf of Mexico, North Brazil Current, hereafter NBC), 7 days error patterns are concentrated in the mesoscale structures: one can see the error variance growing near meanders and strong current path.

What can be seen after 28 days is that the equator behaves like a wave guide for the error variance as it is for the signal (Steger and Carton, 1991). We can clearly see the eastward propagation of some error structures along the equator, their reflection of the African coasts and their westward return between 10°S and 10°N. In the same way, error patterns due to Mediterranean outflow near 1000 m spread to the west of the basin. The NBC and linked eddies make errors grow and carry them northward along the Brazilian coasts to the

Caribbean Sea. Therefore, the error propagation is consistent with the related signal.

#### 3.1. Error doubling time

The error doubling time for  $\Psi$ ,  $T$  and  $S$  is longer than 28 days, as can be seen on Fig. 3. But, the velocity field in the northern equator area between the surface and 100 m depth has a different behavior. In this case, doubling time of error is reached between 15 and 28 days, depending on the period and depth:

- near surface in summer, rms errors can be twice their initial value after 20 days, instead of 12 or 15 days in winter;
- at 100 m, doubling time is about 28 days in summer, but only 17 to 20 days in winter.

The northern tropical region around 10°N is then the only area where some error variance can develop and become twice its initial value, mostly in the first 100 m. We have to notice that time scales of the error growth (between 12 to 28 days) are larger than the assimilation window of the system (7 days). On the other hand, the system also provides 14 days forecasts. The error on the velocity fields may be important, especially if it is a winter forecast.

What can be suspected from Fig. 3 is that ensemble members are not completely initialized. In Etienne and Dombrowsky (2003) preliminary experiments have been made to fit and keep into the initial ensemble only spatial patterns that were consistent with the model dynamics. But we still found in some case a decrease of the error SD, showing either a bad estimation of the initial error variance magnitude or some remaining noise in the perturbation structures. This can be explained by the lifting–lowering method, which has not the same efficiency depending on the area (especially in the tropics). It means that increment distribution between barotropic and baroclinic components of the model state may not be properly ill-built, and introduces structures that would be dissipated by the model dynamics.

We now compute error statistics described above (correlation function and error variance in reduced space) from these results (see Eq. (9)). The following computation is based on the 28-day ensemble forecast, despite the fact that we are interested in the 7 days error statistics for the assimilation. This time period is necessary for the system to dissipate spatial scales that could be introduced into the ensemble initialization by our perturbation method. In the following, only one experiment is discussed, knowing that the 3 experiment results are similar.

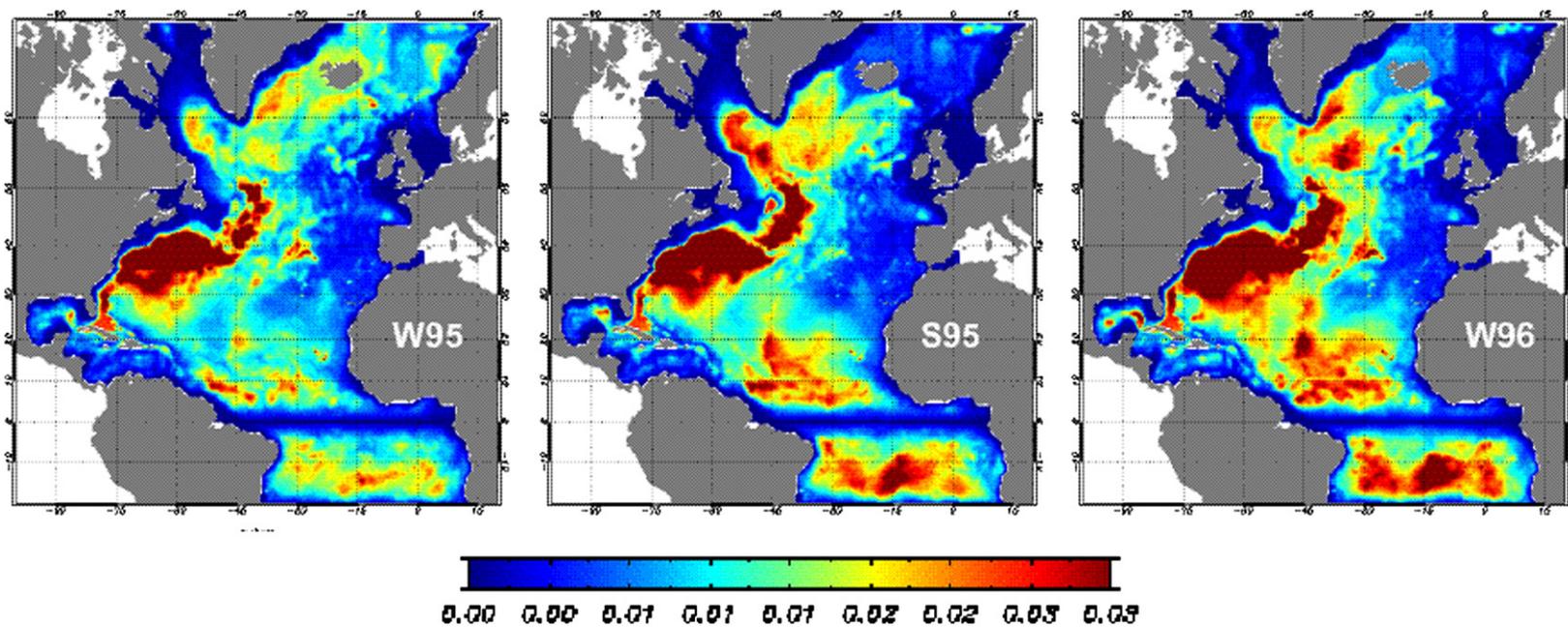


Fig. 2. Barotropic height standard deviation (in m) of the 7-day ensemble for the 3 experiments.

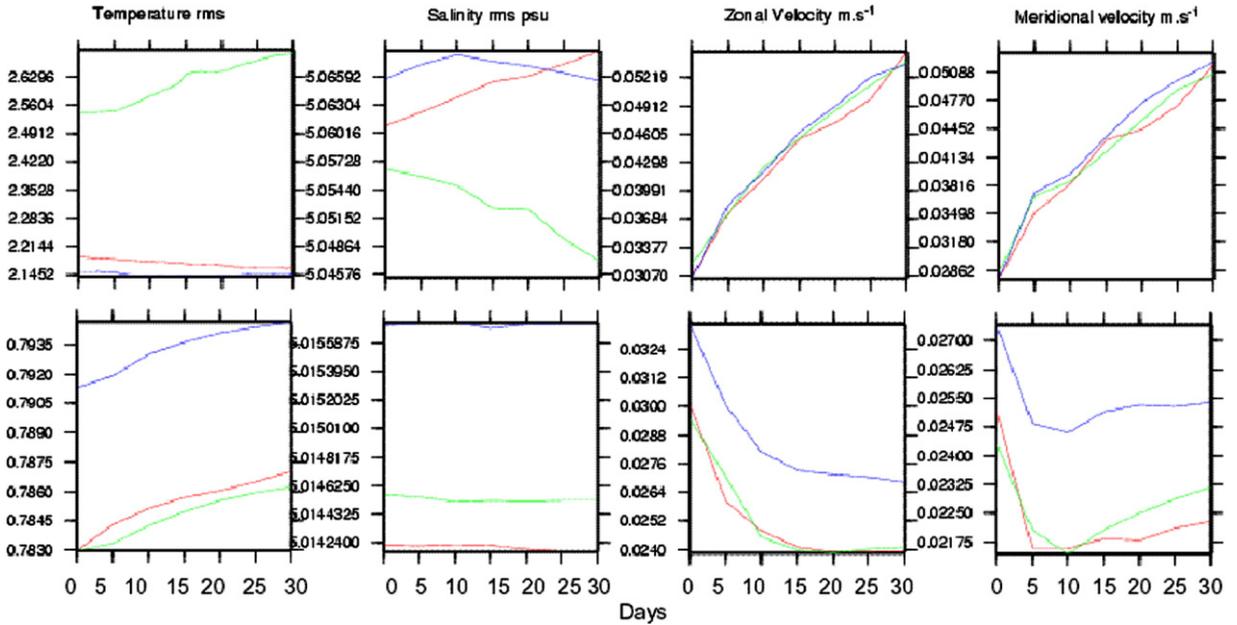


Fig. 3. Time series of the spatial mean of the error standard deviation for temperature (°C), salinity (psu), zonal and meridional velocity (m s<sup>-1</sup>) at the surface (upper plots) and at 1000 m (lower plots). Red line is W95, green line is S95 and blue line is W96.

3.2. Error correlation and variance

The most common analytical correlation functions used in oceanography are performed to represent the mesoscales (Julian and Thiébaux, 1975; De Mey and Robinson, 1987). In general, analytical correlations are computed for stream function (Arhan and Colin de Verdière, 1985) or from SLA signal (Le Traon et al., 1991). In our case, the same function has been used for both SLA statistics in the univariate configuration and in

multivariate one. But correlation radii have been recomputed to be consistent with  $\Psi$ ,  $T$  and  $S$  statistics.

Ensemble correlation of the SLA state (computed from Eq. (9)) brings us to both a spatial and temporal correlation for the initial condition error. We obtain two major patterns over the domain:

- a positive function, similar to the  $C$  function actually set in the system (Fig. 4a). We mostly find this correlation function in the northern part of the

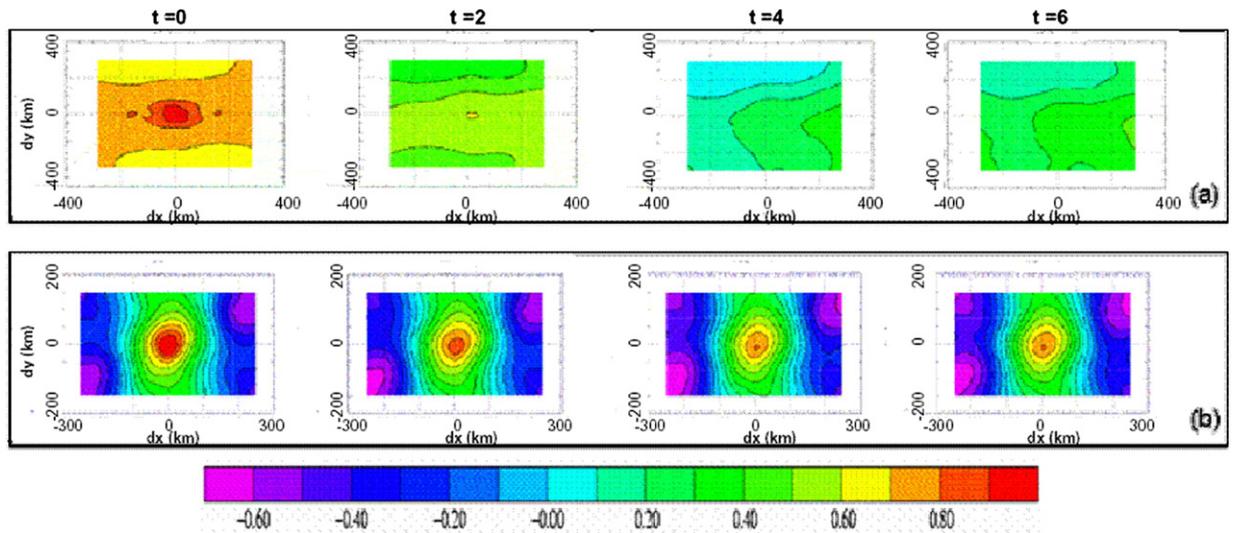


Fig. 4. Spatial correlation of SLA error in W95 at (0.66°W; 17.7°S) (a) and (31.33°W; 33.84°N) (b). Temporal distances of 0, 2, 4 and 6 days are plotted.

domain (north of 50°N), near the equator and at the southern boundary. We can identify a spatial (temporal) decrease with distance (time), but the correlation remains positive.

- A function passing through negative values (Fig. 4b). It can be found in strong current regions (Gulf Stream, Azores Current, North Atlantic Current, tropics, Caribbean Sea and Gulf of Mexico). We still have a spatial and temporal decrease, but some negative values are reached for large spatial distances.

The form of the correlation is a match for the 3 experiments, but correlation length scales seem to evolve with the period (changes in the dynamics and so in the anisotropy of the error). One analytic function exists to describe the negative correlation function (Arhan and Colin de Verdière, 1985). Hereafter, we call it  $C_m$  (Fig. 5). It has the same structure than previously, with a time and spatial exponential decrease, but its magnitude is set by a third degree polynomial (instead of second degree in  $C$ ). Whereas  $C$  is near 0.55 at the correlation length scale ( $r/r_0=1$  in Fig. 5),  $C_m$  is now equal to zero. For  $r/r_0$  between 1 and 3,  $C_m$  is negative. Then it converges to zero by positive value. The ensemble experiments carried out here show that this function is well suited to estimate SLA error correlation in most of domain, especially in dynamical regions. Hence, we can expect from this result a strong spatial inhomogeneity of the correction in some regions. Fig. 5

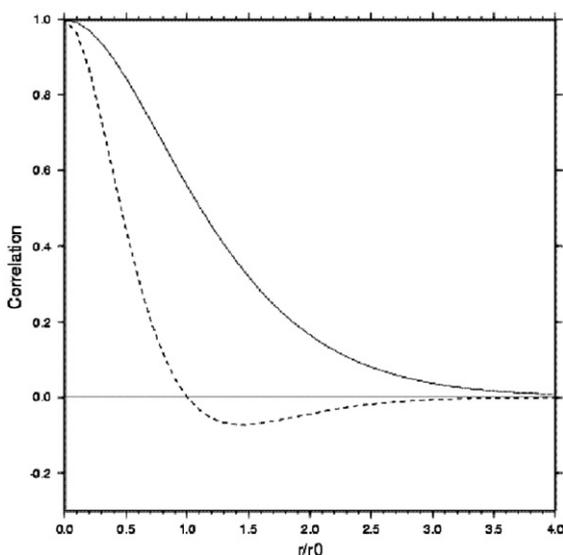


Fig. 5. Spatial part of the  $C$  function (solid line) and  $C_m$  function (dash line).

clearly shows that the impact of a data at  $r_0$  distance of a grid point is really poor considering the  $C_m$  function, whereas it is still about 50% considering the  $C$  function. That is why we may find more localized correction in strong current regions than in the other areas.

Concerning the error variance, it has to be estimated in the reduce space. That is why each ( $\Psi$ ,  $T$ ,  $S$ ) 28-day ensemble member has been projected onto the EOFs base. Then, for each mode, the ensemble spread has been computed (diagonal part of the ensemble covariance matrix, see Eq. (9)). Fig. 6 shows the logarithm of the ratio between ensemble and mode variance for mode 1, 2 and 10 obtained for the S95 experiments. In the actual assimilation scheme, this ratio is taken equal to 0.04 (Eq. (5)). This value corresponds to the white contour in Fig. 6. This state a proportionality relationship between error variance and signal variance (here in the reduced space). This hypothesis is commonly used to estimate error statistics parameters. But, we can see that this ratio is mostly below 0.04 (logarithm below  $-1.39$ , gray line), except in some area where it can be greater than 1 (positive values in Fig. 6). This parameter is then not constant over the domain. This points out that leading signal EOFs are not similar to dominant error EOFs for the whole area. Error variance in the Gulf Stream can be found on mode 2, whereas in the tropics, the modes of order greater than 7 are prevailing. We can then conclude that the signal base is not satisfactory to represent the error statistics. The assumption that the ratio between signal and error is constant in the reduced space is then mistaken.

### 3.3. Discussion

We have just seen that hypothesis made in the multivariate assimilation scheme of Mercator have to be questioned. As shown in previous section, ensemble experiments allow us to estimate error correlation and variance. Considering optimal interpolation scheme, as used in MERCATOR (see Eq. (4)) we now access parameters for error statistics needed in the assimilation scheme. First, we found that a constant correlation function over the domain seems to over estimate the homogeneity of the error structures. Roughly speaking, we found two kinds of patterns that can be correlated to bathymetry and dynamics. In region of high variability (pointed out by larger SD value in Fig. 1), we mostly find a correlation that becomes negative with increasing spatial distance. Along most of shelves and northern part of the domain, correlation function seems to be consistent with the  $C$  function actually parameterized in the system.

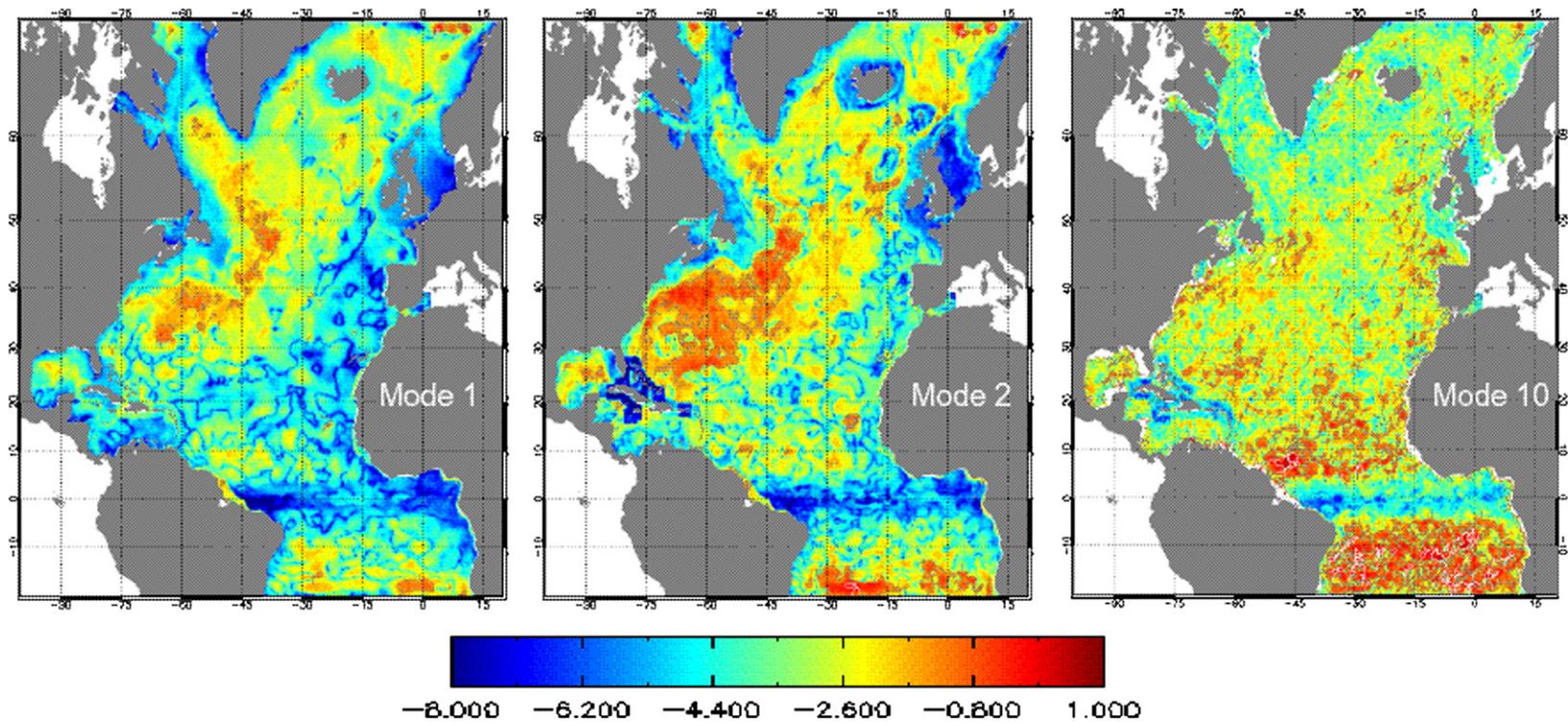


Fig. 6. Log of the ratio between 28-day ensemble variance and mode variance of S95 in the reduced space for mode 1, 2 and 10. White and gray contours correspond to a ratio equal to 1 and 0.04 respectively.

Second, the reduced space is built to depict the signal (EOFs). But we have to nearly rearrange these modes to represent the error. It means, as shown in Fig. 6, that if we want to keep the existing relationship between signal and error statistics in the assimilation scheme (error variance equal to a ratio of the signal variance), we have to deal with both the spatial specificity of the error and their real projection onto the signal EOFs. The ratio between error and signal variance has to be mapped depending on the area and the mode. This shows that the error statistics is not so homogeneous as we expect in our initial scheme.

Two new experiments are now conducted with the multivariate scheme to consider these new pieces of information separately. First, we decide to test the negative correlation function  $C_m$ , then a new error variance with an error/signal ratio computed from an ensemble experiment.

**4. Impact of the correlation function**

Two simulations over year 2003 are compared here. We choose this period because satellite data, in situ observations and a reference run (hereafter  $S_{ref}$ ) were available for the multivariate assimilation scheme. Notice that  $S_{ref}$  is performed using the  $C$  correlation function (Fig. 5). The second experiment called  $S_m$  is performed using the  $C_m$  correlation function over the

whole domain and the same correlation length scales as in  $S_{ref}$ . The choice of the initial set of correlation radii can be questioned. We decided, at first, to estimate new spatial and temporal correlation radii from the computed error correlations. But the use of these new parameters over 2003 did not improve the assimilation and on the contrary made the forecast worse than the  $S_{ref}$  results. This is due to the fact that the new length scales were smaller than the reference one. As mentioned previously, the spatial correlation decrease of  $C_m$  is faster than of  $C$  since  $C_m$  is equal to zero when  $C$  is equal to 0.55 (Fig. 5). Moreover, observations are selected in an influential bubble around each grid point. The size of this area in the zonal and meridian directions is twice the spatial correlation radii ( $r/r_0=2$  in Fig. 5). Added to a reduction of these spatial scales, it leads to very localized correction introducing strong and unrealistic gradients in the model.

*4.1. Assimilation diagnostics*

The first thing to be observed comparing these two experiments is the strong decrease of the increment spatial scales (Fig. 7) from  $S_{ref}$  to  $S_m$ . Whatever the corrected variable, the influence of the data is reduced. Fig. 7 shows the example of the SST increment for the first assimilation cycle, both for the reference and the new simulation. Influence of the data and thus of the

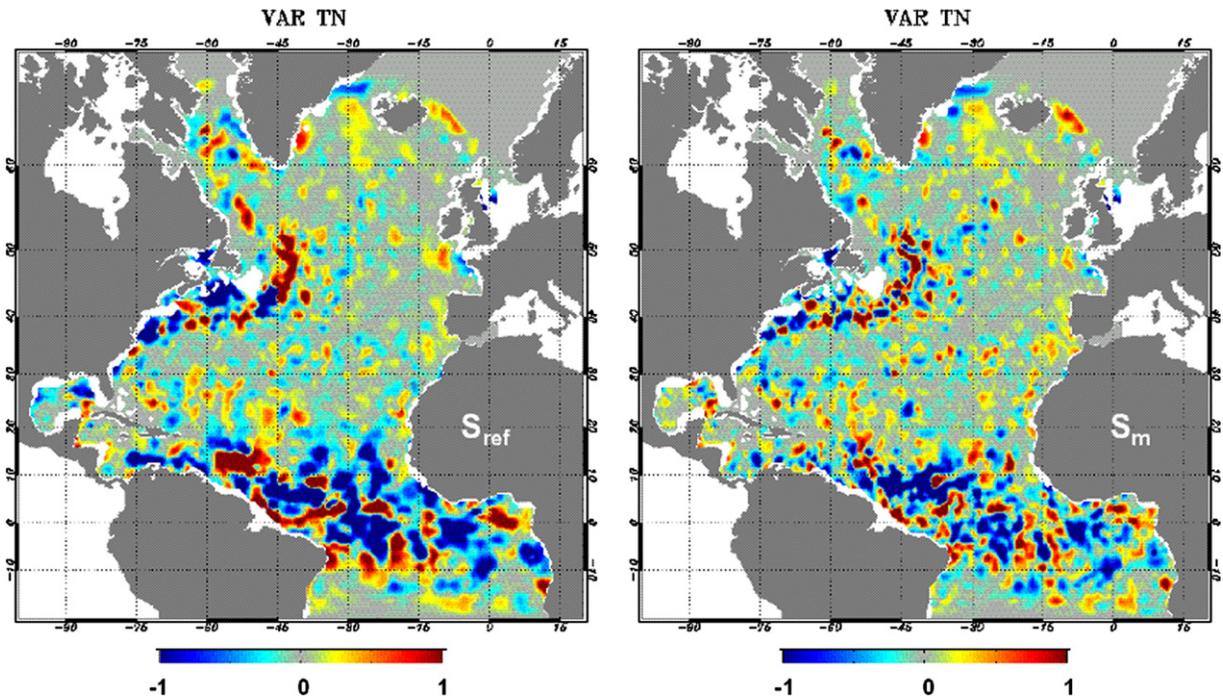


Fig. 7. First SST (°C) increment for  $S_{ref}$  (on the left) and  $S_m$  (on the right).

correction is now more concentrated around each model grid point. The SLA assimilation diagnostics are a little bit better in  $S_m$  than in  $S_{ref}$  (Fig. 8). The rms of the difference between model and observations can be reduced, depending on the satellite, of 0.5 to 1 cm. Usually, this rms is below 8.5 cm in  $S_m$ . The ratio between rms misfit and rms data is always below 1 and is smaller in the new simulation compared to the reference. In other words, the  $S_m$  forecast is statistically closer to SLA data. Of course, this improvement can be very different from one region to another. The greatest impact on the assimilation results can be seen in the Florida Strait, near Puerto Rico (where the misfit can be reduced by 1.5 cm in summer), the Cape Verde zone, and generally between 20°N and 20°S. Assimilation statistics for temperature and salinity (not shown here) are also better than those of the reference, and particularly between 100 and 1500 m.

Maps of temperature and salinity increment rms validate the progress made in regions previously mentioned. Between 10°S and 20°N, this diagnostic has decreased of 0.5 °C and almost 0.15 psu at 100 m

and 0.8 °C and 0.2 psu near 300 m. Great improvements are also made in the Labrador Sea and in the subtropical gyre. Moreover, this new simulation shows a global decrease of all the full state analysis increment, even on baroclinic and barotropic height. Considering temperature and salinity, this is obvious in the North Brazilian Current, in the Caribbean Sea and near the Gulf Stream overshoot.

#### 4.2. $S_m$ forecast versus climatology

In addition of the benefit of the new correlation function on the assimilation statistics, we obtain an improvement of the forecast itself. Temperature and salinity bias between model and climatology (Reynaud et al., 1998) is dramatically reduced in the new simulation. This can be seen all over the domain and along the whole water column, but especially in the tropics (see Fig. 9 for temperature at 100 m). The strong bias of more than 5 °C between 10°N and 20°N (east of Puerto Rico) and 3 °C near the equator has significantly decreased. At this level, salinity difference between

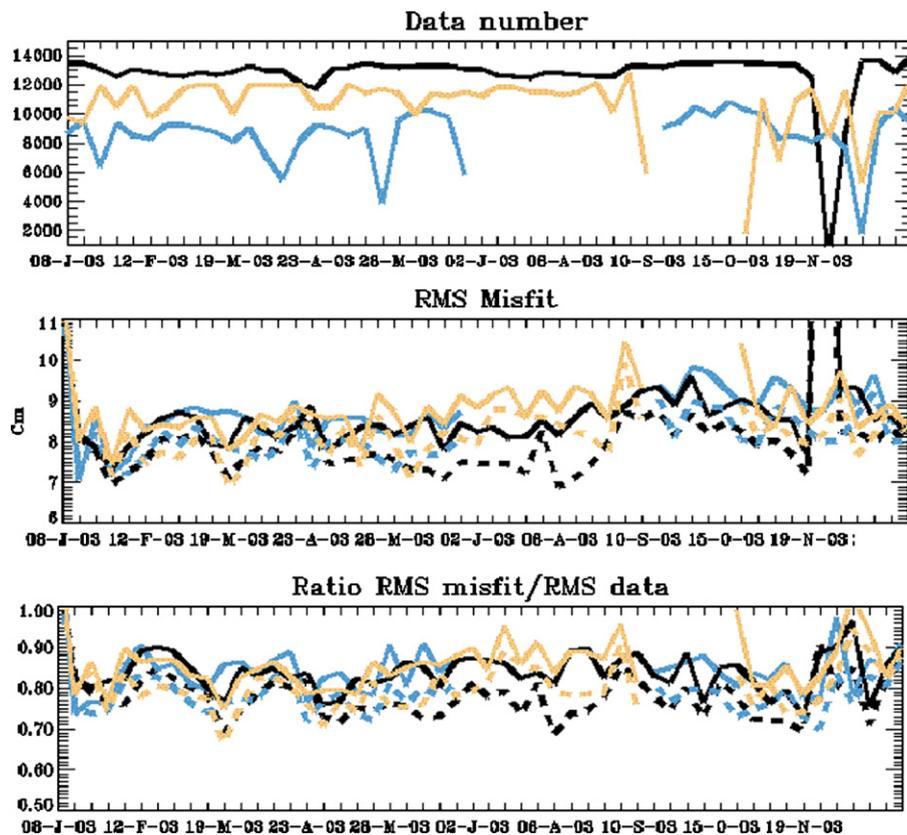


Fig. 8. SLA assimilation diagnostics for  $S_{ref}$  (solid lines) and  $S_m$  (dash lines) over 2003. Data come from JASON (black), ENVISAT (blue) and GFO (orange).

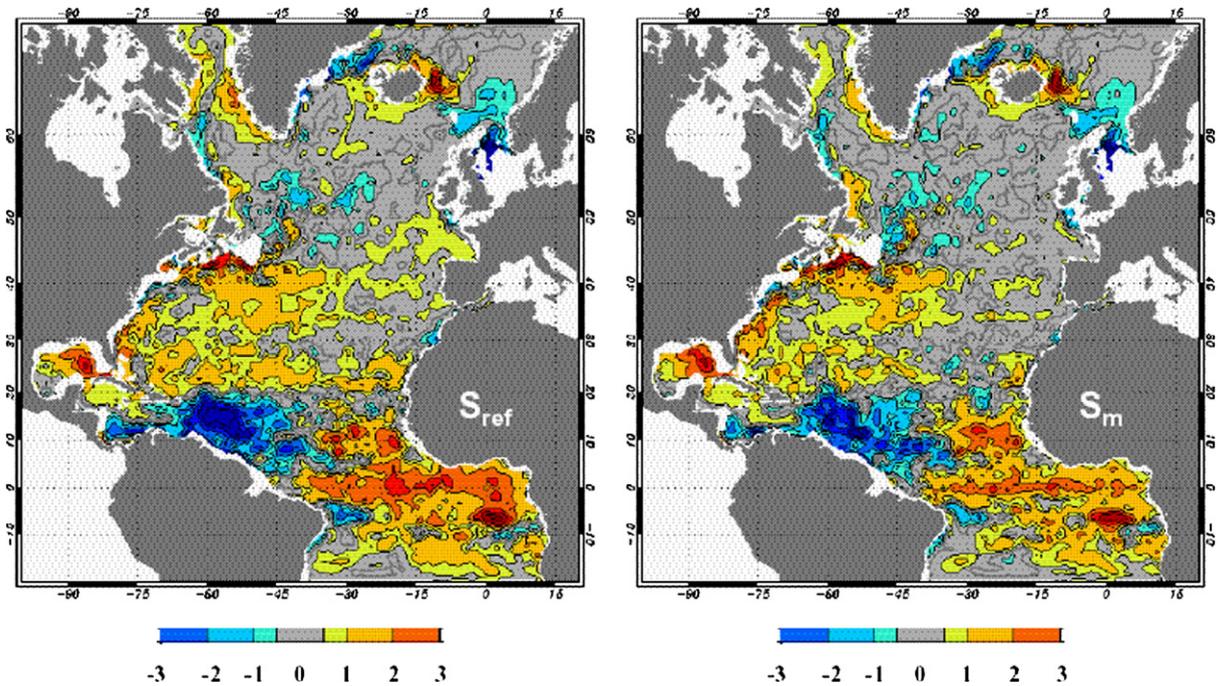


Fig. 9. Mean over 2003 of the difference between model and climatology for temperature ( $^{\circ}\text{C}$ ) at 100 m. Left (right) plot compare  $S_{\text{ref}}$  ( $S_m$ ) temperature output and climatology.

model and climatology has fallen below 1.2 psu (instead of 1.4 psu in the reference). The forecast of the Mediterranean outflow is also strongly enhanced by the change in the correlation function (Fig. 10). Mediterranean water is saltier than in  $S_{\text{ref}}$ . Most of the outflow spread to the North along the Portugal coast. In  $S_m$ , this water concentrates in the Cadix area, explaining the lack of salt between the coast and the Azores. Moreover, the salinity of the water mass that goes to the south is now increased.

However, the use of the new correlation function has strengthened the SST bias already existing in the reference ( $0.25^{\circ}\text{C}$  over the domain). This bias grows near the Brazilian coast to the Caribbean Sea and the Gulf Stream. One explanation is that the SST data use in the assimilation is a low resolution field ( $2^{\circ} \times 2^{\circ}$ ). A small amount of observations is then selected for the assimilation around each grid point, due to the spatial reduction of the correlation. It is probably not enough to constrain the model SST. In spite, bias patterns are located along the slope where a mask is applied on the error variance to slightly reduce the correction there (see M. Benkiran, personal communication). Therefore, in the  $S_m$  simulation, characteristics of the correlation function just intensify the bias structures already existing (from  $0.5$  to  $0.8^{\circ}\text{C}$  more along the Brazilian coast or in the Gulf Stream).

#### 4.3. $S_m$ forecast versus in situ data

$S_{\text{ref}}$  and  $S_m$  simulations are now compared to in situ data. The first data set to be compared are PIRATA (Pilot Research moored Array in the Tropical Atlantic) moorings. They are used inside the assimilation scheme (no independent data). Nevertheless, they can reveal information about how the assimilation acts on the forecast field and if the analysis is consistent with these observations in the tropical area (we saw that this region was sensitive to a change in the correlation function). Ten PIRATA moorings have been deployed since 1997 in the tropical Atlantic (Servain, 1996) to study the ocean–atmosphere interactions in this area using ATLAS (Automated Temperature Line Acquisition System) line. In the first 500 m, ten temperature measurements are made and four in salinity between surface and 120 m. Data are available in real time on the web. The second data sets to be used are measurements from SVP (Surface Velocity Program) drogoue buoys, launch for the Global Drifter Program of GOOS (Global Ocean Observing System). They give current velocity near 15 m depth in the Atlantic Ocean. Trajectories are collected using ARGOS (Niiler et al., 1995) and resampled by Kriggeage technique (Hansen and -M., 1996) every 6 h. Corresponding velocity are computed

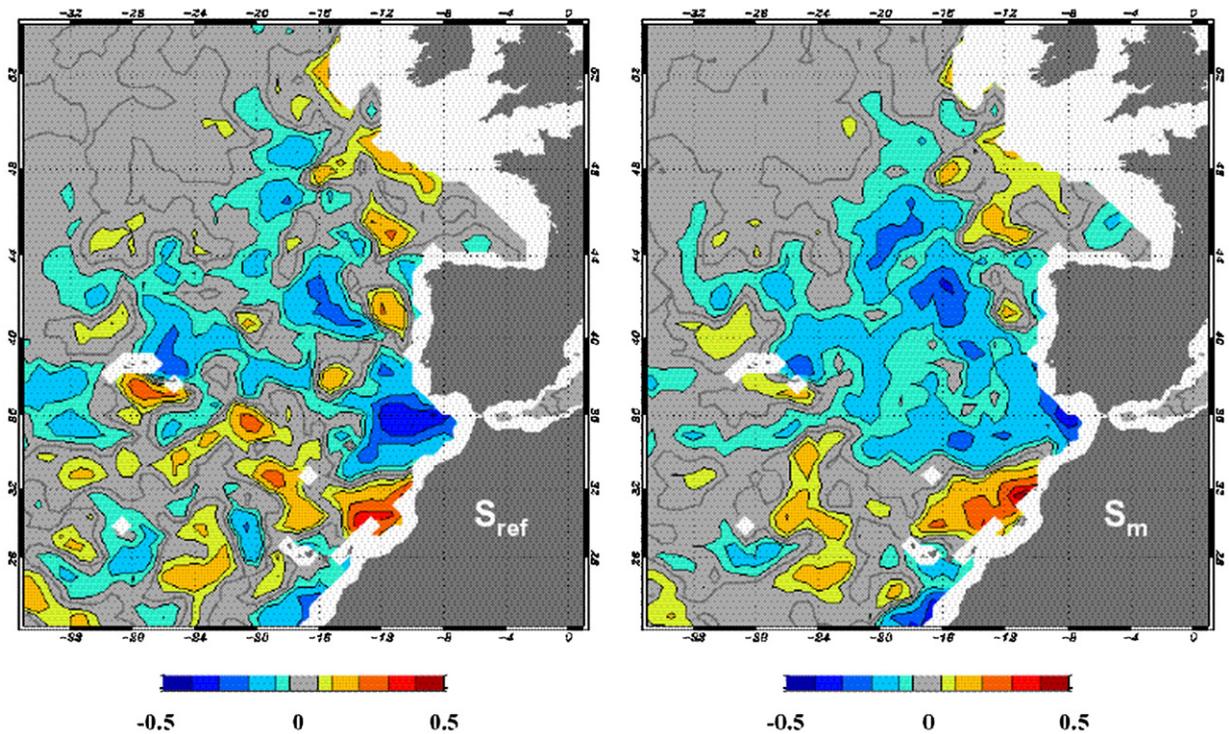


Fig. 10. Difference between model and climatology salinity (in psu) at 1000 m for  $S_{ref}$  and  $S_m$  in the eastern Atlantic.

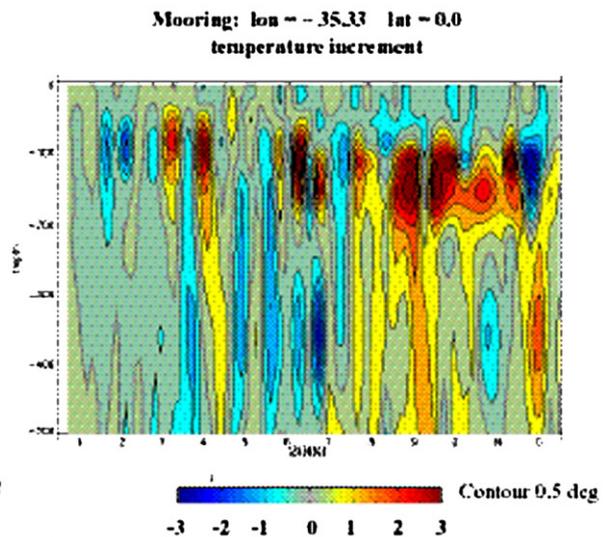
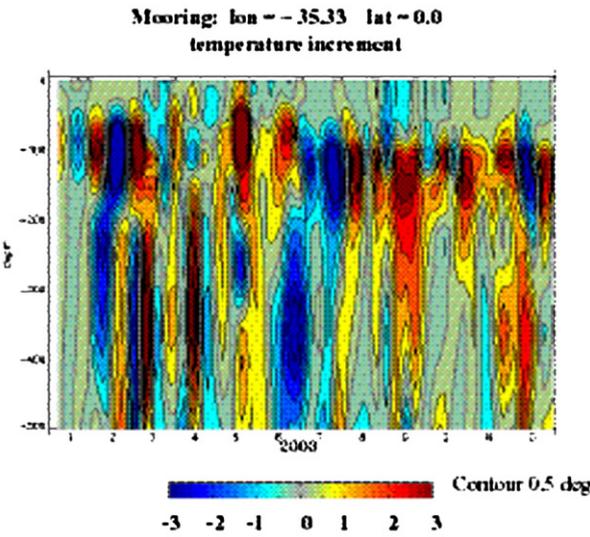
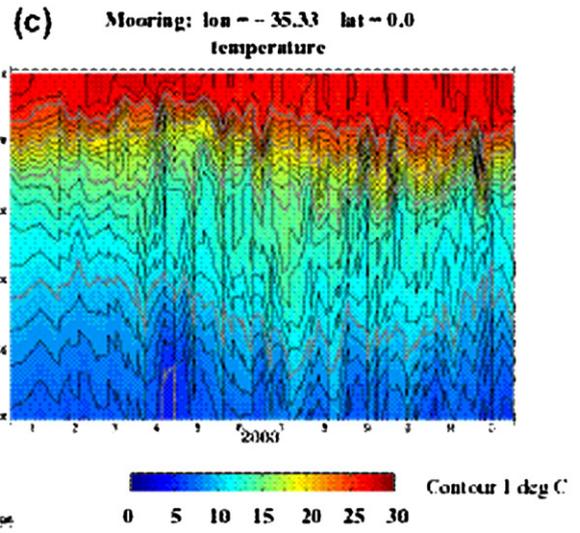
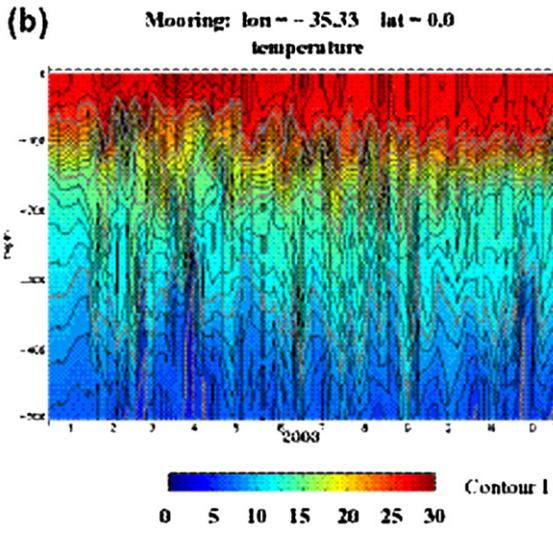
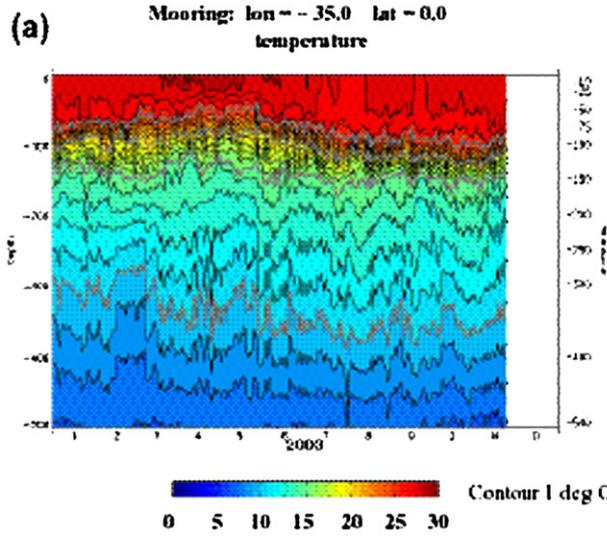
at AOML (Atlantic Oceanographic and Meteorological Laboratory, Miami Florida).

Time series of  $S_{ref}$  and  $S_m$  model forecast at PIRATA mooring show a smoother field in  $S_m$  which is more consistent with the observations, due to the reduction of the weekly gap between forecast and analysis. It makes sense with the decrease of the analysis increment observed in the tropical area and in most of the domain. Sharp events in  $S_{ref}$  (thermocline cooling or rise of cold water) are thus limited in  $S_m$ , which is closer to the observed mooring (Fig. 11). Moreover, major differences in the temperature and salinity increment magnitude are concentrated just below the thermocline and in the first 400 m (Fig. 11, II). As seen previously, these depths are more sensitive to the  $C_m$  correlation function. Even if the SST is colder (and too cold) in  $S_m$  (Fig. 11, I a), both compared to observations and  $S_{ref}$  (Fig. 11, I b), the position of thermocline is more realistic. At depth, some cold events are still over estimated but they are closer to the observations. These results are quite accurate for all PIRATA moorings. At this stage, we have to remind how the analysis is performed in the

mixed layer (for a detailed description, refer to M. Benkiran, personal communication). In the mixed layer, pseudo observations are computed. They are equal to the model forecast, so that the misfits are zero. This means that we consider that the model is perfect between surface and mix layer depth to preserve the structure there. In the meantime, profiles below are taken into account (with very small weight) for the correction above and this is why analysis increments are not equal to zero in the mixed layer. This explains the reason why there is such a difference below the thermocline between the two simulations, added to the undersized correction introduced by the new correlation function.

The new simulation is also able to fit buoy velocity measurements better than the reference. This comparison has been made using an analysis (corrected) velocity field available the day of the measurements and interpolated at the buoy's position. Correlation along the instrument trajectory is computed over the 2003 period. Overall, the correlation between observed and simulated zonal velocity is enhanced in  $S_m$ . This is not very obvious in the meridional field. The tropics

Fig. 11. (a): Temperature ( $^{\circ}\text{C}$ ) time series at PIRATA mooring ( $35^{\circ}\text{W}$ ;  $0^{\circ}\text{N}$ ) for the PIRATA observations (a),  $S_{ref}$  (b) and  $S_m$  (c). (b): time series of the temperature increment at  $0^{\circ}\text{N}$ ;  $35^{\circ}\text{W}$  (PIRATA mooring position) for  $S_{ref}$  (left plot) and  $S_m$  (right plot) simulations.



(between 5°S to 5°N), eastern part of the subtropical gyre and the Caribbean Sea are the areas of great improvement. Elsewhere (as in the NBC or in the Mexico Gulf) only the zonal velocity seems to be better estimated except in the Gulf Stream, where only the meridional component is improved. We use in these experiments the same correlation radii and thus the same directions for the error anisotropy. But doing so, we probably fail to catch the anisotropy contains in the ensemble (and the corresponding error covariance).

Thus what is pointed out in this section is the extreme importance of the choice of the error correlation function. Introduction of negative correlation (Fig. 5) into the error statistics structure is a benefit both for the assimilation results (decrease of the increment rms and mean value) and for the forecast itself (closer to observation and climatology) mostly because of the located increment it produces. As expected, most of the regions identified in the previous section as  $C_m$  regions are strongly improved in the new simulation. The only point is the small SST bias that developed in  $S_m$ .

## 5. Impact of the ensemble error variance

From the results obtained with the 28 days error statistics (Fig. 6), we have tried to find a way to better estimate the reduced order error variance to run a new simulation over 2003. As the system uses seasonal EOFs and corresponding variances, we first estimate new ensemble to deal with the seasonal variation of the error growth, as explained previously. The four experiments are performed as describe in Section 2.3.  $X^a(t_0)$  of Eq. (8) is now taken from the reference at the beginning of January, April, July and October 2003 to catch the seasonal variation of the error growth. The initial perturbation fields  $A = \{\varepsilon_p\}$ ,  $p \in [1, 50]$  are the same in the four experiments. The 7-day ensembles are then use to compute error variance in the reduced space by projection onto corresponding seasonal EOFs. This new set of error variance is introduced in a new experiment over 2003 called  $S_v$ . To clearly identify the impact of the new statistics, the correlation function is equal to the  $C$  function in both  $S_v$  and  $S_{ref}$  experiments. The 0.2 ratio of Eq. (5) is just replaced by the ratio between ( $m$ ) (the ensemble SD on mode  $m$ ) and  $V^{1/2}(m)$  (Fig. 6). To avoid strong amplitude in the correction (and sharp increment gradient that could create numerical problems), the ratio is limited to 1. In this case,  $D(m)$  is equal to the ensemble variance of the mode divided by the total variance of the signal (Eq. (5)). The new simulation, called  $S_v$ , is performed using this new seasonal set of error variance and is compared to  $S_{ref}$ .

### 5.1. The new error variance

As seen in Section 3.2, patterns of the error variance, when obtained by ensemble experiments, are quite different from the signal variance. The error variance has both a new spatial repartition of its amplitude and a new distribution on the mode. On Fig. 12 the winter error SD for mode 1 and 5 of  $S_{ref}$  (normalised EOFs SD) and  $S_v$  are shown. Whatever the mode, spatial repartition in the  $S_{ref}$  set is smoother compared to what we obtain in  $S_v$ . Amplitude of the  $S_v$  error SD can reach 0.9, whereas in  $S_{ref}$  it is limited, by definition, to 0.2 (normalisation of the error covariance matrix). Here, only the winter error SD is shown. But except for the amplitude, the spatial repartition and the mode distribution are similar in the other seasonal EOFs.

Strong differences appear between error variances of the two experiments. Considering the first mode in  $S_{ref}$ , the amplitude is near 0.18–0.19 in most of the major currents of the northern hemisphere. We can see it is not the case in  $S_v$ , where mode 1 is also prevailing in the northern basin but mostly on the western side. Most of all, amplitude is reaching more than 0.6 in the subpolar gyre and in the western side of the subtropical gyre. Another important difference on the first mode can be seen between equator and 10°N. Whereas the error SD is about 0.09 in  $S_{ref}$ , it can be more than 0.2 in  $S_v$ .

Generally in this new simulation, the higher the order of the EOFs, the more southern is the error SD maximum. Modes from 4 to 10 represent error statistics between 10°N and 20°N and south of 10°S. Modes from 1 to 4 are dominant off the European coast, in the Gulf of Mexico. The Cape Verde area is represented by mode from 3 to 5.

Magnitude and spatial distribution of the error variance in the reduced space is completely reviewed. Therefore, we may obtain a new repartition of the correction between barotropic and baroclinic components depending on the area.

### 5.2. Assimilation diagnostics

SLA assimilation diagnostics of this new simulation are better than those of the reference. They are almost similar to those of the  $S_m$  simulation, and the same kind of improvement can be seen. But, in some region as east Puerto Rico and near Dakar differences between model and observations are less weaker in  $S_v$  than in  $S_m$ . Except in the Irminger Sea and near Iceland,  $T$  and  $S$  vertical misfit are also reduced, with global improvement of the increment statistics. This is also the case compared to  $S_m$ .  $T$  and  $S$  mean increments are

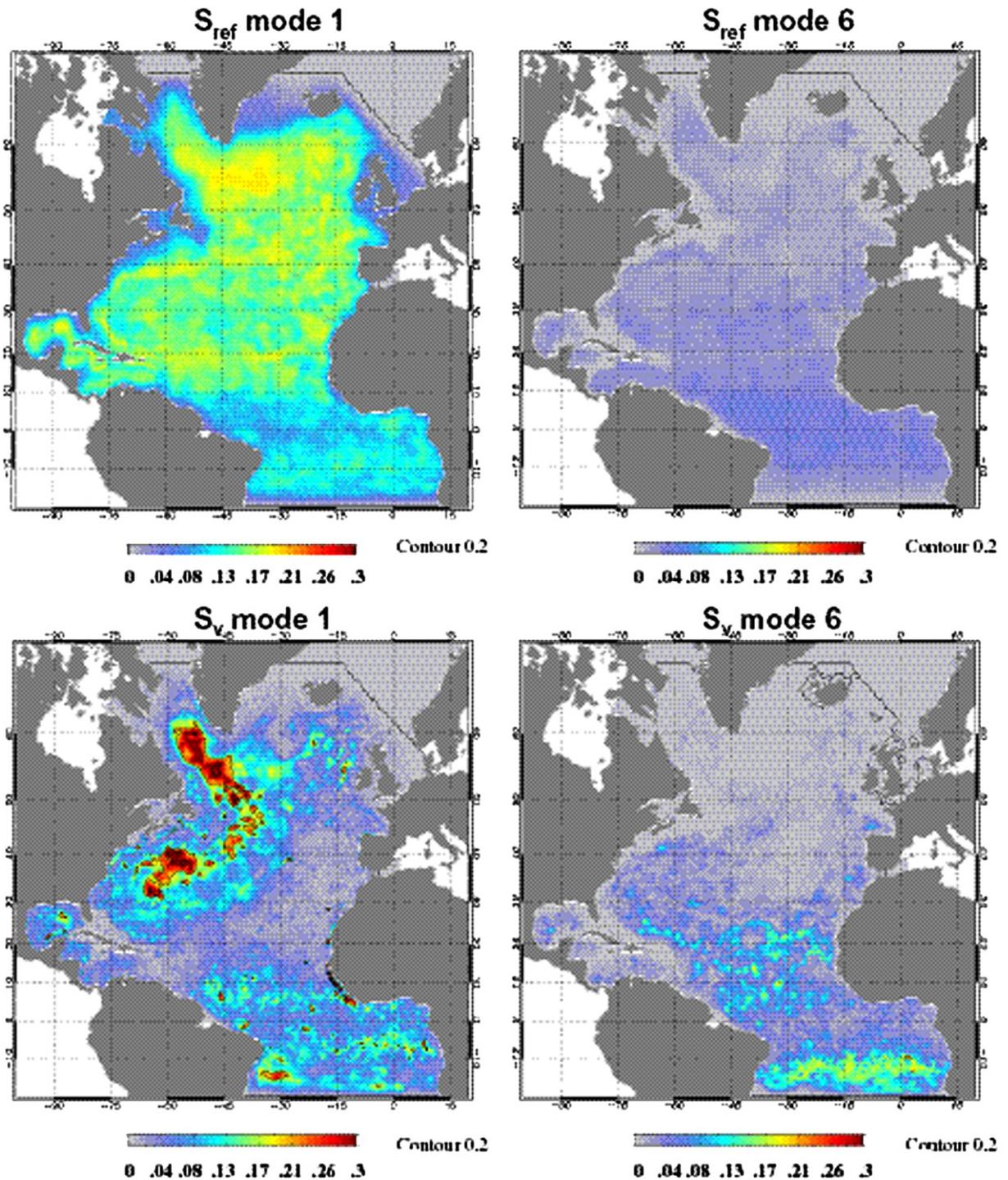


Fig. 12. Error standard deviation of winter period used in  $S_{ref}$  (upper plots) and  $S_v$  (lower plots) for mode 1 (right plots) and mode 6 (left plots).

considerably reduced below 100 m. Fig. 13 shows some  $T$  and  $S$  SD increments of  $S_m$  and  $S_v$  at 100 m and 300 m. Major improvements brought by the new error variance can be seen in the tropics and the Gulf Stream. Region east of Puerto Rico (already identified as a

sensitive area) and the Caribbean Sea, have seen the  $T$  and  $S$  increment statistics strongly reduced in  $S_v$ , with for example, about 0.5 psu and more than 0.5 °C discrepancy between the two. Except for the Labrador Sea, this difference is maximum below 100 m.

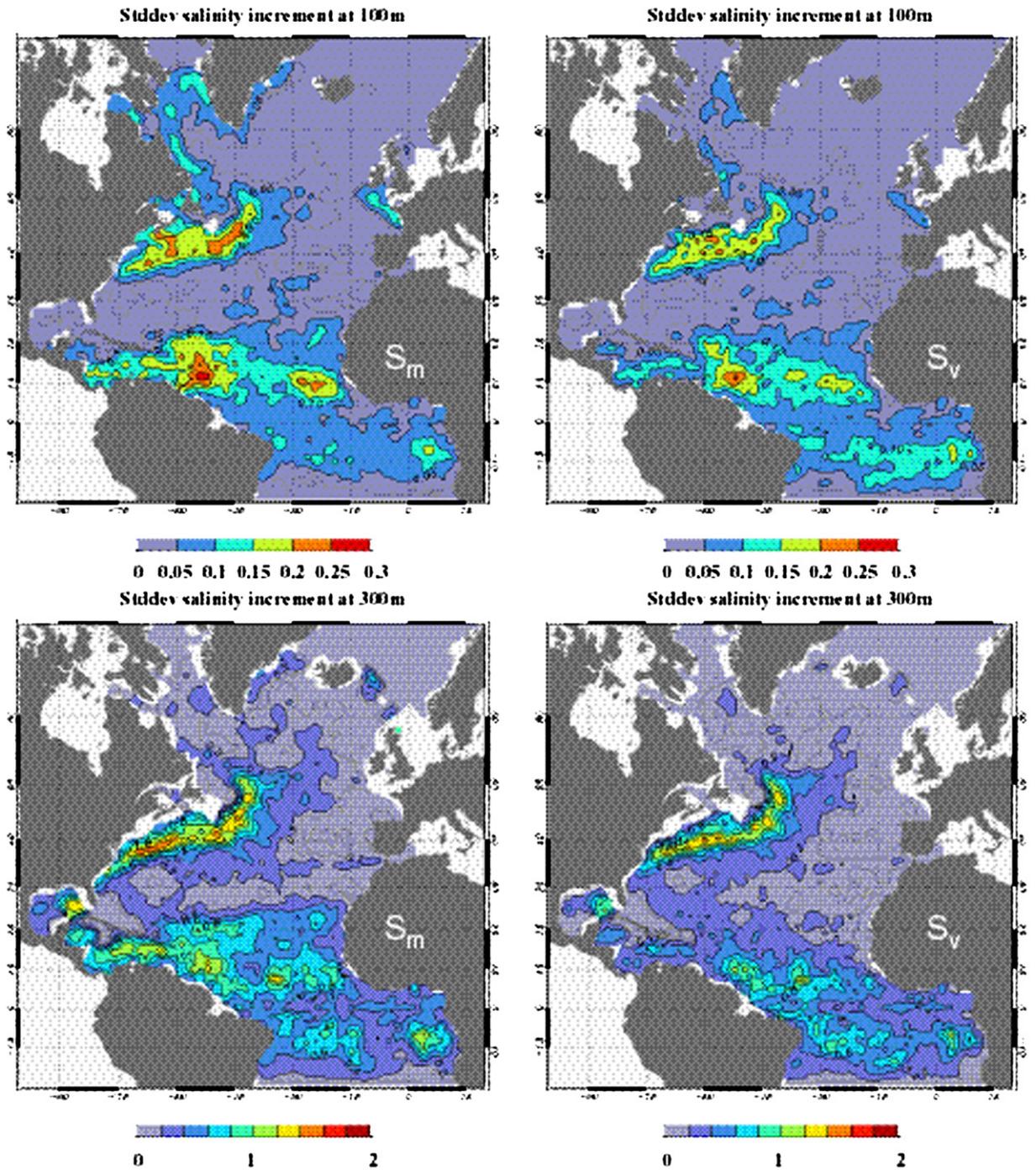


Fig. 13. Increment standard deviation of salinity (psu) at 100 m (upper plots) and temperature (°C) at 300 m (lower plots) for the  $S_m$  and the  $S_v$  simulations.

Overall, the impact of the new error variance on assimilation diagnostics is prevailing in the tropics between 20°N and 10°S. Error variance in this region is mostly expressed on high order modes (between 4 and

10). The  $S_{ref}$  parameterization is obviously not able to vertically constrain the model in the tropics mostly because of the wrong representation of the error on the reduced space. Using the signal EOFs to express the

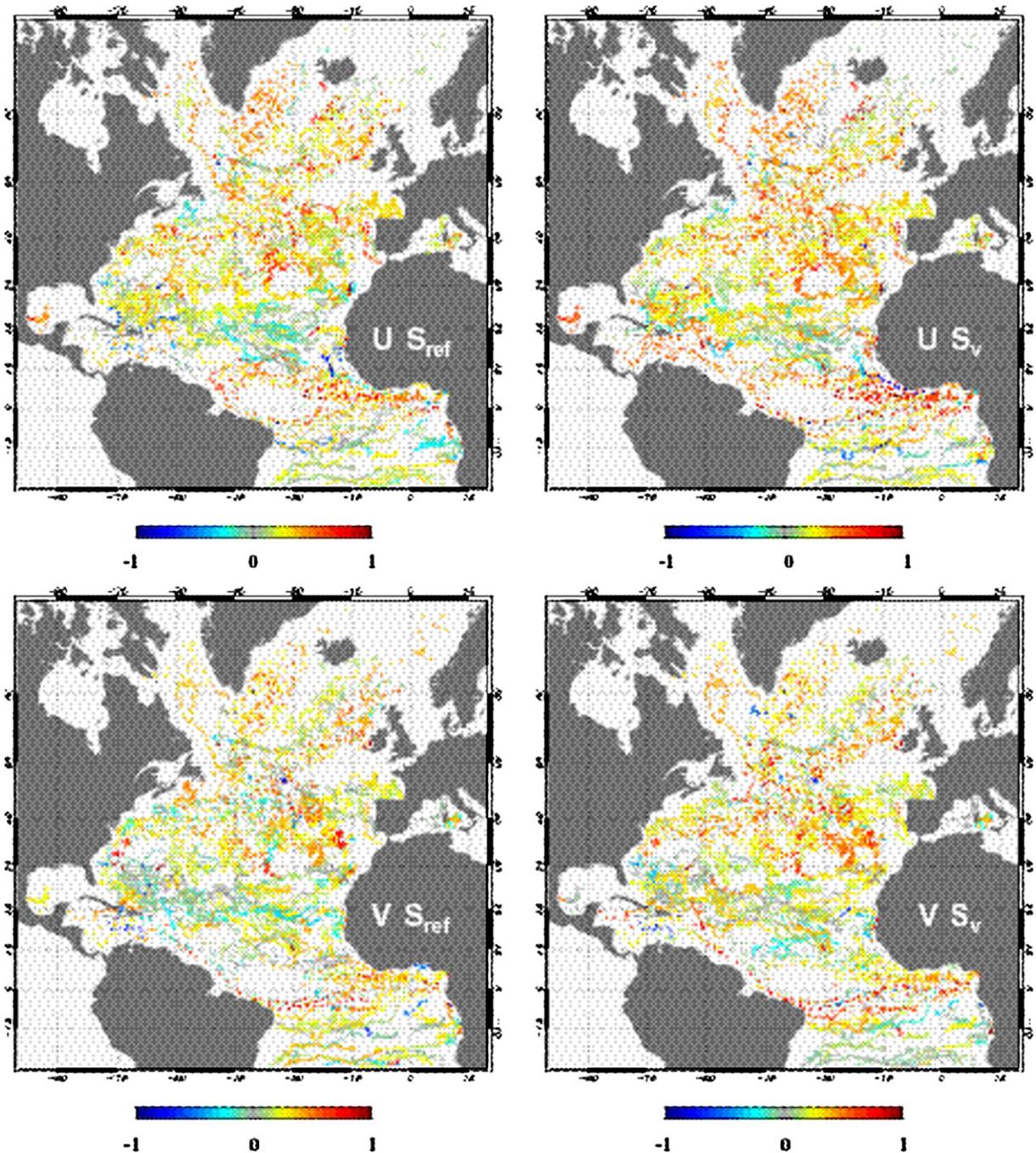


Fig. 14. Correlation between drifter and model analysis velocity components for the reference (left plots) and  $S_v$  (right plots).

error is promising, assuming the ratio between signal and error decomposition as a function of the grid point and the mode. In other word, it has to depend on the local dynamics.

The question is which EOFs is dominant for the error on the signal decomposition? Ten EOFs are enough to reconstruct more than 90% of the signal variance. We

assume a constant ratio (as in  $S_{ref}$ ) between error and signal variances. But we have demonstrated that dominant signal EOFs are not dominant for the error. Truncation of the reduced space to ten modes probably fails to register an important part of the error variance contained on the neglected part of the signal decomposition. Ensemble experiments allow the computation of

a multivariate error subspace fully determining the error. Using the same technique (diagonalisation of the covariance matrix) we can find some orthogonal vectors and corresponding eigenvalue to decompose the error. This has been done, but it has been unsuccessful in improving the assimilation results. Assimilation diagnostics, both for SLA and in situ data, were worse than the reference.

### 5.3. $S_v$ forecast versus climatology and in situ observations

As in the previous  $S_m$  experiment, we found a SST bias in  $S_v$  (no more than 0.25 °C), whereas on the vertical, the difference between  $S_v$  and climatology is strongly reduced compare to the  $S_m$  minus the climatology. One exception is the equator near 100 m depth for  $T$  and  $S$  fields, where a bias remains. Once again, the SST bias structures grows near the continental slope, but this time spatial scales of the correlation are not involved. By projection of surface fields onto the multivariate ( $\Psi$ , SST, SSS) EOFs, we see that the SST is expressed mostly on mode of order greater than 6. But this is also modes for which the error variance decreases on the west continental slope compared to the reference (Fig. 12). This is a reason for the small increase of the SST bias in this area.

At Gibraltar strait near 1000 m, salinity bias has mostly disappeared. There only remains a difference between model and climatology of about 0.2 psu (0.4 psu in  $S_m$  and 0.5 psu in  $S_{ref}$ ). Mediterranean outflow seems to be better estimated in this new configuration, than it was previously. Moreover the temperature bias has totally disappeared at Gibraltar outflow and in the Azores.

The mean increment amplitudes are reduced in  $S_v$ , compared to  $S_m$ , except near the Amazon outflow (along the Brazilian coast between 5°S and 5°N). In this area, temperature increments increase in the new simulation just below the thermocline (around 100 m) and near 400 m. But at the same time, standard deviation of this increment is falling and bias is kept unchanged. Moreover, model temperature at PIRATA mooring positions is particularly consistent with the observation south of the equator (moorings near 10°S and 6°S). In this area, position and gradient of the thermocline is closer to what is observed. North of the equator, improvement can essentially be found above the thermocline. Thus even if the assimilation statistics seem to give better results in  $S_v$  than in  $S_m$ , the equator area remains a particular region that still causes problem for the model and the assimilation.

Considering now the drifter measurements, we can see that  $S_v$  velocities are now more consistent with SVP

measurements than what can be found in the reference (Fig. 14). Correlation is increased by 7% for  $U$  (zonal direction) and 9% for  $V$  (meridional direction). It is obvious in the equatorial and in the north tropical areas (Guinea Gulf, NBC). The meridional velocity in the Gulf Stream seems to be more consistent with the drifter, on the opposite of  $U$ . Velocity is also better estimated in the Sargasso Sea and globally in all the subtropical and subpolar gyres.

## 6. Conclusion

The major problem of the assimilation scheme still consists in the choice of the error statistics, such as guess error covariance matrix  $\mathbf{B}$ . In the MERCATOR multivariate configuration studied in this paper, optimal interpolation is used to compute  $\mathbf{B}$  (Eq. (5)), by separate parameterization of the magnitude (variance) and the spatial and temporal structure ( $C$  function). Moreover, correction to the forecast is performed in a subspace of the model state by projection on the ten first EOFs of the ( $\Psi$ ,  $T$ ,  $S$ ) signal. We assume in Eq. (5) a constant ratio between error and signal standard deviation, which means similar spatial and vertical patterns. Moreover, the correlation function stating the spatial and temporal structure of the  $\mathbf{B}$  matrix is similar on each ten dimension of the reduced space. Hence, barotropic mode (first EOF) and baroclinic component of the guess error covariance are supposed to have the same correlation patterns. This paper aims to assess these assumptions by use of ensemble methods. Knowing a priori information on the error statistics, we perform an ensemble of model initial conditions by random choice of possible states that reconstruct the expected density probability function (cloud of points around the true state). This ensemble is then integrated by the model. Thus we can access the error statistics at every time step and evaluate the time propagation of the error growth.

Ensemble results show that length scale used to compute the covariance matrix are too large and that in some area the correlation can reach negative values. Moreover, the relation between error and signal standard deviation after decomposition on the signal subspace is not as simple as set initially in the assimilation scheme. The ratio between these two parameters has spatial patterns, depending on the area and the mode (dimension of the reduced space). Hence, decomposition of the ( $\Psi$ ,  $T$ ,  $S$ ) signal and decomposition of the guess error using EOF techniques seem to be different while dominant EOFs for the model state are not corresponding to dominant error patterns. Parameters chosen in the assimilation scheme suppose that the error statistics

are quite homogeneous over the domain and the reduced space dimensions. But, we show here that a localized correction gives better results both in the assimilation diagnostics and in the model forecast. This is obvious in the tropics where some  $T$  and  $S$  bias have been dramatically reduced. Thus the optimal interpolation assumption of Eq. (5) can be improved by a better estimation of the error variance and correlation function.

We could think that results could be even better by combining ensemble estimated error variance and correlation function. But further experiments didn't provide any confirmation. We can suppose that the correlation function depends on the mode, as the variance does. This can be supported by the strong correlation between error growth and dynamics (Moore, 1999). Knowing that mode 1 (barotropic mode) represents the large scale and the steric effect and that the mesoscale signal is present on the other modes, we can suppose that the first mode has to be treated separately in term of parameterization.

So, we have shown that ensemble experiments are able to bring reliable information on guess error statistics. The best results have been obtained using the ensemble forecast on the reduced space to estimate the guess error variance (Section 5). But this information is complex and has to be investigated further. Moreover, these statistics have interannual and probably seasonal patterns (Section 3). Longer assimilation experiments extended over 2004 show that error parameters estimated over 2003 are not as consistent with 2004 dynamics as we were expecting. So, even if these results are promising, longer experiments have to be performed to compute error statistics that can be used in any situation.

## Acknowledgements

This study has been made under GIP MERCATOR/CLS contract. The authors would like to thank PIRATA program for providing the data and R. Cousin (from CLS, France) for his help in the comparison between model and PIRATA buoy data.

## References

- Arhan, M., Colin de Verdière, A., 1985. Dynamics of eddy motions in the eastern North Atlantic. *Journal of Physical Oceanography* 15, 153–170.
- Cooper, M., Haines, K., 1996. Data assimilation with water property conservation. *Journal of Geophysical Research* 101, 1059–1077.
- De Mey, P., Benkiran, M., 2002. A multivariate reduced-order optimal interpolation method and its application in Mediterranean basin-scale circulation. In: Pinaridi, N. (Ed.), *Ocean Forecasting: Conceptual basis and application*. Springer Verlag.
- De Mey, P., Robinson, A., 1987. Assimilation of altimetry eddy fields in a limited-area quasi-geostrophic model. *Journal of Atmospheric and Oceanic Technology* 17, 2280–2293.
- Etienne, H., Dombrowsky, E., 2003. Estimation of the optimal interpolation parameters in a quasi-geostrophic model of the Northeast Atlantic using ensemble methods. *Journal of Marine System* 40–41, 317–339.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99 (C5), 10143–10162.
- Gavart, M., De Mey, P., 1997. Isopycnal EOFs in the Azores current region: a statistical tool for dynamical analysis and data assimilation. *Journal of Physical Oceanography* 27, 2146–2157.
- Gavart, M., De Mey, P., Caniaux, G., 1999. Assimilation of satellite altimeter data in a primitive-equation model of the Azores–Madeira Region. *Dynamics of Atmospheres and Oceans* 29, 217–254.
- Hansen, D.V., P. P.-M., 1996. Quality control and interpolation of WOCE/TOGA drifter data. *Journal of Atmospheric and Oceanic Technology* 13, 900–909.
- Julian, P.R., Thiébaux, H.J., 1975. On some properties of correlation functions used in optimum interpolation schemes. *Monthly Weather Review* 103, 605–616.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *Journal of Physical Oceanography* 23, 2541–2566.
- Kuragano, T., Kamachi, M., 2000. Global statistical space-time scales of oceanic variability estimated from the TOPEX/POSEIDON altimeter data. *Journal of Geophysical Research* 105, 955–974.
- Le Traon, P.Y., Boissier, C., Gaspar, P., 1991. Analysis of errors due to polynomial adjustment of altimeter profiles. *Journal of Atmospheric and Oceanic Technology* 8, 385–396.
- Madec, G., Delecluse, P., Imbard, M., Lévy, C., 1998. OPA8.1 ocean general circulation model reference manual. Notes du pôle de modélisation IPSL 11 (91 pp.).
- Maes, C., Benkiran, M., De Mey, P., 1999. Sea level comparison between Topex/Poseidon altimetric data and a global general circulation model from an assimilation perspective. *Journal of Geophysical Research* 104, 15575–15585.
- Moore, A.M., 1999. The dynamics of error growth and predictability in a model of the Gulf Stream. Part II: ensemble prediction. *Journal of Physical Oceanography* 29, 762–778.
- Mureau, R., Molteni, F., Palmer, T.N., 1993. Ensemble prediction using dynamically conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society* 119, 299–323.
- Murphy, J.M., 1988. The impact of ensemble forecasts on predictability. *Quarterly Journal of the Royal Meteorological Society* 114, 463–493.
- Niiler, P.P., Sybrandt, A.S., Bi, K., Poulain, P.-M., B.D., 1995. Measurements of the water-following characteristics of Tristar and Holey-sock drifters. *Deep Sea Research* 42, 1951–1964.
- Reynaud, T., Le Grand, P., Mercier, H., Bamier, B., 1998. A new analysis of hydrographic data in the Atlantic and its application to an inverse modeling study. *International WOCE Newsletter* 32, 29–31.
- Servain, J., 1996. PIRATA, science and implementation plan for an observing system to support tropical atlantic climate studies 1997–2000. Internal Report.
- Steger, J.M., Carton, J.A., 1991. Long waves in the tropical Atlantic ocean: 1984–1990. *Journal of Geophysical Research* 96, 15161–15171.
- Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological Society* 74 (12), 2317–2330.