

# Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions

Francis Auclair\*, Patrick Marsaleix<sup>1</sup>, Pierre De Mey<sup>2</sup>

*Pôle d'Océanographie Côtière de l'Observatoire Midi-Pyrénées, LA/LEGOS,  
UMR CNRS/UPS 5560–5566, 14, Avenue E. Belin, 31400 Toulouse, France*

## Abstract

The probability density function (pdf) of forecast errors due to several possible error sources is investigated in a coastal ocean model driven by the atmosphere and a larger-scale ocean solution using an Ensemble (Monte Carlo) technique. An original method to generate dynamically adjusted perturbation of the slope current is proposed. The model is a high-resolution 3D primitive equation model resolving topographic interactions, river runoff and wind forcing. The Monte Carlo approach deals with model and observation errors in a natural way. It is particularly well-adapted to coastal non-linear studies. Indeed higher-order moments are implicitly retained in the covariance equation. Statistical assumptions are made on the uncertainties related to the various forcings (wind stress, open boundary conditions, etc.), to the initial state and to other model parameters, and randomly perturbed forecasts are carried out in accordance with the a priori error pdf. The evolution of these errors is then traced in space and time and the a posteriori error pdf can be explored.

Third- and fourth-order moments of the pdf are computed to evaluate the normal or Gaussian behaviour of the distribution. The calculation of Central Empirical Orthogonal Functions (Ceofs) of the forecast Ensemble covariances eventually leads to a physical description of the model forecast error subspace in model state space. The time evolution of the projection of the Reference forecast onto the first Ceofs clearly shows the existence of specific model regimes associated to particular forcing conditions. The Ceofs basis is also an interesting candidate to define the Reduced Control Subspace for assimilation and in particular to explore transitions in model state space.

---

\* Corresponding author. Tel.: +33-561-33-27-75; fax: +33-561-33-27-90.

*E-mail addresses:* auclair@aero.obs-mip.fr (F. Auclair), patrick.marsaleix@aero.obs-mip.fr (P. Marsaleix), pierre.de-mey@cnes.fr (P. De Mey).

<sup>1</sup> Pôle d'Océanographie Côtière de l'Observatoire Midi-Pyrénées, LA, UMR CNRS/UPS 5560, 14, Avenue E. Belin, 31400 Toulouse, France.

<sup>2</sup> Pôle d'Océanographie Côtière de l'Observatoire Midi-Pyrénées, LEGOS, UMR CNRS/UPS 5566, 14, Avenue E. Belin, 31400 Toulouse, France.

We applied the above methodology to study the penetration of the Liguro-Provençal Catalan Current over the shelf of the Gulf of Lions in north-western Mediterranean together with the discharge of the Rhône river. This region is indeed well-known for its intense topographic and atmospheric forcings.

*Keywords:* Ensemble forecast; Coastal oceanography; Errors; Functional analysis; Ocean circulation; Mediterranean

---

## 1. Introduction

A universal but very challenging problem in earth science modelling and prediction is the characterisation of both conceptual and numerical model errors. Indeed whatever the model used to mimic nature is, the forecast must always be compared to observations and be adjusted to fit them within the limit of their own observational error estimates. However, to achieve this, the trust in the model and in the observations must be evaluated, this being closely related to the determination of the probability density function (pdf) associated with both model and observational errors. If one could specify such statistical information, the Bayes theorem would provide with an optimal solution at the forecast time. Without any drastic assumptions this is yet unrealistic and this theorem is nothing but a mathematical tautology (Tarantola, 1988). The main reason for that is the difficulty to evaluate quantitatively the errors resulting on one hand from the modelling assumptions and on the other hand from the various forcings that were specified to obtain the forecast.

If one makes the assumption that the processes under study are linear or can be considered as quasi-linear then the very popular Kalman filter and its extensions, the so-called Extended Kalman Filters (EKF) (Ghil et al., 1981), offer several tools to combine model forecast and observations. The error pdfs are described by their mean and covariance and higher-order moments are neglected. However, even with these strong assumptions the computation of the complete covariance evolution equation is beyond actual computational possibilities and, even if such computations could be carried out, several problems would still remain. On one hand, the initial error covariances are unknown and, as importantly, very few estimates of model errors are available, and on the other hand, it was shown that over long periods the forecast error covariance would increase to reach unrealistic and unbounded values due to the rejection of higher-order moments in the evolution equation (Evensen, 1992). Miller et al. (1994) used different versions of both the EKF and of Variational methods to deal with the simple double-well potential and Lorenz equations and showed the important limitations of such techniques in strongly non-linear cases.

The EKF and the equivalent linear filters were, however, implemented for short term forecasts based on very crude statistical assumptions of homogeneity, isotropy and quasi-linearity (Lacarra and Talagrand, 1988). If the initial state and forcing noise statistics are for instance supposed to be Gaussian then, under linearity assumptions and with an appropriate choice of the model state vector, the process remains Gaussian and can be represented by its first- and second-order moments. However, in coastal ocean modelling, the forecast error statistics remain largely unknown (they are probably not even exactly Gaussian) and such well-behaved

homogeneous (or even regionally homogeneous) and isotropic statistical description of the dynamics could seem rather over-simplistic (Echevin et al., 2000; Auclair et al., 2000a). The severe forcing imposed by the topography, the atmosphere, the land and the open ocean are to a large extent responsible for these difficulties. Both the time evolution (predictability time scale) and the spatial distribution (predictability space scale) of the forecast errors must thus be evaluated carefully before any robust control of the model trajectory based on observations be settled.

The objective of this paper is to describe the forecast error density probability of a coastal ocean high resolution model using an Ensemble technique (Evensen, 1994). This Monte Carlo approach deals with model and observation errors in a natural way. Indeed statistical assumptions are made on the uncertainties related to the various forcings, to the initial states and to other model parameters. Randomly-perturbed forecasts are carried out in accordance with the a priori error pdf. The evolution of the errors can then be traced in space and time and the a posteriori error pdf can be explored. Miller et al. (1994) showed that these Ensemble techniques are well-adapted to non-linear cases in so far as higher-order moments are implicitly retained in the covariance equation. This is true concerning both Ensemble assimilation techniques and Ensemble predictability and observability studies.

Even if they do not give a complete picture of the forecast error statistics in non-linear cases, the first- and second-order moments are yet giving the “mean forecast” path and the associated uncertainties in model state space. The study of the first- to fourth-order moments of the forecast errors gives both focus and limitations to this paper. Under such limitations, error covariances are explored by means of their Central Empirical Orthogonal Functions (Ceofs), leading to physical characterisation of the forecast error subspace in model state space.

This qualitative and quantitative error characterisation can help reducing the order of an assimilation problem, a suitable control subspace being chosen within this forecast error subspace.

This Ensemble approach is applied to the north-western Mediterranean known for the strength of its north-western and northern wind (respectively, the Tramontane and Mistral winds) and north-eastern winds. Northern winds blow all year long almost every other day for the Tramontane and are associated to cold and dry atmospheric conditions which are responsible for intense heat fluxes from the surface of the ocean to the lower atmosphere. This region is also interesting for the interactions of its slope current (the northern Mediterranean Current or Liguro-Provençal Catalan current—hereafter LPC) with a strong topography gradient: the Gulf of Lions shelf break. It is also a region of intense river forcing associated in particular to the Rhône river and its delta. The strength of these various forcings together make up for a very interesting case study for the evolution of model forecast errors and the proposed Ensemble technique clearly requires a good modelling of the slope LPC current.

Section 2 gives a description of the model implementation and of the types of perturbations brought to the model in order to obtain four trajectory Ensembles. The first and second moments of the forecast errors are studied for each Ensemble in Section 3 together with an evaluation of the Gaussian assumption through the evaluation of third- and fourth-order moments for each Ensemble. A physical characterisation of the leading error Ceofs is carried out in Section 4 for each Ensemble. In Section 5, the time evolution of the Reference

forecast and of the forecast errors are described based on their respective projections onto the associated Ceof basis.

## 2. Model implementation and Ensemble generation

### 2.1. Model implementation

The region of interest is the north-western Mediterranean. The domain (Fig. 1a) includes the Gulf of Lions, its shelf break and part of the Ligurian Sea. The Ligurian extension is aimed at correctly representing the upstream characteristics of the LPC current which follows the shelf break from the Ligurian sea into the Balearic sea (Millot, 1990). Its signature in surface elevation anomaly, temperature and depth averaged velocity can be observed on the April 8, 1998 field (Fig. 1b–d) which we will use as our initial field for the simulations. Fig. 1c also shows a steep horizontal temperature gradient above the shelf break that is vertically homogeneous and corresponds to the regular strong and dry northern and north-western wind conditions (Mistral, Tramontane) over the Gulf of Lions in the wintertime. This temperature front is also associated to a salinity front with fresh water spreading across the shelf due the various river discharges.

The 3D primitive equation free surface model *Symphonie* (Marsaleix et al., 1998; Estournel et al., 1997, 2001a) was implemented in order to study a 4.5-day period beginning on April 8, 1998 (at 1930Z) and including a long-lasting, strong Mistral wind event over the shelf after Day 3 (Fig. 2a and b). The atmospheric forcings were taken from the ALADIN model from Météo-France, using analyses over a  $10\text{ km} \times 10\text{ km}$  grid every 3 h. The ocean-atmosphere parameterisations we used are described in Estournel et al. (2001b). The Rhône, Petit Rhône and Aude river daily runoff values are provided by the *Compagnie Nationale du Rhône* and the *Directions Départementales de l'Équipement* (Fig. 2c). The characteristics of the model grid are shown in Table 1 together with the main modelling parameters.

Table 1  
Main characteristics of model implementation

Horizontal grid scale	$3\text{ km} \times 3\text{ km}$
Number of layers	19
Sigma layers	Two-fold Sigma layers (Beckers, 1995)
Number of model variables (m): surface elevation anomaly, temperature, salinity and velocity components	566,258
External mode time step	5.8 s
Internal mode time step	174.2 s
Advective scheme	Mixed (Estournel et al., 2001a,b)
Horizontal diffusion	$45\text{ m}^2\text{ s}^{-1}$
Turbulent scheme	(Estournel et al., 2001a,b)
Open-boundary conditions	Characteristics and relaxation
Bottom Rugosity length ( $Z_0$ )	$10^{-3}\text{ m}$
Coriolis parameter	Beta-plan

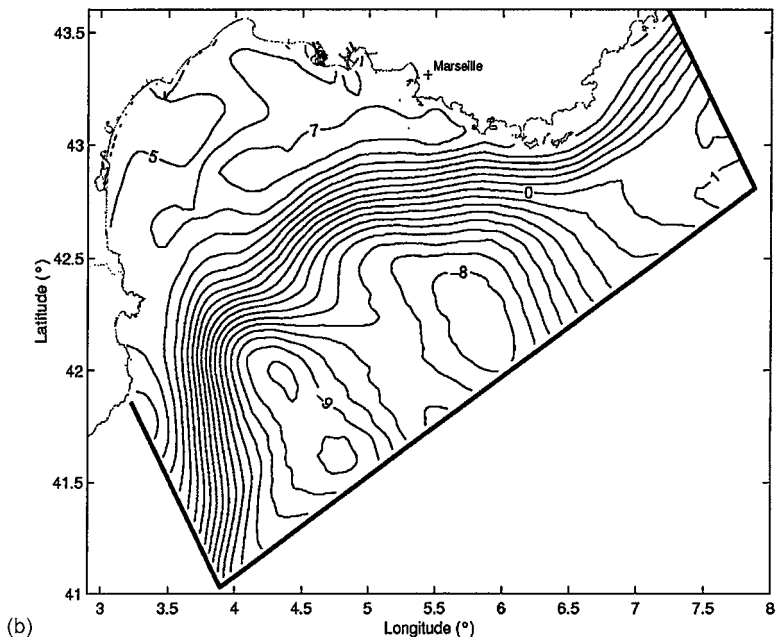
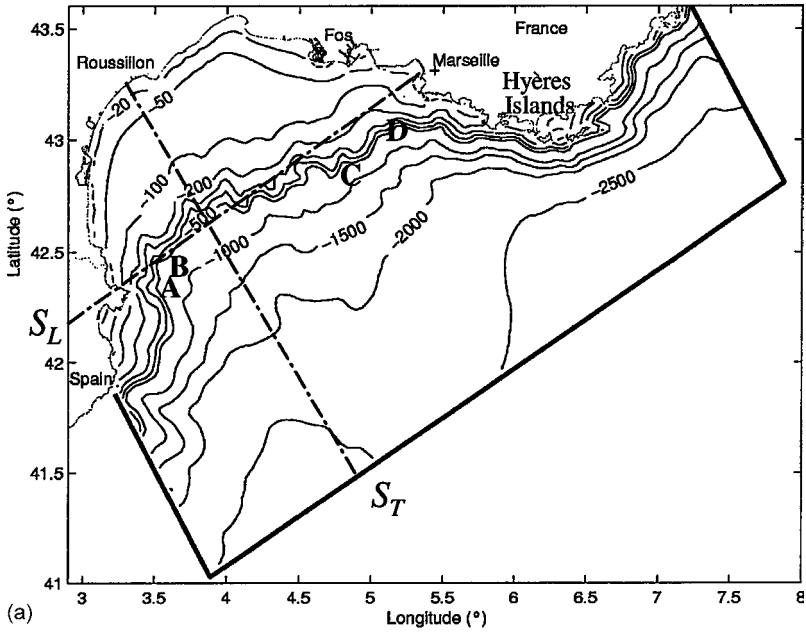


Fig. 1. Bathymetry and initial state of the Reference Forecast for 8 April 1998 at 1930Z. (a) Bathymetry in meters,  $S_T$  and  $S_L$  indicate the location of the Transversal and Longitudinal sections. The position of the main canyons is indicated by capital letters: (A) Cabo Creus, (B) Lacaze-Duthiers, (C) Grand Rhône, (D) Marseille. (b) Initial surface elevation anomaly. (c) Transversal section of temperature (along  $S_T$ ). (d) Depth average velocity.

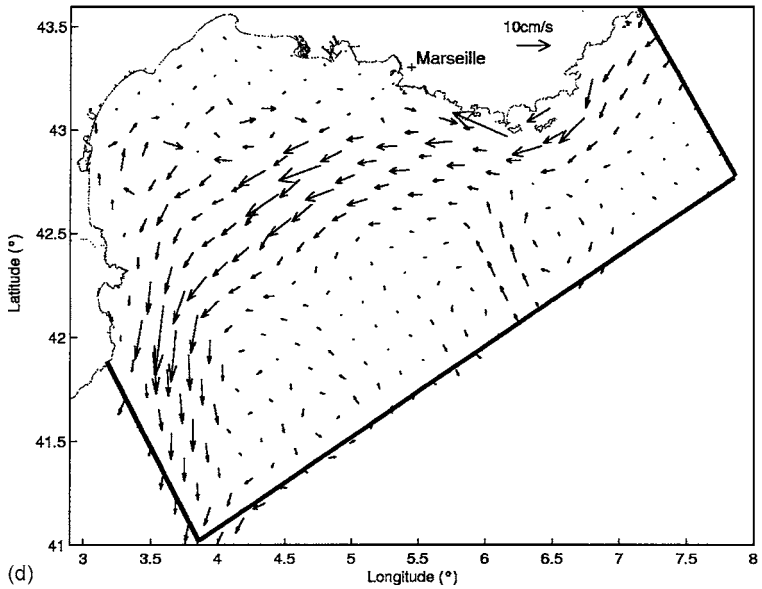
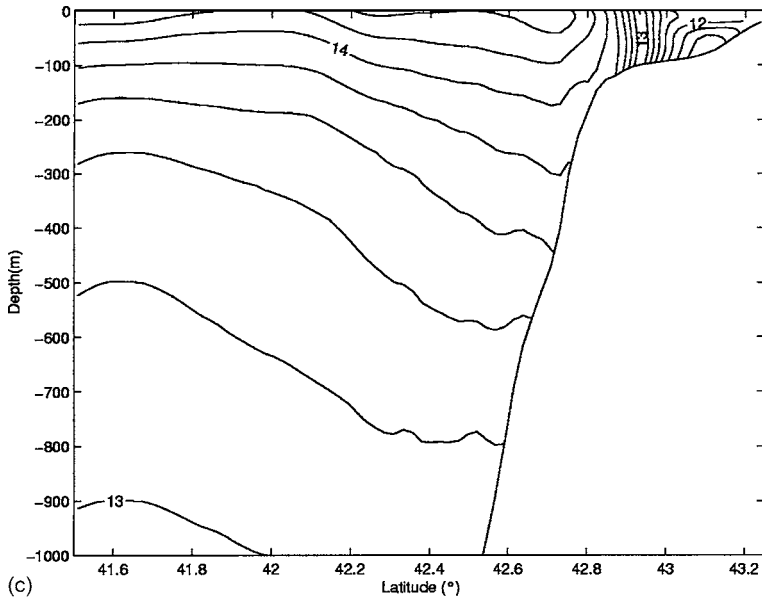


Fig. 1. (Continued).

The initial and external boundary condition fields originate from a simulation of the rigid-lid Ocean General Circulation Model (OGCM) MOM (De Mey and Benkiran, 2002), provided by the authors on the initial date. A Variational Initialisation technique (Auclair et al., 2000a,b) is then used to dynamically adjust the large scale circulation to the high

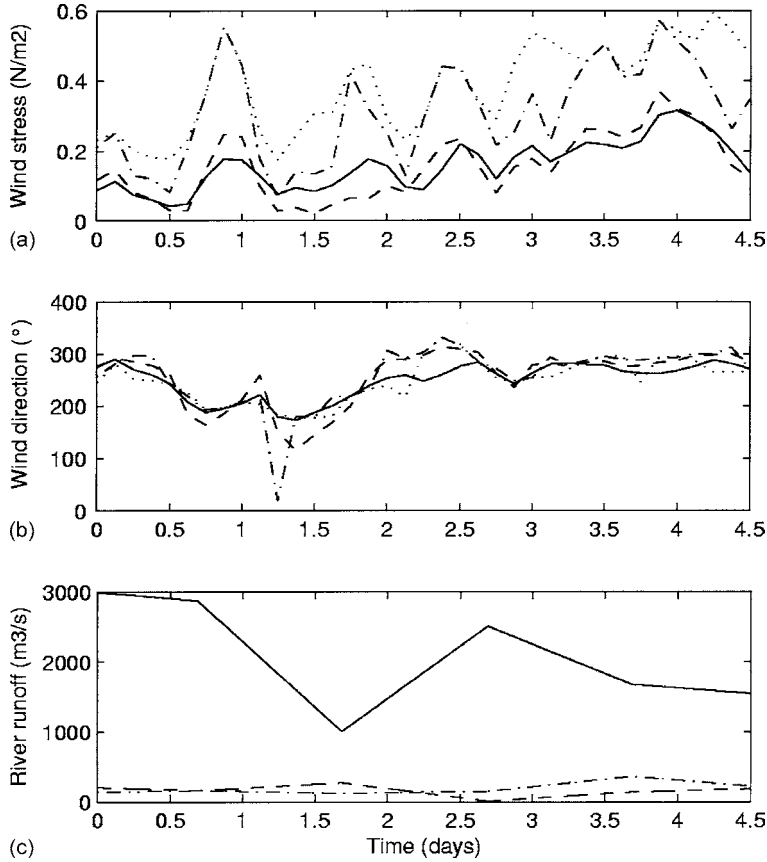


Fig. 2. Atmospheric forcing and river runoffs during the forecasted period. (a) and (b) Wind stress ( $\text{Nm}^{-2}$ ) and wind direction ( $^{\circ}$ ): solid line, mean over the domain; dashed line, mean over the shelf (north to section  $S_L$ ); dotted line, maximum over the domain; dash-dotted line, maximum over the shelf. (c) Fresh water discharge in  $\text{m}^3 \text{s}^{-1}$ : solid line, Rhône; dashed line, Petit Rhône; dash-dotted line, Hérault, Aude and Orb.

resolution bathymetry and to the free surface model constraints. The implementation of the Variational Initialisation technique is similar to the one used in Auclair et al. (2001): the external mode and the horizontal pressure gradient truncation errors are optimised. Then, in order to further adjust the plume of the Rhône river and the turbulent mixing and to avoid any spurious numerical adjustments, a 2-day spinup period ending on April 8 is carried out. Based on this atmospheric and river forcing and on the large scale circulation given by the OGCM, a Reference forecast starting on April 8, 1998 at 1830Z and ending 4.5 days latter on April 13, 1998 is carried out. The open boundary condition numerical scheme is similar to the one used in Estournel et al. (2001b). In particular, a radiative boundary condition associated with the surface elevation anomaly ( $\eta$ ) is used for the normal component of the depth-averaged velocity ( $U^{\perp}$ ):

$$(U^{\perp} - U_0^{\perp}) = \pm \sqrt{gH}(\eta - \eta_0) \quad (2.1)$$

where  $H$  is the bottom topography, and the zero subscript refers to OGCM external fields. Such a scheme prevents spurious reflections of the external mode, and also enables the specification of the external fields in 10 km-wide relaxation bands along the open boundaries for incoming conditions. A vanishing gradient is specified along the open boundary condition for the internal mode velocity. Large scale temperature and salinity fields extracted from the OGCM output fields are advected into the domain through the open boundaries under inflow conditions based on an upstream condition.

Clearly, due to the numerous uncertainties associated with the modelling procedure, many other forecast solutions could have been possible. The model variables evolve in the  $m$ -dimensional model state space as a “cloud” along the time axis. Our stochastic forecasting procedure can be viewed as the modelling of a cloud of possible states evolving with time under the influence of errors that we explicitly included, in particular those associated with the various forcings and with the initial state. Clearly, the procedure consisting in issuing only a Reference forecast misses a lot of information. The possible forecasts are more adequately described by characterising their statistical moments (their dispersion, to get started) with respect to a particular reference estimate in the model forecast error subspace: mathematical expectancy, statistical mean, median, or a particular mode if the system is multi-modal. Based on a primitive equation OGCM, (Power and Kleeman, 1993) showed that the coupled atmosphere-ocean pdf in model state space was indeed multi-modal. In this work, we will limit ourselves to the determination and EOF analysis of the first- and second-order moments. To reach such an objective, we shall explore the third- and fourth-order moments of the Ensemble pdf.

## 2.2. Ensemble generation

The modelling strategy can be described by a vector Itô stochastic differential equation:

$$d\mathbf{x}^f = M(\mathbf{x}^f, F, t) dt + G(\mathbf{x}^f, F, t) d\mathbf{w} \quad (2.2)$$

where  $M$  is the non-linear coastal ocean model and  $G$  a key non-linear function from the  $n_f$ -dimensional forcing error subspace to the  $m$ -dimensional model state space,  $\mathbf{x}^f$  the statistical variable associated to the  $n$ -dimensional Reference forecast  $x^f$ ,  $\mathbf{w}$  a  $m_f$ -dimensional Wiener process and  $F$  stands for the various model forcing. At initial time, the statistical variable  $\mathbf{x}^f$  is given by:

$$\mathbf{x}^f(0) = \mathbf{x}^b \quad (2.3)$$

where  $\mathbf{x}^b$  is the initial state statistical variable originating from the analysis of the latest model forecast at  $t = 0$  or directly from observations.

The evolution of the corresponding probability density function (pdf) is given by the Kolmogorov Fokker–Plank equation and the evolution equation of the second-order moments of the discrete pdf evolution equation can be written as:

$$\begin{aligned} P^f(k+1) &= E \left[ (x^f(k+1) - x^t(k+1))^2 \right] \\ &= \Omega(x^f(k) - x^t(k), M) + \Gamma(x^f(k) - x^t(k), \mathbf{w}(k), G) \end{aligned} \quad (2.4)$$



where  $k$  is the discrete time index,  $x^t$  the true state and  $\Omega$  and  $\Gamma$  are two non-linear functions. Explicit Taylor decompositions of  $\Omega$  and  $\Gamma$  can be found in matrix form in Canizares (1999) and in an analytical form in Miller et al. (1994). The first part of the right hand side in Eq. (2.4) ( $\Omega$ ) stands for the evolution through model dynamics of the forecast error covariance and the second part ( $\Gamma$ ) is associated with the addition of noise through the various forcings and through model inconsistencies.

The basic approximation of the Ensemble formulation of (2.4) is:

$$x^t = \lim_{n \rightarrow \infty} x_n^c \quad (2.5)$$

where  $x_n^c$  is the Ensemble mean forecast or Ensemble Central forecast for the  $n$ -member Ensemble, i.e.  $x_n^c = (1/n) \sum_{i=1}^n x_i^f = \overline{x^f}$ , the overbar standing for the Ensemble mean over the  $n$  members. This assumption raises several non-trivial limitations concerning in particular the convergence of (2.5) and the number of Ensemble forecast needed.

### 2.3. Forcing and initial state error modelling

In practice, the noise that is being added is chosen in stochastically modelling the random variables  $\overline{w}$  and  $\underline{x}^b$  (Evensen, 1994). The evolution of the Ensemble forecast error covariance  $\overline{P^f}$  through  $\Omega$  is thus implicitly carried out in the Ensemble model integration without any additional assumptions concerning the closure of the covariance evolution equation and the order of truncation of the non-linear evolution of  $\overline{P^f}$  as in the EKF.

Perturbations of the atmospheric forcing fields and of the specification of the external fields along the lateral open boundaries and of the initial state are generated in order to evaluate the forecasting capabilities of the model and to physically characterise the model forecast error subspace.

The random noise used to generate the Ensembles is Gaussian with a vanishing mean and a standard error equal to 1, and when time varying parameters are perturbed the noise is not correlated in time. Each Ensemble forecast is perturbed using a series of such random numbers  $(R_{i,q})_{n,m_q}$  where  $n$  is the number of members of the Ensemble forecast and  $m_q$  is the number of perturbed parameters used to generate the Ensemble ( $m_q \ll n$ ). The noise generator used (MATLAB) is converging fast enough to ensure that the mean of any group of 300 numbers is at least two-orders of magnitude smaller than its standard deviation. These properties ensure that the Ensemble mean of any of the perturbations is close to zero. As a consequence the posterior discrepancies between the Central Forecast and the Reference unperturbed forecast would vanish for a linear model.

As shown in the previous section, the output of the Ensemble strategy mechanically depends on the modelling of  $G$ ,  $\overline{w}$  and  $\underline{x}^b$ . In this respect, our approach is close to a sensitivity analysis using an Ensemble method, since the a posteriori pdf only contains the response of the system to the a priori pdf. However, as physicists, we are interested in covering the spectrum of the most significant errors for our model. The choice of ‘‘sensitivity’’ parameters of course depends on the domain and on the period studied. Based in particular on the results of Millot (1990) and Estournel et al. (2001a,b), the shelf of the Gulf of Lions hydrodynamics appears to be controlled primarily by the wind and the atmospheric forcing, the river discharges and the intrusions of the LPC current over the shelf. These control

parameters are randomly perturbed in turn, to generate four Ensemble forecasts made of 300 forecasts each.

The initial density field is a key parameter of the forecast since it determines, together with the surface elevation anomaly, the strength and the direction of the geostrophic component of the currents. A random perturbation of the density field offers the opportunity to evaluate the impact of the small scale processes not being present in the OGCM forecast, and the impact of the numerical noise induced by the interpolation of the OGCM solution onto the high resolution grid. Following Evensen (1994) and Echevin et al. (2000) and in order not to offset the water masses, the temperature and salinity of Ensemble A initial state are changed coherently by moving isopycnal surface up and down depending on a random but correlated displacement field. This perturbation field is correlated horizontally (respectively vertically) with a correlation length of 30 km (respectively 100 m), and uniformly vanishes in the surface mixed layer (assumed 50 m thick). Unlike what is done to generate the other Ensembles, this first type of perturbation is not dynamically adjusted, indeed the perturbed field is not a solution of the model equations. As a consequence, the initialisation is followed by a period of adjustment.

In addition to these potential errors on the density field, the OGCM shelf currents are often subject to large error amplitudes due to the low-resolution or wrong bottom topography. As a consequence, we use a new, consistent perturbation of the LPC current to open several additional degrees of freedom aimed at evaluating the impact of such errors in the coastal model. The objective of the generation of Ensemble B is to randomly vary the characteristics of the LPC current: its depth, strength and upstream offshore distance from the coastline (another idea, not followed here, would have been to directly perturb the bottom topography). To achieve this, dynamically-adjusted perturbations of the regional circulation (it is a summation of solutions of the model equations) are added to both the initial and open boundary external fields. For each OGCM forecast, three perturbation vertical density profiles (Fig. 3a) are first computed by averaging the large scale density over the three areas bounded by the 0, 100 and 400 m isobaths. The resulting profiles correspond respectively to shallow, medium and deep profiles that are specific to the studied region. This choice of these profiles is obviously not unique as several other profiles could have been considered to study other aspects of the problem. In the present case, this choice of profiles is related to a fundamental problem of the circulation in the north-western Mediterranean: the depth of the slope current (the LPC current) is a key parameter of the inflow and outflow over the shelf through its interactions with the bottom topography. In this sense we can consider that Ensemble B corresponds to a perturbation of the LPC current in so far as the resulting perturbation of the regional circulation is viewed as a consequence of the interaction of the current with the bottom topography.

A Variational Initialisation technique is then used to compute three multivariate adjusted perturbation fields for each OGCM forecast and the amplitude of these perturbations are multiplied by a random number from  $(R_{i,q})_{n,m_q}$  with  $(n = 300$  and  $m_q = 3)$  and added to the corresponding initial or open boundary external field. A description of the method used is given in Appendix A.

Ensemble C is based on random, but dynamically-consistent perturbations of the atmospheric forcing and more specifically the wind stress, which is one of the major sources of uncertainties and errors in the Gulf of Lions. In the present study, the heat fluxes

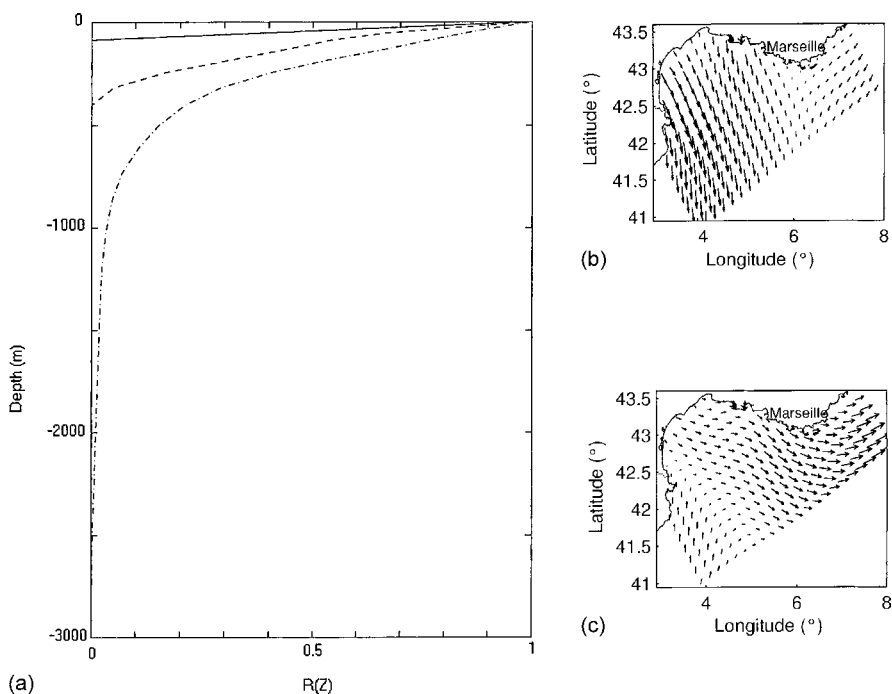


Fig. 3. Density profiles and wind perturbations used to generate respectively Ensemble B and C. (a) Profiles of the similarity function for the regions limited by isobaths 0 and 200 m (solid line), isobaths 200 and 400 m (dotted line) and off shore isobath 400 m (dash-dotted line). (b) and (c) First and second wind stress Eofs with respect to the time averaged wind stress.

were not perturbed in order to clearly separate the consequences of each forcing on the forecast and to open only a limit number of degrees of freedom for each Ensemble. However, they might be as important as the wind stress and should thus be perturbed in the future. First of all, the 10 dominant Empirical Orthogonal Functions (Eofs) of wind stress with respect to its time average over the forecasting period are calculated (Fig. 3b and c). Every 3 h (the forcing period), the amplitude of the perturbations can be written as:

$$\mathfrak{J}_{j,k'} = R_{i,j \times k'} \times \frac{\sigma_j}{\sum_{j=1}^{10} \sigma_j} \quad (2.6)$$

where  $\sigma_j$  is the singular value associated to the  $j$ th wind stress Eof and  $k'$  the time index. A linear combination of these Eofs with random amplitude is then added to the original wind stress. It can be noted that the very first Eof (explaining about 63% of the total variance) is characteristic of the northern wind blowing after Day 3 (Fig. 3b), Eof 2 (18% of the total variance) being associated to the variation of its direction (Fig. 3c). Using this Eof perturbation technique, it is thus possible to selectively vary the amplitude of the wind components without creating unrealistic wind conditions.

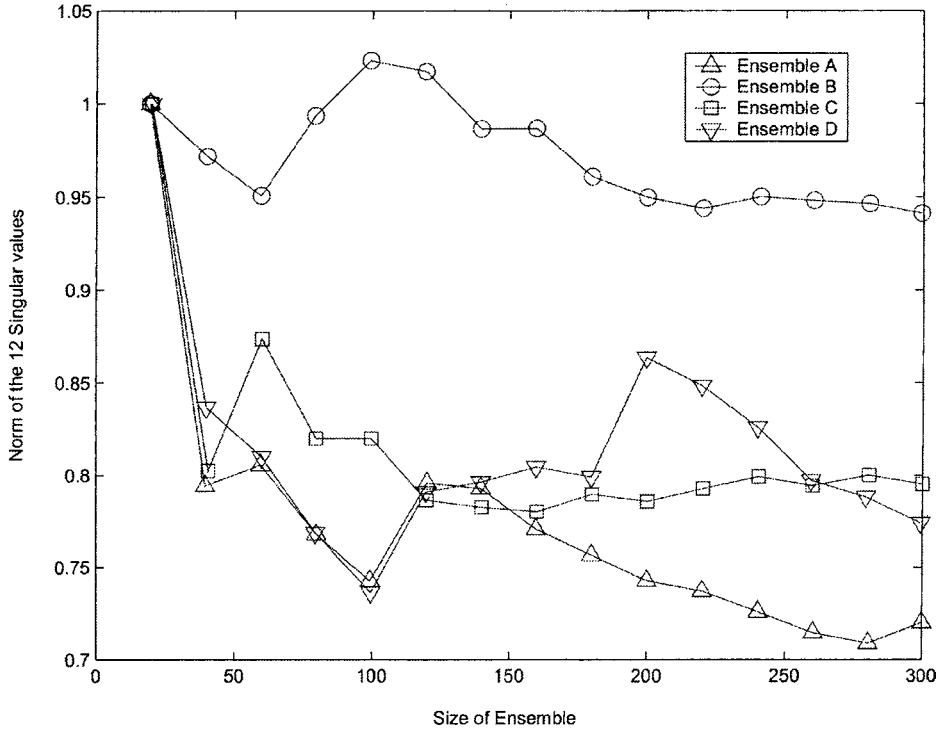


Fig. 4. Evolution with time of the norm of the first 12 singular values for Ensembles A–D.

Ensemble D is generated by adding random noise to the Rhône, Petit Rhône and Aude river total fresh water discharge. The amplitude of the perturbations is proportional to the river runoff and a random number taken from  $(R_{i,q})_{n,m,q}$ .

For each of these four perturbations of the forcing fields, aimed at approximating four different a priori error pdfs, 300 members are generated in the form of 4.5-day forecasts. The calculation of so many forecasts was made possible thanks to the use of a cluster of 16 PCs under LINUX with 512Gb of RAM each. One of the powerful characteristics of the Ensemble approach is the straightforward use of multiprocessing.

Fig. 4 shows the evolution of the norm of the first 12 singular values of the Ensemble forecast multivariate covariance matrix with respect to the Central forecast as functions of the number of computed forecasts for each Ensemble. Each forecast is characterised by its temperature, salinity, surface elevation anomaly and velocity fields, i.e. about 567,000 variables. Depending on the characteristics of the added perturbation, the number of required forecasts in each Ensemble to reach convergence varies between 100 and 200, Ensembles B and D converging the slowest. However, the structure of the corresponding left singular vectors is found to converge faster: it can indeed be obtained with about 100 forecasts for each one of the four Ensembles. To generate the present Ensembles, no particular threshold has been imposed, but an efficient way to accelerate the convergence is for instance to reject

among the random numbers  $(R_{i,q})_{n,m,q}$  those whose absolute value is larger than three times the specified variance.

### 3. First- and second-order moments of the Ensemble forecasts

#### 3.1. Reference forecast and Ensemble Central forecast

The transversal section of the temperature are presented on Fig. 5 for the Reference forecast and for the Ensemble A Central forecast. While the surface elevation anomalies are rather similar (not shown), large discrepancies can be found between the forecast temperatures. The signature in temperature of the LPC current and the most salient physical structures have in particular been smoothed out in Ensemble A Central forecast. In comparison, Ensembles B, C and D Central forecasts are very similar to the Reference one (not shown).

An interesting conclusion that can be drawn from this comparison of the temperature fields is that when the Central forecast is obtained by perturbing members with a random noise which is not dynamically adjusted (Ensemble A), it exhibits a large smoothing of the main physical structure whereas when well-adjusted perturbations are randomly added to the same initial field (Ensemble B), the Central forecast is very close to the Reference forecast. The similarity of the Central and Reference forecasts is obviously not a validation criterion. Indeed they are only supposed to be identical if the model is linear which is not the case in the present study. However, the dampening of the coherent physical structures in Ensemble A seems to indicate that the perturbation strategy is responsible for the smoothing of the spatial energy spectrum, the density perturbations showing a strong component perpendicularly to the attractor. On the other hand, Ensemble A Central forecast for the surface elevation anomaly compares well with the Reference forecast, which seems to indicate that only the randomly perturbed baroclinic component is concerned. The addition of spatially coherent density perturbations which are not dynamically adjusted, i.e. which generate large adjustments during the initialisation period, leads to a partial destruction of the LPC current density structures through the generation and the dissipation of baroclinic gravity waves. The comparison of the evolution of Ensemble A temperature and salinity fields with the same Reference fields (not shown) indicates that these spatially coherent density perturbations disappear during the very first inertial period. However, the density structure of the LPC current is obviously not recovered during the same period. No reflection of these gravity waves have been evidenced along the open boundaries.

#### 3.2. Ensemble variance

The second-order moments for Ensembles A, B, C and D (Fig. 6) show that the variance of the first two Ensembles remains much larger and spread than the Ensemble variance of the latest two after 4.5 days. Ensemble A exhibits a larger variance over the shelf break and its canyons (Fig. 6a and b) and more specifically in the region of the large density gradient separating the cold and fresh winter shelf waters from the LPC warm and salty waters (see Fig. 1c). Three maxima can in particular be observed above the canyons which

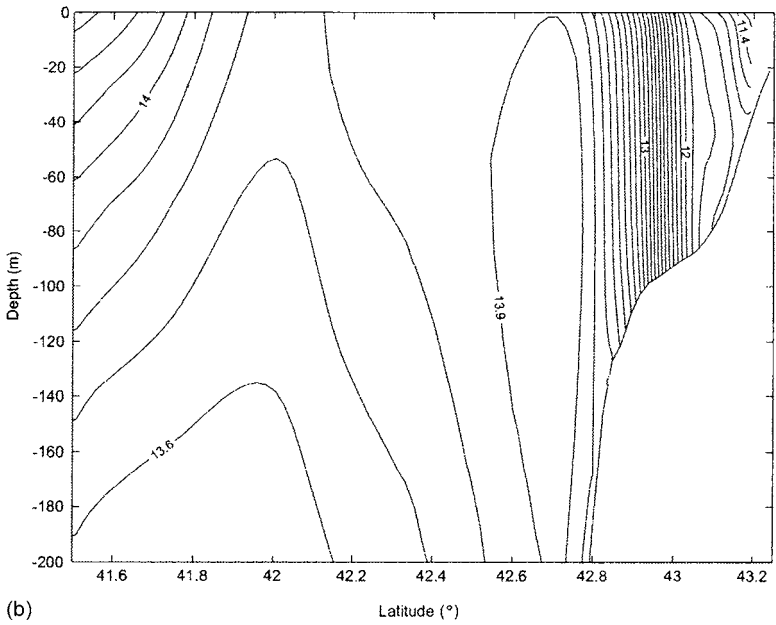
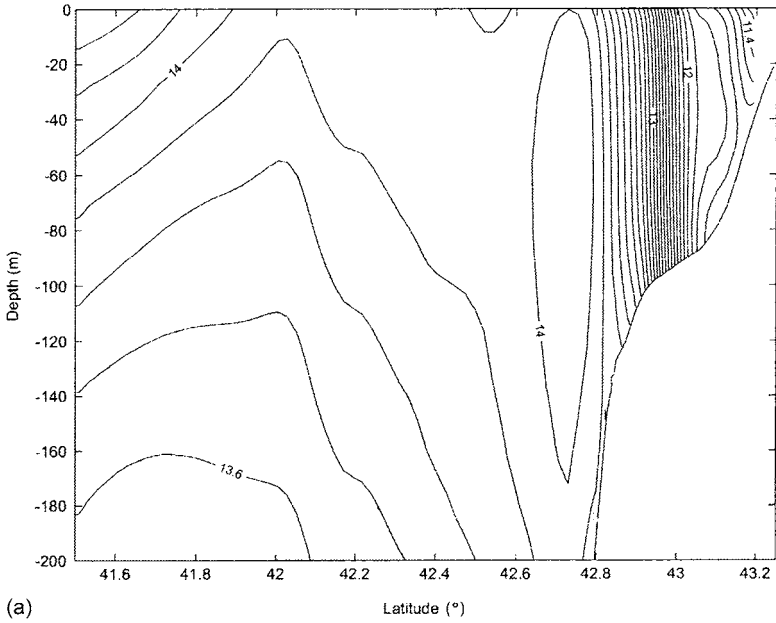


Fig. 5. Forecasts for 13 April, 0730Z: transversal section of temperature ( $^{\circ}\text{C}$ ). (a) Reference forecast. (b) Ensemble A Central forecast.

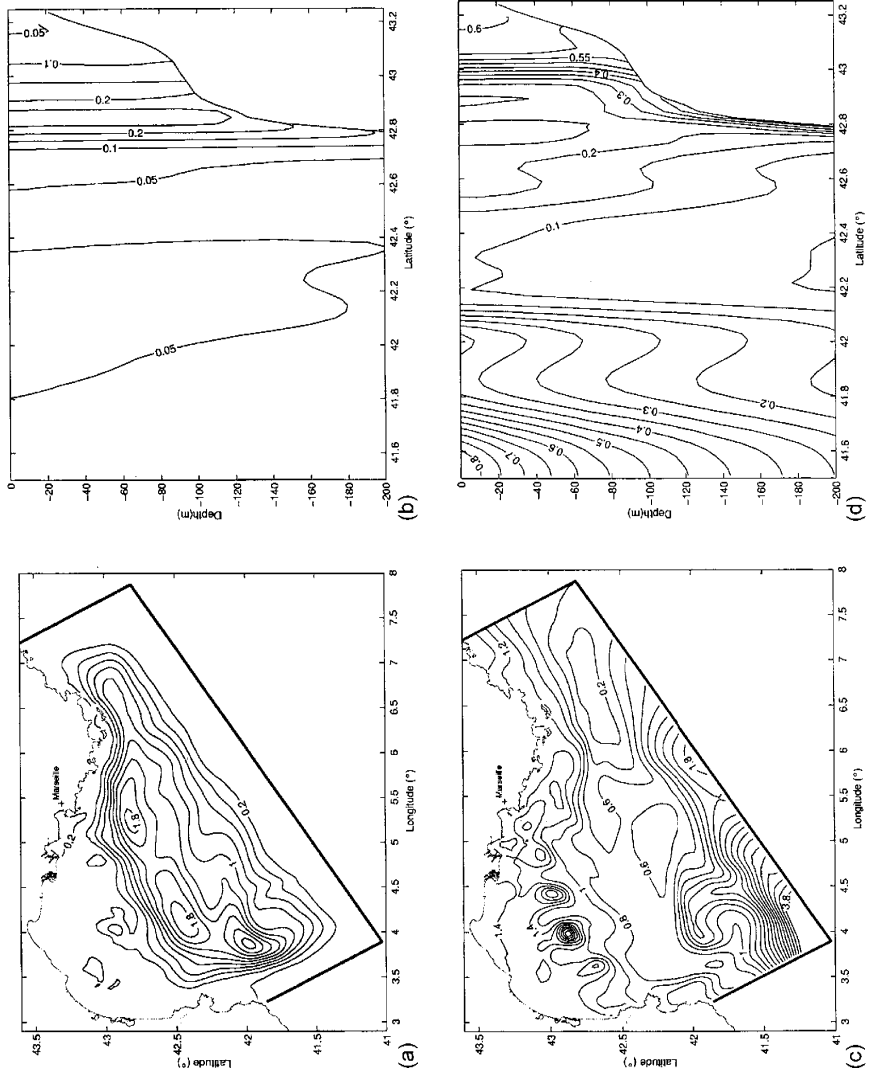


Fig. 6. Ensemble standard deviation for 13 April, 0730Z: surface elevation anomaly (cm) and transversal section of temperature ( $^{\circ}\text{C}$ ). (a) and (b) Ensemble A. (c) and (d) Ensemble B. (e) and (f) Ensemble C. (g) and (h) Ensemble D.

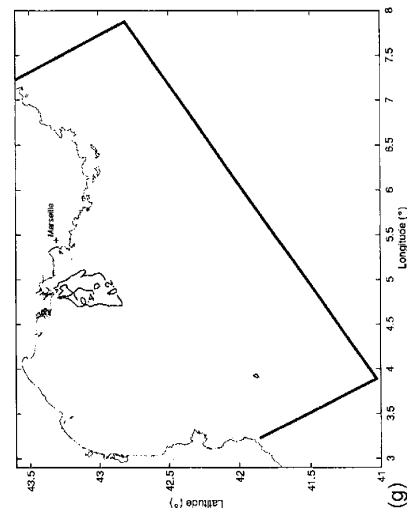
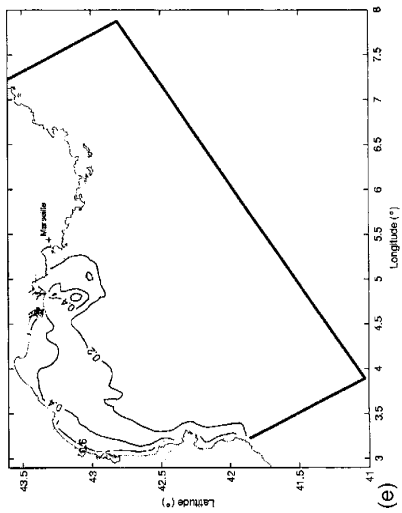
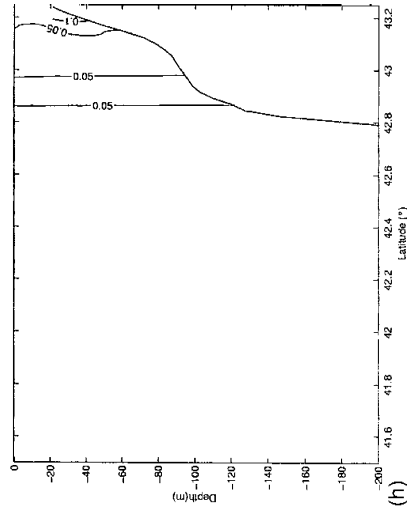
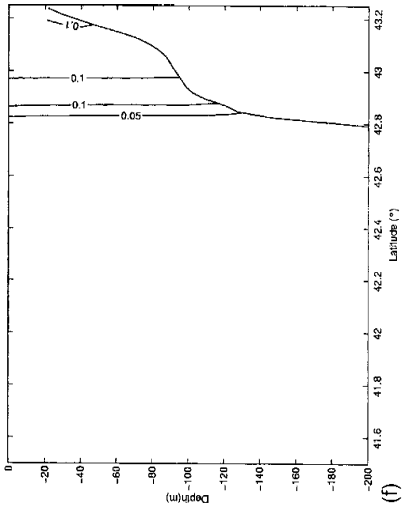


Fig. 6. (Continued).



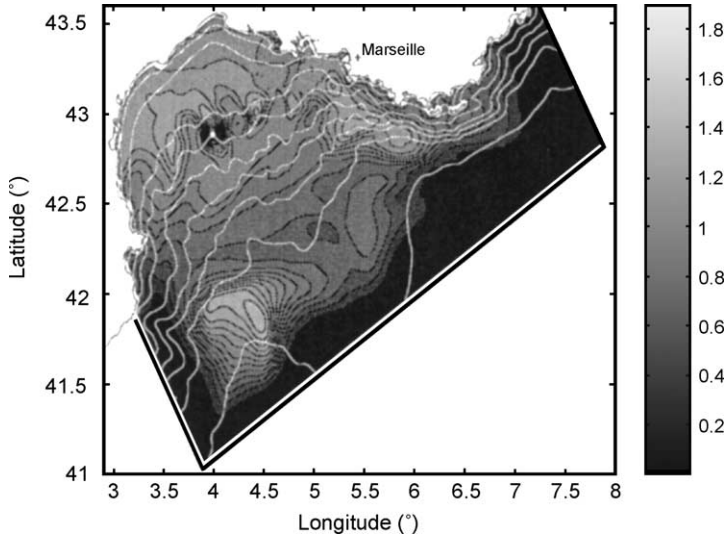


Fig. 7. Ensemble standard deviation for 13 April, 0730Z for the surface elevation anomaly (cm) in Ensemble B'.

are spread over the shelf break between isobaths 500 and 1000 m. Every forecast having the same exact boundary conditions, a region of very low variance can be observed along the open boundaries (Fig. 6a and b). In Ensemble B, the maximum variance is found along the open boundaries for both temperature and surface elevation anomalies. A large variance of the LPC current surface signature can in particular be observed at the eastern and western boundaries of the domain (Fig. 6c) together with a shift of its temperature signature (Fig. 6d). In order to understand the impact of the open boundaries, Fig. 7 shows the Ensemble variance for a fifth experiment similar to Ensemble B (named Ensemble B') except that the boundary conditions are the same for each member of Ensemble B'. The relaxation toward unperturbed external fields at the open boundary results in a vanishing variance of the surface elevation anomaly and of the temperature along this open boundary in Ensemble A and in Ensemble B', while non-vanishing covariance are observed in Ensemble B when the external fields are perturbed. Fig. 6d clearly displays several distinct regions with large Ensemble temperature variance: along the open boundary, at the southern edge of the LPC current and over the shelf break. Several maxima can also be found above the main canyons at the Gulf of Lions shelf break but, unlike in Ensemble A, they are located in shallower waters (between the isobaths 100 and 500 m), in the shelf break winter density front. As could be expected, the perturbation of the wind field (Ensemble C) has a maximum impact on the shelf processes: the plume of the Rhône river and its spreading along the Roussillon coast (Fig. 6e and f). Finally, most of the variance in Ensemble D is in the plume of the Rhône river (Fig. 6g and h), as could be expected. At the scale of 4–5 days, the perturbation of the plumes (Ensemble D) has only a slight impact on the shelf break winter horizontal density gradient (Fig. 6g).

The amplitude of the Ensemble variance is, to a great extent, proportional to the amplitude of the perturbation. So, in order to compare different Ensembles, the corresponding variance

should for instance be normalised by the total energy of the perturbation. This was not done for the plots on Fig. 6 and they are thus only indicative for each Ensemble of the dispersion of the “clouds” of forecasts. The perturbations of the initial and open boundary external fields have thus a greater impact on the resulting Ensemble variance at least at the time scale of the present forecasting period, i.e. 4.5 days. We can indeed easily presume that longer forecast periods would see the impact of the initial state perturbation decrease while the consequences of the perturbation of the various forcings would increase with time. This is confirmed by the evolution of the model forecast error displayed in Section 5.

The displacement of the maxima of variance inside the canyons between Ensemble A and B can be explained by the perturbation strategy. In Ensemble A, the density field is perturbed everywhere below 50 m whereas in Ensemble B only the structure of the LPC current is disturbed, i.e. in the first 500–800 m. In this latter case, the perturbations of the upsloping and downsloping currents inside the canyons together with the resulting barotropic adjustment (via the surface elevation anomaly) are thus found in shallower waters (Fig. 6a and c). The final forecast comes after a strong Tramontane event that induces an anticyclonic circulation over the western part of the shelf (Estournel et al., 2001a,b). This circulation is constituted of a south-easterly current over the shelf together with a strong northerly coastal jet along the Roussillon coast. A longitudinal velocity section across the canyons (not shown) indicates that the water exchanges through the shelf are located inside the canyons and corresponds to the upsloping and downsloping schemes given by Klinck (1996). By Tramontane wind conditions the upsloping to the shelf appears over the western edge of the canyons while the downsloping out of the shelf is located over the eastern side. These signatures on the surface elevation anomaly can be seen on Fig. 6a with positive anomalies over all three main canyons. Depending on the position, thickness and strength of the LPC current the amplitude of these exchanges varies among the members of the Ensemble leading to a maximum in Ensemble variance (Fig. 6c).

Another region with maximum Ensemble variance can eventually be found near Marseille. This corresponds also to an interaction between the LPC current and the bottom topography where the slope current passes the Hyères Islands. This anomaly reaches a maximum by south-western wind, i.e. before Day 2. Indeed, in this case, the LPC current enters the shelf through the wide Grand Rhône and Marseille canyons (the easternmost canyons). Ensemble A variance maxima (Fig. 6a) are located offshore and correspond to the amplitude of the surface elevation anomaly signature of the LPC current stationary meanders in this region. The core of the LPC current is located between isobaths 1000 and 2000 m and follows the bottom topography of this region and in particular the deepest part of the main three canyons. Since, in Ensemble A, density is perturbed from 50 m to the bottom, the LPC current shows important differences in its interaction with the bottom topography in this deep region. This results in large Ensemble variance of the surface elevation anomaly. As a consequence the physical processes leading to the Ensemble variance maxima in Ensembles A and B are similar but the locations differ due to the perturbation strategy. In Ensemble A only the offshore interactions are perturbed and the fluxes through the shelf are more or less similar among the members while in Ensemble B the penetration of the LPC current over the shelf is significantly perturbed.

The maximum in temperature variance over the shelf break (Fig. 6b, d, f and h) can be explained by a displacement of the steep temperature gradient located in winter over the

shelf. This latest observation can be made for all type of perturbations (Ensembles A–D) leading to the conclusion that the exact location of this horizontal gradient is a rather critical diagnostic in model forecast. This corresponds to the northern edge of the LPC current above the shelf break and the Ensemble variance maximum is a consequence of the interaction of the LPC current together with the bottom topography and the winter cold and fresh shelf waters.

As far as the perturbation of the open boundary conditions is concerned, a comparison of Figs. 6c and 7 shows that after 4.5 days the influence of the specified external conditions of the LPC current along the north-eastern boundary is felt all the way to the Hyères islands. Indeed, Fig. 7 shows very low Ensemble variance anomalies in this region while Fig. 6c exhibits large Ensemble variance anomalies. This gives an idea of the importance of the specification of the initial conditions. One could think that after a very strong wind event such as the Tramontane event between Day 2 and the end of the forecast, the circulation over the shelf of the Gulf of Lions is more or less totally controlled by the atmospheric forcing (Estournel et al., 2001b). This is, however, not totally true insofar as the LPC current still controls penetration over the shelf break through the canyons and as a consequence influences the circulation over the shelf. An interesting consequence is that even with a perfectly-known very strong offshore blowing wind over the shelf one cannot hope to forecast the circulation over the shelf without a good general circulation.

### 3.3. Skewness and Kurtosis

Based on the information given by the Ensembles, it is also possible to evaluate the “linearity” of the main dynamical error processes involved during the period studied. An interesting diagnostic concerns the higher-order moments and in particular the statistical Skewness (third-order) and Kurtosis (fourth-order) based on these moments (D’Agostino, 1986). These diagnostics offer a measure of the divergence with respect to Gaussian statistics. Indeed the larger the absolute value of the Skewness the more asymmetric the distribution, the larger the Kurtosis the larger the peakiness of the distribution and the smaller the Kurtosis (negative) the larger the spreading. The Skewness can be defined by:

$$\bar{S}(k) = \frac{n \times \overline{M}_3(k)}{(n-1)(n-2)\bar{\sigma}^3} \quad (3.1)$$

where the Kurtosis by:

$$\bar{K}(k) = \frac{n(n+1)\overline{M}_4(k) - 3(n-1)(\overline{M}_2(k))^2}{(n-1)(n-2)(n-3)\bar{\sigma}^4} \quad (3.2)$$

where  $k$  is the discrete time index,  $\bar{\sigma}$  the evaluated standard deviation and the  $j$ th-statistical moment  $\overline{M}_j$  is defined by:

$$\overline{M}_j(k) = \sum_{i=1}^n (x_i^f - x_n^c)^j. \quad (3.3)$$

Fig. 8 shows the Skewness and Kurtosis for each Ensemble forecast for the surface elevation anomaly. For Ensemble A, Fig. 8a and b exhibit an important patchiness. The

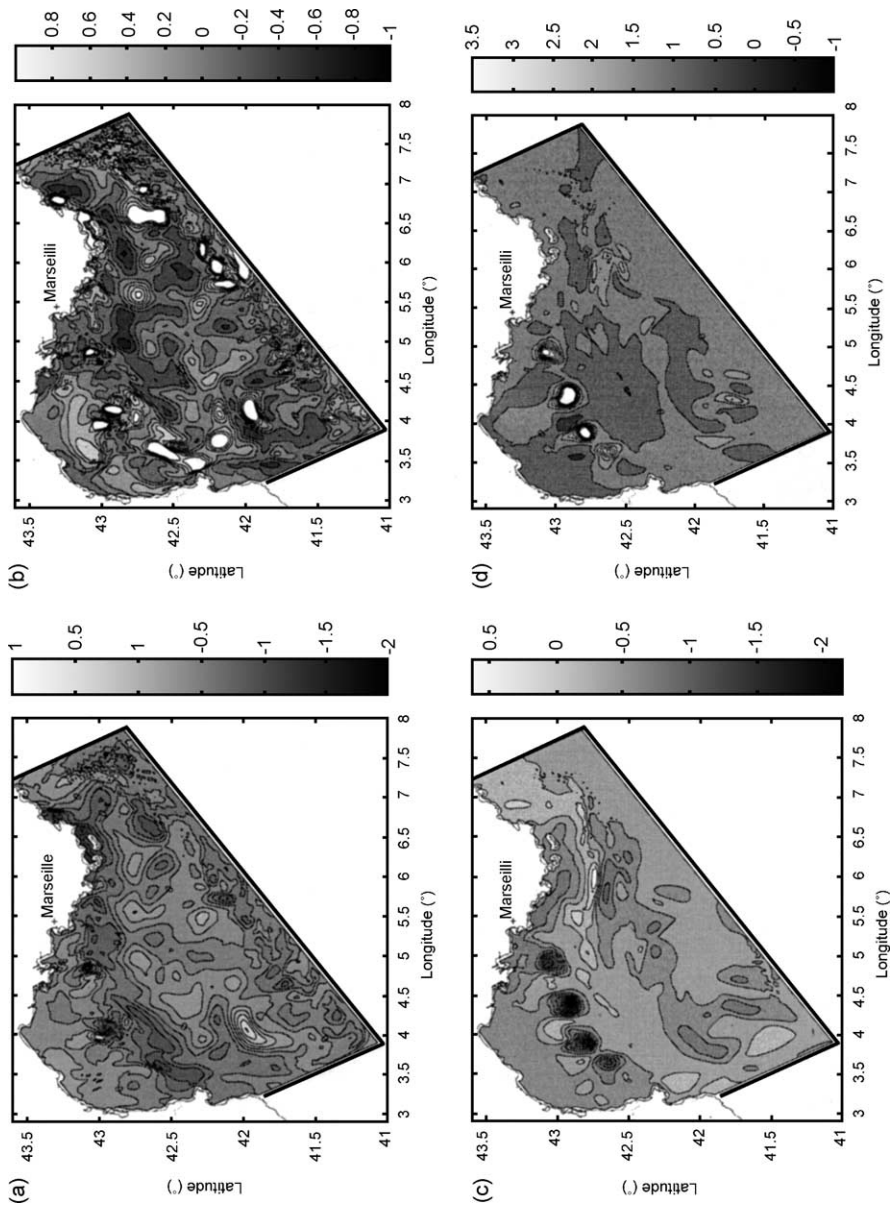


Fig. 8. Ensemble Skewness (a, c, e, f) and Ensemble Kurtosis (b, d, f, h) for 13 April, 0730Z and for surface elevation anomaly. (a) and (b) Ensemble A. (c) and (d) Ensemble B. (e) and (f) Ensemble C. (g) and (h) Ensemble D.

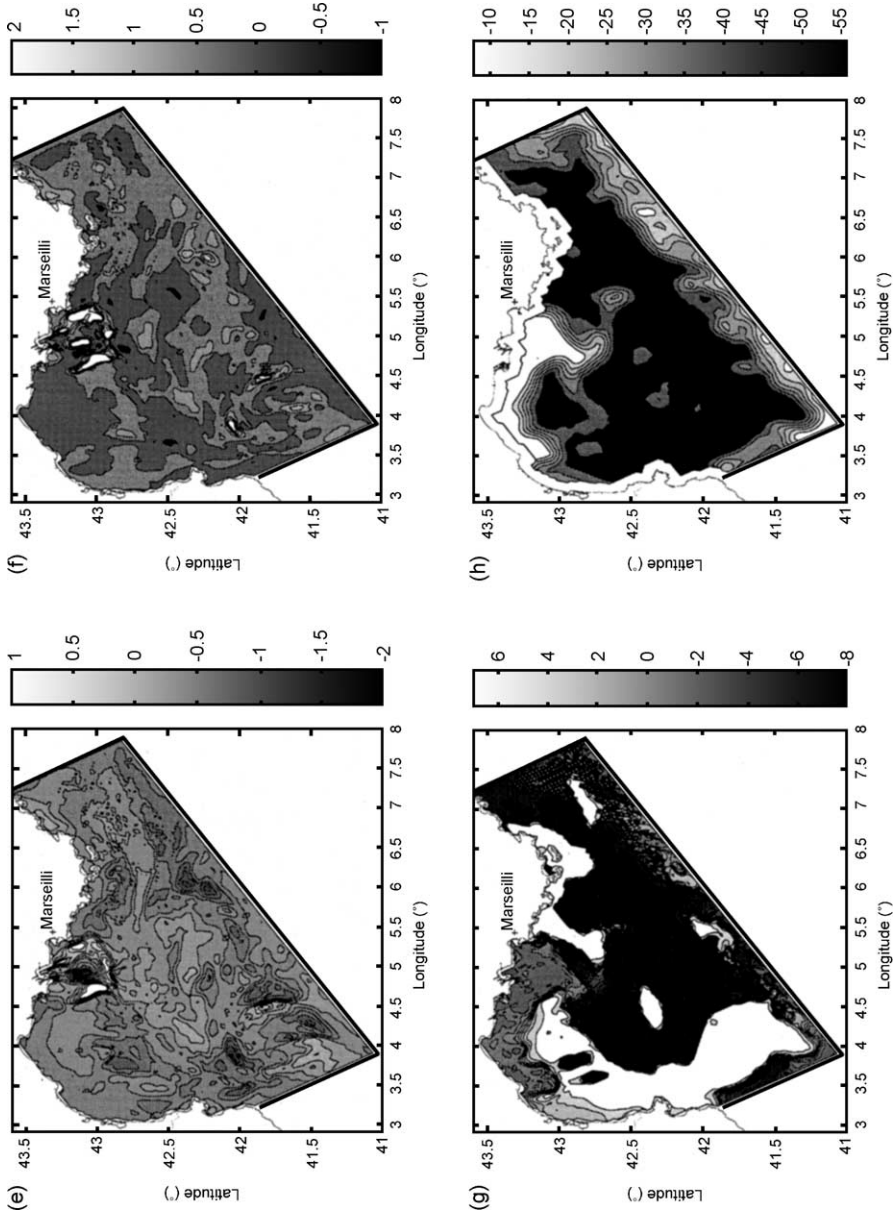


Fig. 8. (Continued).

Skewness is negative along the northern edge of the LPC current and positive along its southern edge which is largely confirmed by the transversal sections of the temperature Skewness and Kurtosis (not shown). The Skewness and Kurtosis for Ensemble B (Fig. 8c and d) are globally much smaller except over the canyons where they both reach larger values. When the wind stress is perturbed with Gaussian noise (Ensemble C), the higher values of Skewness and Kurtosis can be found in the Rhône river plume and at the bottom of the Cabo Creus and Lacaze–Duthiers canyons (Fig. 8e and f). In these regions, the pdf is peaked and diverges slightly from the Gaussian distribution. Ensemble D exhibits the largest divergence with respect to the Gaussian distribution. Except in the plume region, the Skewness (Fig. 8g) shows very large values with a tendency toward lower values offshore and a tendency toward higher values in the shelf break region. As no perturbation other than the river discharges has been implemented, the Kurtosis (Fig. 8h) displays a very peaked distribution: all the members have nearly the same value of the surface elevation anomaly. Along the open boundaries the distribution is very close to a Gaussian distribution which can most probably be explained by the accumulation of external gravity waves generated by the perturbation of the river discharge. We can note the Gaussian structure of the plume regions and, as a consequence, the linearity of the processes involved in its spreading.

The pdf for Ensemble A is tilted toward smoothed. This corresponds to a decrease of both the barotropic and baroclinic horizontal gradients sustaining the geostrophic component of the LPC current (Fig. 8a). This effect is to be associated with the observations concerning the smoothing of the LPC characteristics in the Ensemble A Central forecast. The predominantly negative Kurtosis confirms the general trend toward a smoothing of the Ensemble probability density function with only small-extent regions of positive Kurtosis. This also confirms the importance of the generation and dissipation of the baroclinic gravity waves in this experiment. On the opposite the Ensemble B surface elevation anomaly Central forecast exhibits the greatest global agreement with the Gaussian distribution. On top of the canyons however non-linear behaviours have large amplitude, the distribution is peaked and asymmetric. This is clearly related to the forcing by the topography which is larger and controls the current penetration onto the shelf.

The maxima observed for Ensemble C can also be explained by a very strong forcing: the river fresh-water discharge in the plume region and by the bathymetry in the canyon. The general tendency in Ensemble C is the spreading of the distribution in the Ekman layer due to the perturbation of the wind stress (Fig. 8f). This is confirmed by the vertical structure of the temperature Kurtosis (not shown) with low values above the surface (around  $-1$ ). This cannot be a direct and linear effect of the wind stress perturbation which is Gaussian in nature, but can be explained by the non-linear turbulent mixing particularly large during the strong wind event following Day 3. It would not be very surprising that in such regions an assimilation scheme based on a linear filter gives only very poor results if the model errors contain such processes.

As a consequence the higher-order moments give interesting hints concerning the limitations of any Bayesian approach associated to the Gaussian approximation. The canyons and more generally the bathymetry appeared to be associated with peaked and asymmetric distributions.

This local discrepancy with regard to the Gaussian approximation clearly deserves further investigation. Indeed this means that locally the interpretation of the variance is not trivial,

and the significance of the usual 67% confidence interval for a Gaussian pdf has to be replaced by some other statistical approach. This is not the central focus of the present paper and to further investigate the statistics of the distribution we chose to study a special Gaussian approximation to this locally non-Gaussian distribution. It is trivially obtained by truncating the higher-order moments and keeping unchanged moments of orders 1 and 2 (although we agree that this could have been achieved differently). The second-order moments of this “ideal Gaussian” distribution are studied by computing the singular values and the singular vectors of the covariance matrix.

#### 4. Physical description of the departures from the “Central forecast”

We now explore the Ensemble covariances by means of multivariate Central Empirical Orthogonal Functions. By “Central”, we mean that statistics are computed with respect to the Central forecast. In this section, the dominant Ceofs are used to characterise the dynamical structures that are spatially coherent in the forecast error.

The first 12 Ceofs have been computed for each Ensemble using a Lanczos technique (Toumazou and Créteaux, 2001). The surface elevation anomaly for each Ceof together with a transversal or longitudinal section of temperature are presented in Figs. 9–12 for two

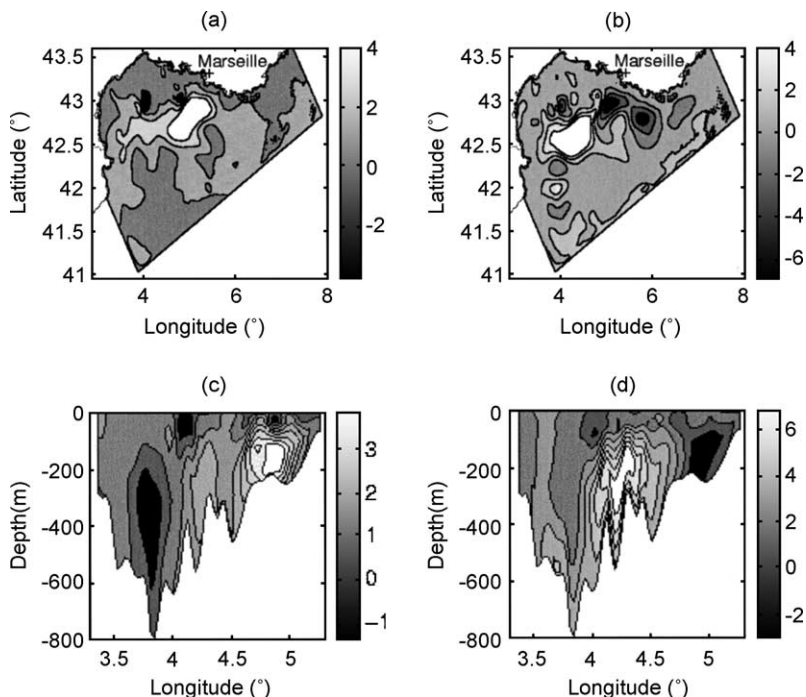


Fig. 9. Multivariate Central Empirical Orthogonal Functions 1 and 2 for 13 April, 0730Z and for Ensemble A (Normalised Variance Units). (a) and (b) Surface elevation anomaly. (c) and (d) Longitudinal section of temperature.

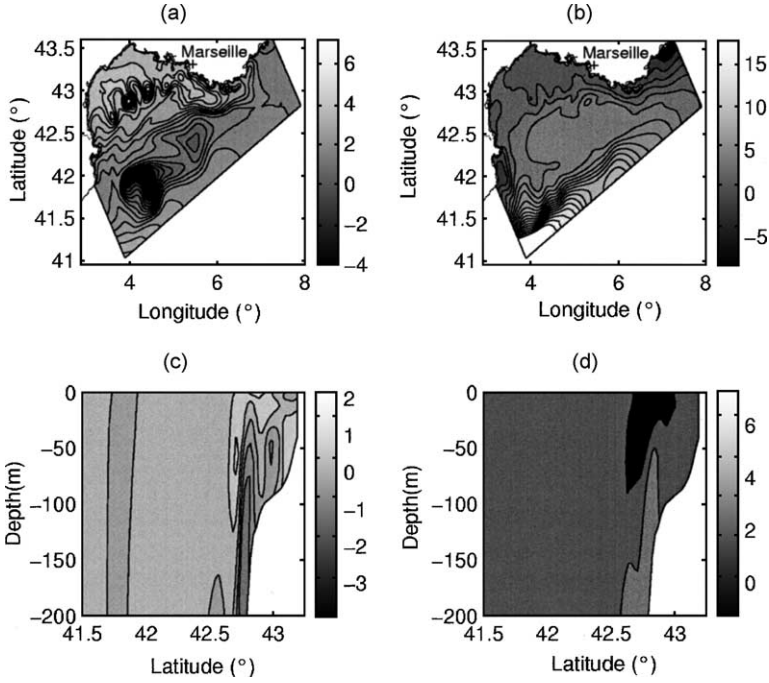


Fig. 10. Multivariate Central Empirical Orthogonal Functions 1 and 2 for 13 April, 0730Z and for Ensemble B (Normalised Variance Units). (a) and (b) Surface elevation anomaly. (c) and (d) Transversal section of temperature.

Ceofs. Each Ceof is normalised and as a consequence the amplitude of the variance is not meaningful.

The corresponding Central singular values are given for each Ensemble in Table 2. The spectrum of Ensemble A singular values is nearly white no Ceof explains a large percentage of variance. At the opposite Ensemble B exhibits two dominant modes explaining a large percentage of variance. For each Ensemble, two Ceofs out of the 12 computed are presented. They have been chosen based on their physical consistency rather than on the percentage of variance they explained. Their rank is given among each Ensemble Ceofs.

Ensemble A Ceofs 1 and 2, as the remaining 10 Ceofs, display alternate maxima and minima of the variance of the surface elevation anomaly and the temperature located along

Table 2  
Percentage of variance explained by each Ceof for the four Ensembles

Ceof	1	2	3	4	5	6	7	8	9	10	11	12
Ensemble A	3.26	2.74	2.57	2.41	2.35	2.29	2.15	2.07	1.98	1.80	1.77	1.74
Ensemble B	57.37	15.32	6.91	3.89	3.12	1.90	1.58	1.11	0.89	0.70	0.57	0.51
Ensemble C	19.29	16.55	11.22	6.47	5.07	3.98	3.70	2.47	2.19	2.10	1.70	1.56
Ensemble D	20.32	15.63	14.55	8.08	6.72	5.43	4.67	3.16	2.73	2.20	1.74	1.53



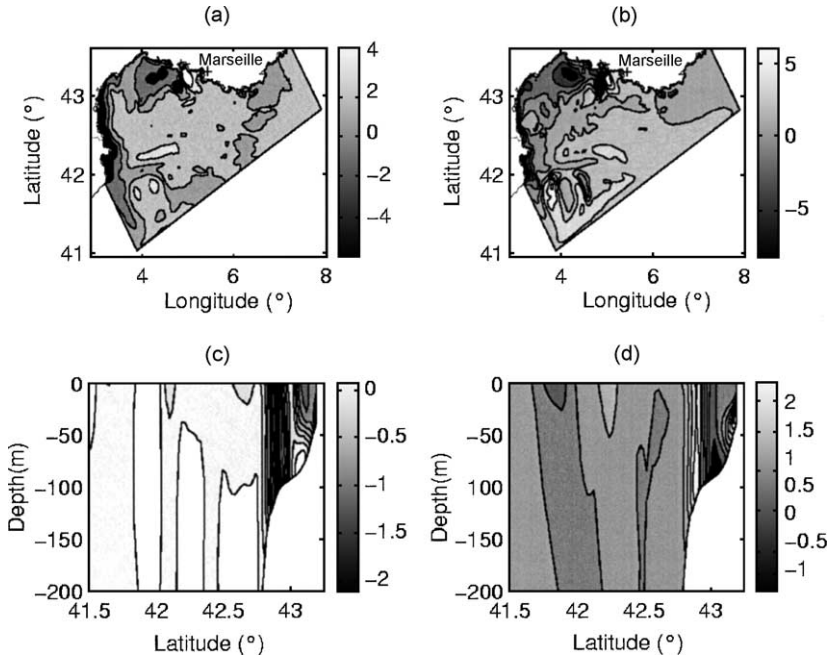


Fig. 11. Multivariate Central Empirical Orthogonal Functions 1 and 2 for 13 April, 0730Z and for Ensemble C (Normalised Variance Units). (a) and (b) Surface elevation anomaly. (c) and (d) Transversal section of temperature.

the shelf break above the canyons (Fig. 9). The temperature section (Fig. 9c and d) indicates in particular that these extrema are located above the lateral slopes of the canyons and the corresponding transversal section of the Ceof horizontal velocity shows that exchanges occur in opposite directions depending on the slope of the canyon considered. The perturbation of the density field has thus a significant impact in controlling the interactions of the LPC current with the canyons. Klinck (1996) showed that a crucial parameter controlling the generation and the amplitude of the upsloping or downsloping is indeed the strength of the vertical density gradient together with the main geometrical characteristics of the canyon. Depending on the perturbation, the LPC current deflection is changed at the canyon upstream (downward deflection) and downstream (upward deflection) edges. A depth integrated vorticity balance (not shown) confirms this conclusion by exhibiting positive vorticity anomaly over the upstream edge and negative vorticity anomaly over the downstream edge.

Ceofs 1 and 2 are presented for Ensemble B on Fig. 10. Ceof 1 (Fig. 10a and c) shows large amplitude values over the canyons superimposed on a transverse surface elevation anomaly gradient with positive values over the shelf and negative off shore. Fig. 10c confirms the location of the temperature anomaly over the canyons following the processes studied by Klinck (1996). The surface elevation anomaly further confirms that the Ensemble variance maxima are located in shallower waters. Interestingly enough the Ceof decomposition enables us to separate clearly the influence of the initial condition perturbation (Ceof 1) from

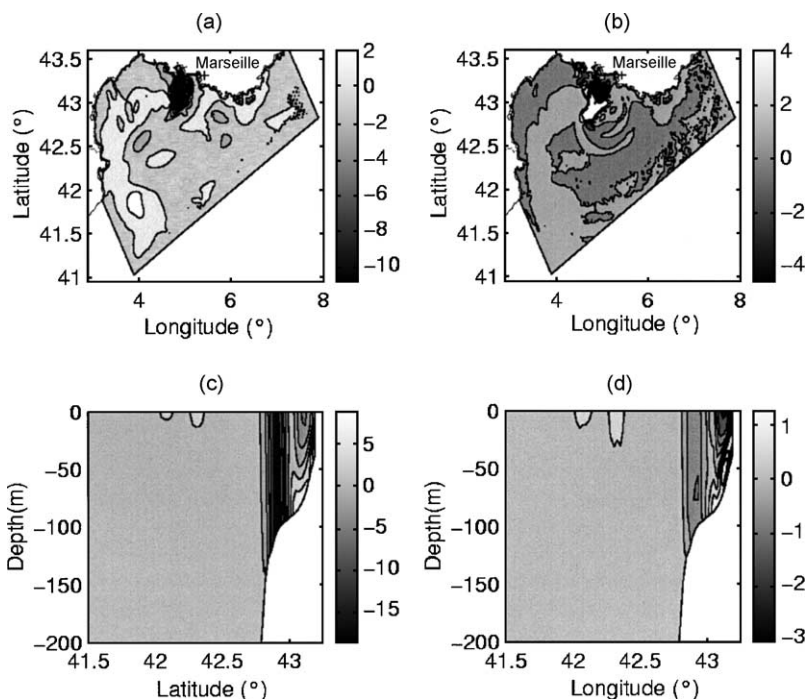


Fig. 12. Multivariate Central Empirical Orthogonal Functions 1 and 5 for 13 April, 0730Z and for Ensemble D (Normalised Variance Units). (a) and (b) Surface elevation anomaly. (c) and (d) Transversal section of temperature.

the influence of the open boundary conditions. Ceof 1 shows large amplitude maximum over the canyons whereas Ceof 2 is related to the open boundary conditions and shows large gradient of the surface elevation anomaly along the north-western and south-eastern open boundaries. This is correlated in the interior of the domain with a transverse surface elevation gradient and over the shelf with an amplification of the structure corresponding to the LPC current inflow. Both Ceofs 1 and 2 have a strong baroclinic temperature structure over the shelf break (Fig. 10d for Ceof 2) which is related to the temperature front observed in winter at this location.

Ensemble C Ceofs 1 and 2 (Fig. 11) deal with the observed correlation between several error processes when the wind stress is perturbed: the spreading and position of the plume of the various rivers (Fig. 11a and c), the amplitude of the upwelling along the western Roussillon coast (Fig. 11a and c), the interaction of the LPC current with the Cabo Creus and Lacaze–Duthiers canyons in the south-western part of the domain (Fig. 11c), the position of the temperature gradient over the shelf break and the vertical structure of the temperature field over the shelf (Fig. 11b and d). As a consequence, an erroneous specification of the wind stress should result in a bad forecast of the amplitude and occurrence of these processes.

Ensemble D Ceofs 1 and 5 (Fig. 12) are associated with the Rhône river. Ceof 1 (Fig. 12a and b) is directly associated with the strength of the plume while Ceof 5 (Fig. 12c and d)

shows interesting concentric structure associated with the propagation of the information related to the perturbation of the river runoffs through external gravity waves. The transversal section of temperature shows a large temperature variance above the shelf break still associated with the horizontal winter gradient. Such transient processes are obviously very dependent on the date of the forecast and the Ceof 5 is thus localised in time.

Each Ceof has consequently a strong physical interpretation. It can be associated with a particular physical error process (or several error processes emanating from a single error source) and can therefore be useful in a predictability study. For instance, if the wind forcing is associated with large errors, then the forecasting of the plume position and extension will necessarily be associated with strong errors and Ensemble C Ceofs 1 and 2 (Fig. 11) will show where these errors are likely to be located and where the model is to be corrected in an assimilation configuration. The correlation of the general circulation with the shelf processes appears clearly. Ensemble A and more specifically Ensemble B Ceofs clearly indicate that the canyons are the main locations of the exchanges through the shelf break. They also show that these exchanges are restricted to the western side of the canyon in the direction of the shelf (upsloping) and to the other side in the other direction (downsloping). However, nothing indicates that this relation between the Ceofs and the dynamics is a one to one relation. Indeed Ensemble C Ceof 1 is only supposed to explain the largest part of the Ensemble variance among Ensemble C Ceofs. This obviously does not mean that some other Ceof cannot explain part of the same physical process.

Each of these Ceofs can easily be related to a representer of a particular observation via its observation operator (Echevin et al., 2000). Lets for instance call  $S$  the matrix of the first  $n$  Ceof:

$$S = [\text{Ceof } 1, \dots, \text{Ceof } 3]. \quad (4.1)$$

$H$  the observation operator associated to any observation in the region and  $\text{Pr}^f$  the forecast error covariance matrix in the reduced space spanned by the  $S$ . Then,  $HS^T\text{Pr}^fSH^T$  is a reduced-order representer matrix and gives information about the observability of the Ceofs and as a consequence about the observability of the corresponding error processes. However, the fact that a given error process is not observable in the reduced space does imply that it is not observable at all.

## 5. Error growth and predictability

The Ceofs calculated in the previous section offer for each Ensemble a natural basis for the forecast error subspace, a major advantage of this basis being its physical and dynamical consistency. It was indeed shown that each Ceof could be interpreted individually in terms of the processes it stands for. This Ceofs basis can now be used as a diagnostic tool to project both the Reference forecast and the associated Ensemble forecasts. The coefficient of projection is given by:

$$\alpha_j(\cdot) = \langle \cdot, S_j \rangle \quad (5.1)$$

where  $\langle \cdot, \cdot \rangle$  is the natural scalar product in the  $m$ -dimensional model forecast error subspace.

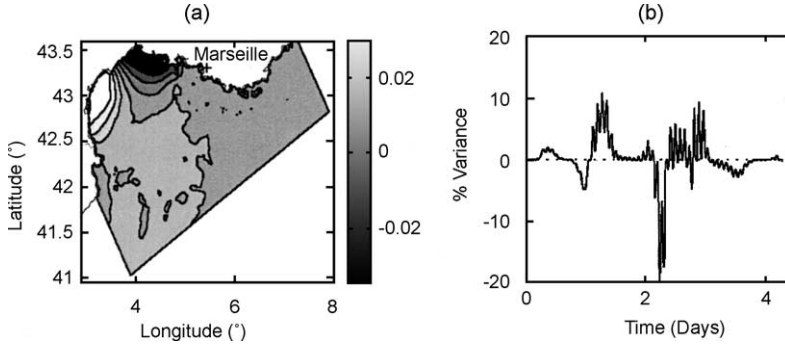


Fig. 13. Monivariate Central Empirical Orthogonal Functions 1 for 13 April, 0730Z and for Ensemble C (Normalised Variance Units). (a) Surface elevation anomaly. (b) Temporal evolution of the projection of the Reference forecasted surface elevation anomaly onto Eof (a).

### 5.1. Free mode of the Gulf of Lions

This projection coefficient offers a valuable tool of investigation to understand the dynamics of the Reference model forecast. Fig. 13 displays for instance the monivariate surface elevation anomaly Ceof 2 for Ensemble C. The surface elevation anomaly has a bipolar structure over the shelf which corresponds to the free mode of the Gulf. Indeed, the projection of the Reference surface elevation forecast ( $\alpha_2^C(\Delta\eta^f)$  with  $\Delta\eta^f = \eta^f - \bar{\eta}^f$  and  $\bar{\eta}^f$  the time average of the Reference forecast over the period) shows characteristic oscillations whose period is found between 2 and 2.5 h (Fig. 13b). Each wind burst occurring during the period is associated with an excitation of the free mode and a transfer of energy that is dissipated within a few hours. This confirms the observations made by several authors in the region based on tide gauges (Millot, 1982). This mode is obviously present among Ensemble C multivariate Ceofs (Ceof 4), but its structure is easier to understand in a monivariate mode.

### 5.2. Model regimes

Fig. 14 displays, still for Ensemble C, a three-dimensional plot of the time evolution of ( $\alpha_1^C(\Delta\eta^f)$ ,  $\Delta\alpha_1^D(\eta^f)$ ,  $\Delta\alpha_2^D(\eta^f)$ ). These coefficients correspond to the projection of the Reference simulation onto the monivariate Ceofs for the surface elevation anomaly plotted on the same figure along the corresponding co-ordinate axis. The first one (along the  $x$ -axis) is associated with Ensemble C, i.e. with a perturbation of the wind field while the remaining two ( $y$ - and  $z$ -axis) are associated with the perturbation of the river runoffs (Ensemble D). This plot shows aggregates of points in two regions of the forecast error subspace. The first one corresponds to the first 2 days of forecast and quiet wind conditions over the shelf, while the second one is associated with the end of the period and the strong wind event. When the model trajectory first penetrates into the second basin of attraction, it follows a first loop that corresponds to a first short wind burst after 2.5 days. Then with the strengthening of the northern wind, the trajectory comes back in the region of positive coefficients. The

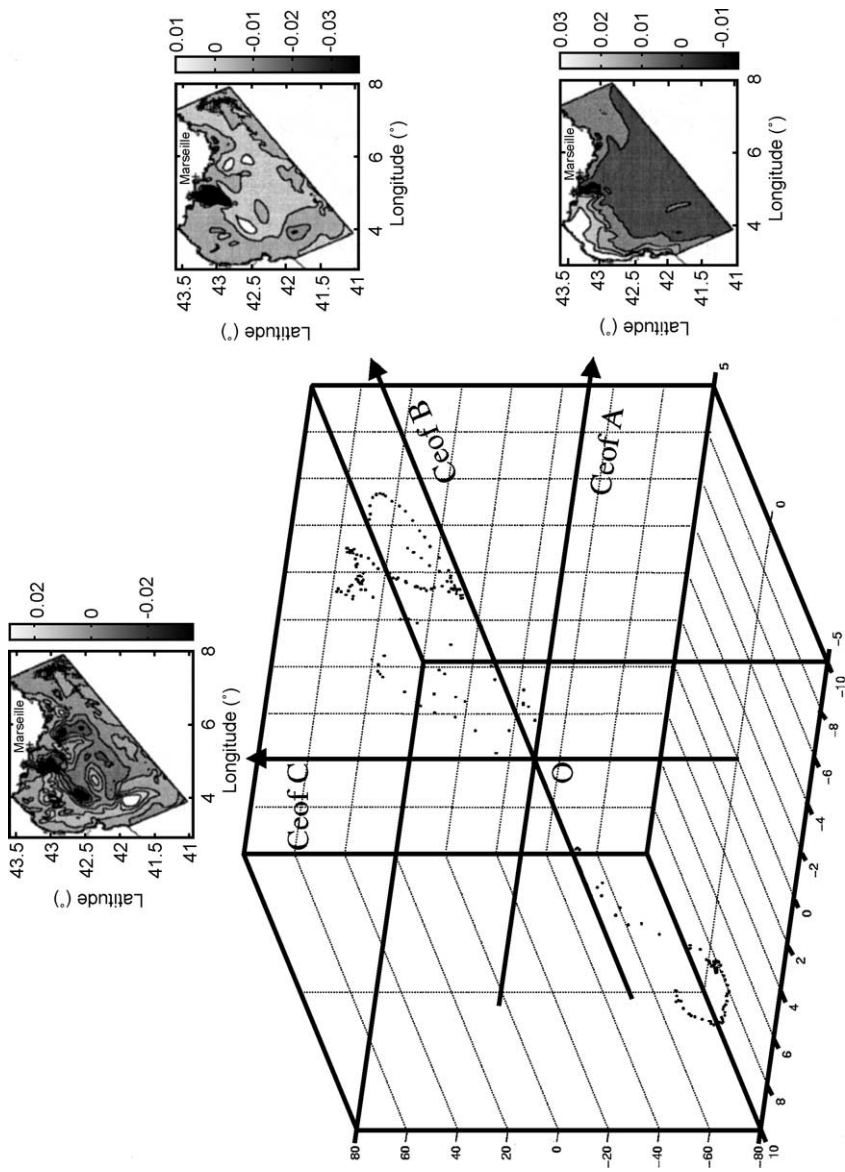


Fig. 14. Trajectory of the Reference forecast in the projection of model forecast error subspace onto the monovariate Central Empirical Orthogonal Functions for 13 April, 0730Z for Ensembles C and D (Normalised Variance Units) plotted along each axis. X-axis: surface elevation anomaly of Eof 1 for Ensemble C. Y-axis: surface elevation anomaly of Eof 1 for Ensemble D. Z-axis: surface elevation anomaly of Eof 2 for Ensemble D.

points being plotted regularly (every 3 h), the larger distance that separates two neighbours in the transition region indicates a larger rate of change along the trajectory in forecast error subspace resulting in a short transition period lasting less than a day. This plot shows that a particular subset of the Ceof basis is able to capture the dynamical transition that occurs when the wind strength increases. It also shows the strong correlation between the position of the Rhône plume, the thickness, position and strength of the LPC current and the northern wind. Fig. 14 is thus an indication of the existence of basins of attraction in model state space corresponding to specific regimes: in the present case, one corresponds to low wind conditions and the other to strong northern wind.

### 5.3. Evolution of the forecast error

The time evolution of the model forecast error  $\Sigma_j$  along the  $j$ th Ceof can be written:

$$\Sigma_j(t) = \frac{1}{n \times m} \left[ \sum_{i=1}^n \langle x_i^f(t) - x_n^c(t), \mathbf{S}_j \rangle^2 \right]^{1/2} \quad (5.2)$$

with  $m$  the total number of grid points. On a statistical point of view,  $\Sigma_j$  can also be understood as the variance of the projection coefficients onto the Ceof basis. If  $x_i^f$  are the Ensemble forecasts at final time, then  $\Sigma_j$  is the  $j$ th singular value of the corresponding covariance matrix. This diagnostics give an indication of the dispersion of the corresponding Ensemble and consequently of the evolution of the forecast errors with time. The curves corresponding to the Ceofs described in Section 4 are presented on Fig. 15.

In Ensemble A, the forecast error along the first two Ceofs increases quickly to reach a threshold value after less than a day (Fig. 15a). The errors do not evolve significantly until Day 3 and the arrival of the strong northern wind conditions over the shelf. After Day 3, the standard deviation increases sharply until the end of the studied period. This results shows first that the initial adjustment of the density perturbation is coherent with the inertial time scale (about 17 h in this region) and that the appearance of strong wind condition, by deepening the mixing layer, spreads the forecast errors that were located in the lower layers. A similar experiment has been carried out where the Ceofs are based on the initial field covariance (such Ceofs have by construction vanishing surface elevation anomaly and velocity) (not shown). The time evolution of the forecast errors for each Ceof decreases exponentially from the initial value with a e-folding time of less than a day. This confirms that the initial perturbations are not consistent with the model dynamics but are used as “seeds” before energy is transferred to dynamically adjusted processes.

The evolution of the forecast errors in Ensemble B is somehow different (Fig. 15b). The forecast error along Ceof 1 (associated to the entrance of the LPC current over the shelf) quickly reaches a threshold value at the beginning of the forecast that corresponds to the propagation of the perturbation of the characteristics of the LPC current in terms of exchanges through the shelf break. Then Fig. 15b shows low amplitude inertial oscillations until Day 3 and the Tramontane wind burst. The strengthening of the wind corresponds to an increase of the exchanges in the western part of the shelf and to an increase of the forecast errors associated to inertial oscillations. The forecast errors along Ceof 2 (associated with

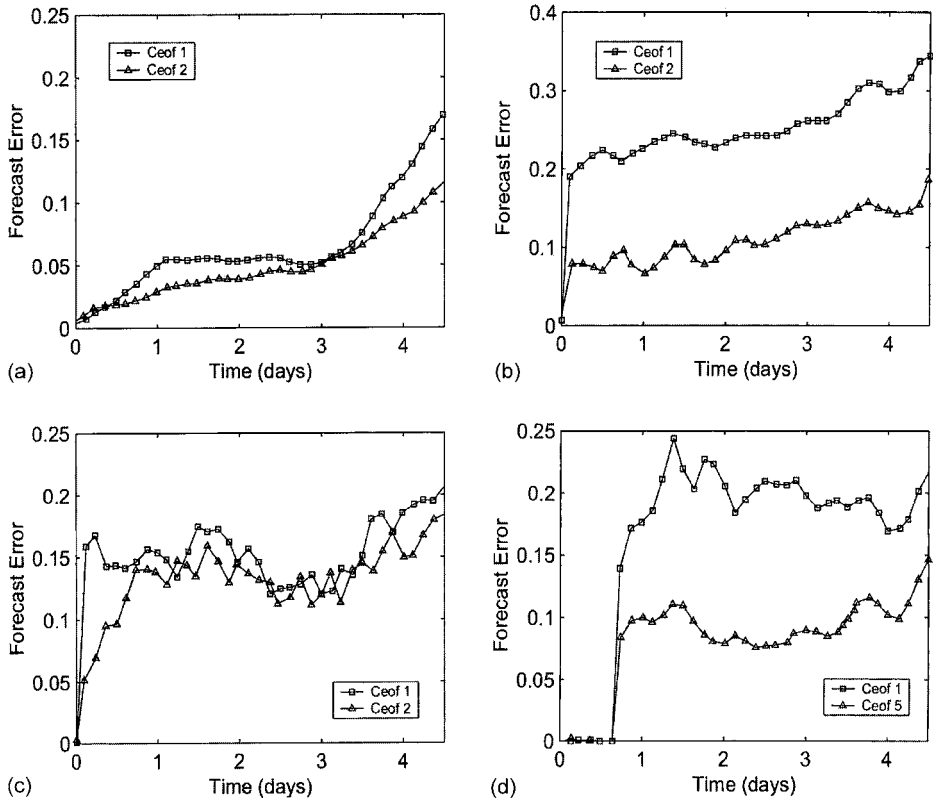


Fig. 15. Time evolution of the model forecast error. (a) Ensemble A. (b) Ensemble B. (c) Ensemble C. (d) Ensemble D.

the open boundaries) increase with time with several inertial oscillations. Forecast errors are to a certain extent continuously included in the domain through the perturbation of the open boundary external fields. However, the amplitude of the inertial oscillations is larger during the very first 2 days associated to the weak wind conditions. It can eventually be observed that Ceof 1 also explains part of the variance along the south-western boundary which also leads to a slight increase of the forecast errors during the first 2 days. This increase is not observed for Ensemble B' Ceofs (without any perturbation of the open boundary forcing) (not shown).

The Ensemble C error forecast sharply increases during the very first hours, then it converges toward a threshold value around which it oscillates (Fig. 15c). Ceof 3, associated to the position of the Rhône river plume along the coast (not shown), shows a sharp increase of the corresponding forecast errors after Day 3, i.e. in strong wind conditions. This means that for subsequent forecasts this Ceof would probably become the dominant empirical function and would explain a large percentage of the total error variance.

The evolution of the forecast errors in Ensemble D (Fig. 15d) is closely related to the dynamics of the plume of the Rhône river. Indeed, no significant errors are found before

the end of the first inertial period. Then, errors increase sharply to reach a threshold and do not evolve significantly before the arrival of the strong wind burst after Day 3.

## 6. Discussion and conclusion

In the present paper, four Ensembles were computed in order to understand the impact on the ocean forecast in the coastal area of the uncertainties on respectively the initial density field, the initial large scale circulation together with its forcing along the open boundaries, the wind field and the river discharges. As a test case, we have studied the penetration of the LPC current over the shelf of the Gulf of the Lion in the north-western Mediterranean.

For each Ensemble, the Central forecast and the second-, third- and fourth-order moments were computed and compared and after 4.5 days (at forecasting time) leading to interesting conclusions concerning the pdf of the coastal ocean. Large amplitude non-linear processes leading to peaked distribution have been found along the shelf break at the top of the canyons in regions where the LPC current is interacting with the bottom topography. The study of the third- and fourth-order moments of the distribution further confirms the spreading of the pdf due to the wind induced mixing (Ensemble C) and the smoothing of the physical processes by generation and dissipation of baroclinic gravity waves when random noise is added (Ensemble A). The presence of several basins of attraction in model state space associated to the various regimes clearly show the multimodal character of this distribution already described by Power and Kleeman (1993). In the present study, all the Ensemble forecasts eventually ended in the same basin of attraction due to the strong northern wind forcing. However, this is not necessarily the case and when the final Ensemble forecast distribution is itself multimodal, the procedure of assimilation becomes non-linear by nature and rather complex. This leads to serious limitations concerning the use of linear filters to assimilate observations in coastal ocean models and in particular concerning the traditional Gaussian assumption for the error forecast pdf.

To go further, the spatial structure of the variance of an “equivalent” Gaussian distribution was studied using Empirical Orthogonal Functions of the Ensemble variance/covariance matrix  $((x_{i,j}^f - x^c)_{i \in [1,m], j \in [1,n]})$  (the so-called Ceofs). The “equivalent” Gaussian approximation was trivially obtained by truncating the pdf to its second-order moment. The computation of the Ceofs shows very promising possibilities both to study the forecast error subspace and the evolution of identified physical processes. We must, however, keep in mind that, based on our results on the third- and fourth-order moments, the Ceofs only provide information about the (non-unique) normal part of the pdf. Locally and in particular in regions like canyons non-linear processes are very likely to be large.

For such an equivalent distribution, the Ceofs provide a physical and coherent description of the forecast error subspace in the form of an orthogonal basis of that subspace. The interaction of the slope current (the LPC current) with the canyons over the shelf, the spreading of the river fresh water discharge, the position of the density winter front over the shelf break or the free oscillation of the Gulf of Lions were shown to have a clear signature on the Ensemble variance and as a consequence on the Ceofs. These latter mathematical tools yield a three-dimensional multivariate description of the processes clearly showing for instance the correlation of the circulation over the shelf with the large scale general



circulation. The importance of the canyons as means of communication between the shelf and the deep ocean has in particular been evidenced and the upsloping/downsloping scheme (Klinck, 1996) has been confirmed. This has important consequences in terms of coastal and shelf modelling. Indeed a good modelling of the topographic structure of the shelf canyons is indeed required to accurately represent the shelf processes. Even in the case of a very strong wind northern wind event over the Gulf of Lions, the shelf circulation was shown to be partly controlled by the entrance of the LPC current through the western canyons (Estournel et al., 2001a,b). The thickness, the position and the strength of that current were also found to be crucial. Indeed a coherent, dynamically adjusted perturbation of these basic properties leads to a modification of the water exchanges through the shelf break by modifying the upsloping/downsloping conditions.

This Ceof basis has also been used to investigate the time evolution of various processes in the Reference forecast (free mode oscillation, influence of the wind on the characteristics of the plume of the Rhône and of the LPC current, etc.) and to understand the evolution of the model forecast errors. Ceofs based on Ensemble methods appear as valuable tools in the understanding of the predictability of the coastal ocean circulation. It was shown for instance in Section 5 that following a first increase associated to the adjustment, the evolution of the forecast error was very dependent on the dominant physical and numerical processes. The forecast error time scales associated to the open boundary, to the initial state or to other forcing are several candidate to become the limiting factors. An interesting observation is then the fast increase of the errors following the strong wind burst following Day 3. Based on the Ceofs, it was first possible to evidence the appearance of two regimes in model forecast error subspace: one corresponding to low wind conditions over the shelf, the other to the strong northern wind burst. The evolution of the forecast errors is very dependent on the regime: the errors associated to the perturbation of the density field (Ensemble A) are for instance increasing much faster in the second regime. A crucial consequence is that the predictability time scale is closely associated to the coastal ocean regime and it is in particular shorter under strong forcing conditions.

The Ensemble method can be used to provide an estimate of the representer of a given observation operator. It is thus possible to understand the observability of a particular physical error process (such as a Ceof) and the impact of an observation made over the shelf or over the deep ocean. The perspective of a local typology of transitions in model state space offers an opportunity to set-up an observing strategy for a future assimilation tool. Some information about the predictability and controllability of the physical and numerical components of the system were also given by the evolution of the forecast errors along each of these Ceofs. Such an Ensemble study could be pursued over longer periods associated for instance to different seasons in order to yield some kind of typology of coastal ocean error regimes in the region. Several processes were indeed found to be associated with sharp error increases in particular during the northern wind burst. The consequences of a bad specification of the open boundary conditions have also been translated in terms of forecast errors. We, thus, better understand how the model is to be controlled which is the very first step toward the implementation of a realistic forecasting tool in the coastal region associating observation and modelling through a well adapted assimilation scheme. The Ceof basis is an interesting candidate to define the Reduced-order Control Subspace in which observations can be assimilated in order to correct the model trajectory and in particular to adjust such

transitions in model state space. An important topic is the observability of subsets of the Ceof basis, another is its time stationarity. We did not provide any information to understand this stationarity in the present paper. We indeed focus on a particular period without exploring the pertinence of the computed basis for different seasons nor even for different periods. Such an approach was already followed by Ballabrera et al. (2001) based on the SEEK filter.

In the present work, the perturbation of the initial state proposed by Evensen (1994) and Echevin et al. (2000) has been extended to the various forcings: open boundary conditions, river freshwater discharge and atmospheric forcings. A promising development concerns the construction of perturbations which are physically coherent and dynamically adjusted. We have shown that such perturbations limit the period of adjustment when they are applied on the initial state and are particularly interesting in terms of the energy which is eventually transferred to the system. Several Ensembles have been generated: in Ensemble A, the initial density field is perturbed by adding correlated noise and the resulting perturbations along the Ceofs computed for the initial perturbed state decrease exponentially with time while energy is non-linearly transferred to other modes (evidenced by the forecast Ceofs). The situation is different in Ensemble B where the position, thickness and strength of the LPC current are varied based on dynamically adjusted perturbations, and the amplitude of the initial perturbations increases with time. A dynamically adjusted perturbation is thus a mean to vary specifically the characteristics of a few selected processes and to carry out only a few well-identified Ensemble forecasts.

## Acknowledgements

We gratefully thanks *Météo-France* for the atmospheric forcings based on the Aladin model during the period and the *Compagnie Nationale du Rhône* and the *Directions Départementales de l'Équipement* for the Rhône, Petit Rhône and Aude river daily runoffs. No Ensemble computation would also have been possible without the support of the *Laboratoire d'Aérodynamique* (Toulouse, France) Linux cluster people: Serge Prieur, Didier Gazen and Juan Escobar nor without the support of Vincent Toumazou (*LEGOS*, Toulouse, France) as far as the computation of the Empirical Orthogonal Functions is concerned.

## Appendix A. Perturbation of the LPC current

The position of the LPC current across the shelf break depends essentially on the bottom topography and on the hydrodynamic characteristics of the current: its depth and thermal signature. As a consequence, the OGCM can only give an approximate picture of such a current insofar as its interaction with the bottom topography is based on a coarse modelling of the shelf break. By adding an adjusted perturbation of the position, thickness and strength of the LPC current it is thus possible to open some new degrees of freedom in Model state space corresponding to error processes.

To generate these perturbations, the model proposed by Estournel et al. (2001a,b) is used. It is assumed that the geostrophic current vanishes uniformly at some depth  $Z_0$

which leads to a relation between the surface pressure gradient and the weight of the water column:

$$\nabla p_s = -g \int_{-Z_0}^0 \nabla \rho \, dz. \quad (\text{A.1})$$

Following Cooper and Haines (1996), it is assumed that the density increments  $\nabla \rho$  are obtained from a horizontally homogeneous density field  $\rho_{\text{ref}}(z)$  by a vertical displacement of each individual water column. The density can thus be written:

$$\rho(x, y, z) = \rho_{\text{ref}}(z) + \frac{\partial \rho_{\text{ref}}(z)}{\partial z} \Delta h(x, y) \quad (\text{A.2})$$

where  $\rho_{\text{ref}}$  is a reference profile of density and  $\Delta h$  the vertical displacement at location  $(x, y)$ . A consequence of this assumption is the homogeneity of the perturbation potential vorticity over the domain (Cooper and Haines, 1996). This enables the definition of a similarity function  $R$  given by:

$$R(z) = \frac{\rho_{\text{ref}}(z) - \rho_{\text{ref}}(z_0)}{\rho_{\text{ref}}(0) - \rho_{\text{ref}}(z_0)} \quad (\text{A.3})$$

and satisfying:

$$\nabla p(z) = R(z) \nabla p_s. \quad (\text{A.4})$$

By averaging the density profiles over the areas bounded by the 0, 100 and 400 m isobaths (associated to three values of  $Z_0$ : 100, 400 and 2500 m), three reference profiles ( $\rho_{\text{ref}}$ ) were computed and used to generate three adjusted stationary circulations satisfying:

$$-\frac{\partial(f/\bar{R})}{\partial y} \frac{\partial \psi}{\partial x} + \frac{\partial(f/\bar{R})}{\partial x} \frac{\partial \psi}{\partial y} = \frac{\partial}{\partial y} \left( \frac{-\tau_{bx}}{\rho_0 \bar{R}} + \frac{F_x}{\bar{R}} \right) - \frac{\partial}{\partial x} \left( \frac{-\tau_{by}}{\rho_0 \bar{R}} + \frac{F_y}{\bar{R}} \right) \quad (\text{A.5})$$

where  $f$  is the Coriolis parameter,  $\psi$  the barotropic current function,  $(F_x, F_y)$  is the diffusion Laplacian operator and  $\bar{R}$  the integral of the similarity function  $R$ :

$$\bar{R} = \int_{-h}^0 \frac{\rho_{\text{ref}}(z) - \rho_{\text{ref}}(z_0)}{\rho_{\text{ref}}(0) - \rho_{\text{ref}}(z_0)} \, dz \quad (\text{A.6})$$

$\tau_b$  is the linear bottom friction given by:

$$\tau_b = \rho_0 C_d U \begin{pmatrix} u_b \\ v_b \end{pmatrix} = \rho_0 C_d U \frac{R_{(z=-h)}}{\bar{R}} \begin{pmatrix} -\frac{\partial \psi}{\partial y} \\ \frac{\partial \psi}{\partial x} \end{pmatrix} \quad (\text{A.7})$$

with  $C_d = 2.6 \times 10^{-3}$  and  $U$  is a reference velocity equal to 0.5 m/s.

Eq. (A.5) being a linear equation in  $\psi$ , it is solved numerically on the 3-km grid over a domain slightly larger than the modelled one and using the OGCM output as boundary conditions. The solutions are thus added to the initial field as small adjusted perturbations of the velocity, temperature, salinity and surface elevation anomaly fields with random amplitude.

Fig. 16 shows the depth average velocity for each circulation. As a consequence of the conservation of potential vorticity, the shallow reference profile leads to a LPC current that penetrates over the shelf (Fig. 16a) while the slope current based on the deeper reference profile follows the 2000 m isobath (Fig. 16b). Each individual perturbation can thus be

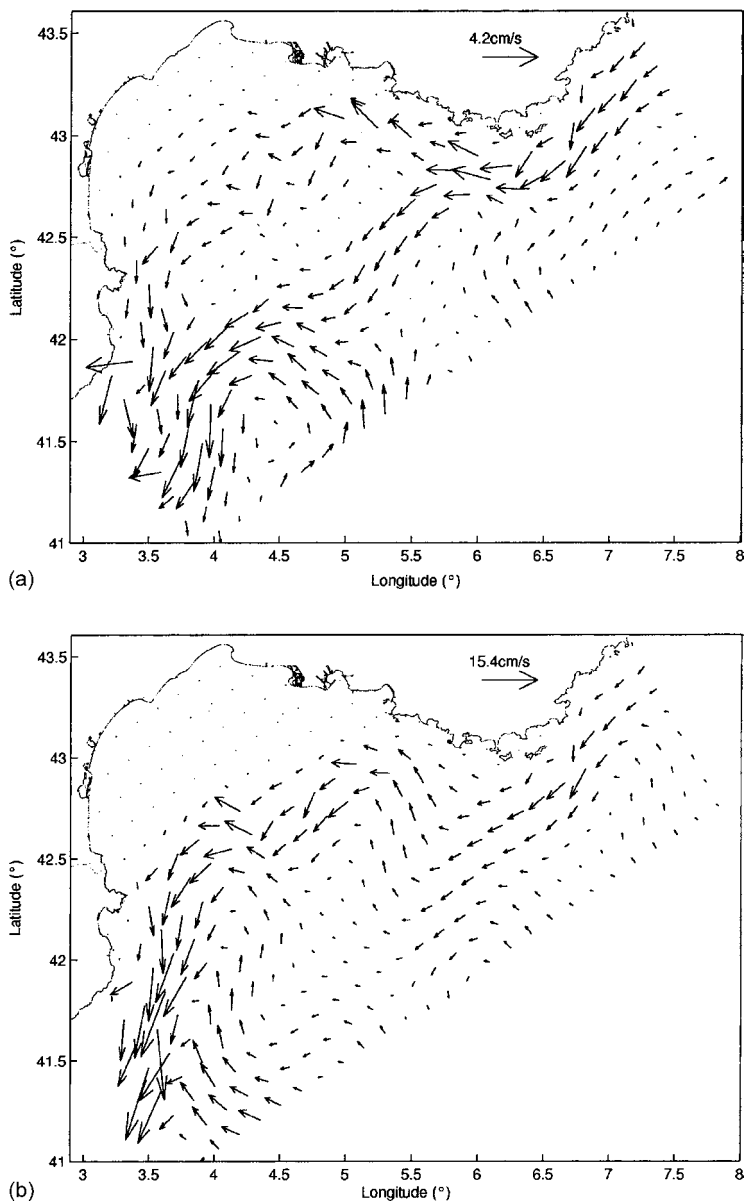


Fig. 16. Mean velocity of the dynamically adjusted circulations obtained with profiles 1–3 of Fig. 4a: (a) Profile 1 for  $Z_0 = 100$  m; (b) Profile 2 for  $Z_0 = 400$  m; (c) Profile 3 for  $Z_0 = 2500$  m.

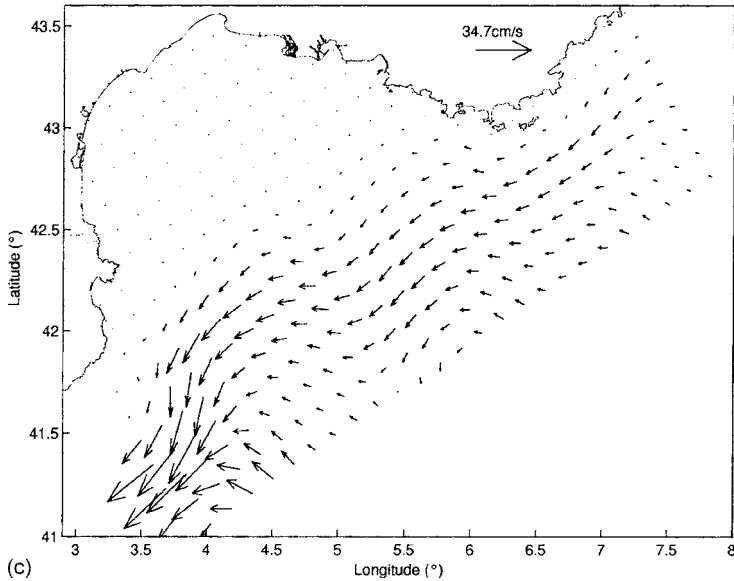


Fig. 16. (Continued).

viewed as a perturbation of the upstream potential vorticity of the LPC current leading to a perturbation of the penetration of the current over the shelf.

## References

- Auclair, F., Casitas, S., Marsaleix, P., 2000a. Application of an inverse method to coastal modelling. *J. Atmos. Oceanic Technol.* 17, 1368–1391.
- Auclair, F., Marsaleix, P., Estournel, C., 2000b. Truncation errors in coastal modelling: evaluation and reduction by an inverse method. *J. Atmos. Oceanic Technol.* 17, 1348–1367.
- Auclair, F., Marsaleix, P., Estournel, C., 2001. Specification of the exchanges through the continental shelf break in high resolution coastal models: penetration of the LPC current over the Gulf of Lion (Mediterranean). *Oceanol. Acta* 24 (6), 529–544.
- Ballabrera, J., Brasseur, P., Verron, J., 2001. Dynamical evolution of the error statistics with the SEEK filter to assimilate altimetric data in eddy-resolving ocean models. *Q. J. R. Met. Soc.* 127, 233–253.
- Beckers, J.M., 1995. La Méditerranée occidentale: de la modélisation mathématique à la simulation numérique. Collection des publications de la Faculté des Sciences Appliquées no 136. Université de Liège, B-4000 Liege, Belgique, p. 342.
- Canizares, R., 1999. On the application of data assimilation in regional coastal models. Ph.D. thesis, Delft University of Technology, Delft.
- Cooper, M., Haines, K., 1996. Altimetric assimilation with water property conservation. *J. Geophys. Res.* 101 (C1), 1059–1077.
- D’Agostino, R.B., 1986. Tests for the normal distribution. In: D’Agostino, R.B., Stephens, M.A. (Eds.), *Goodness-of-Fit Techniques*. Marcel Dekker, New York, pp. 367–419.
- De Mey, P., Benkiran, M., 2002. A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. In: Pinardi, N., Woods, J.D. (Eds.), *Ocean Forecasting, Conceptual Basis and Applications*, Springer, Berlin, ISBN: 3-540-67964-2.

- 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 (10), 2627–2635.
- Estournel, C., Kondrachoff, V., Marsaleix, P., Vehil, R., 1997. The plume of the Rhone: numerical simulation and remote sensing. *Continental Shelf Res.* 17 (8), 899–924.
- Estournel, C., Broche, P., Marsaleix, P., Devenon, J.L., Auclair, F., Vehil, R., 2001a. The Rhone river plume in unsteady conditions: numerical and experimental results. *Estuarine Coast. Shelf Sci.* 53, 25–38.
- Estournel, C., Durrieu de Madron, X., Marsaleix, P., Auclair, F., Julliand, C., Vehil, R., 2001b. Observation and modelling of the winter coastal ocean circulation under wind conditions influenced by the continental orography (Fetch experiment). *J. Geophys. Res.*, in press.
- Evensen, G., 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.* 97 (C11), 17905–17924.
- 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.
- Ghil, M., Cohn, S., Travantzis, J., Bube, K., Isaacson, E., 1981. Application of Estimation Theory to Numerical Weather Prediction. *Dynamic Meteorology: Data Assimilation Methods*. Springer, Berlin.
- Klinck, J.M., 1996. Circulation near submarine canyons: a modelling study. *J. Geophys. Res.* 101 (C1), 1211–1223.
- Lacarra, J.F., Talagrand, O., 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus* 40A (2), 85–95.
- Marsaleix, P., Estournel, C., Kondrachoff, V., Vehil, R., 1998. A numerical study of the formation of the Rhône River plume. *J. Mar. Res.* 14, 99–115.
- Miller, R.N., Ghil, M., Gauthiez, F., 1994. Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.* 51, 1037–1056.
- Millot, C., 1982. La dynamique marine du plateau continental du Golfe du Lion en été. Ph.D. thesis. University Pierre et Marie Curie, Paris 6.
- Millot, C., 1990. The Gulf of Lions' hydrodynamics. *Cont. Shelf Res.* 10 (9–11), 885–894.
- Power, S.B., Kleeman, R., 1993. Multiple equilibria in a global ocean general circulation model. *J. Phys. Ocean.* 23, 1670–1681.
- Tarantola, A., 1988. *Inverse Problem Theory*. Elsevier, Amsterdam, pp. 613.
- Toumazou, V., Créteaux, J.F., 2001. Using a Lanczos eigensolver in the computation of Empirical Orthogonal Functions. *Monthly Weather Rev.* 129 (5), 1243–1250.