A Semi-Evolutive Filter with Partially Local Correction Basis for Data Assimilation in Oceanography

Ibrahim Hoteit, Dinh-Tuan Pham, Jacques Blum

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Semi-Evolutive Filter with Partially Local Correction Basis for Data Assimilation in Oceanography

Ibrahim Hoteit, Dinh-Tuan Pham and Jacques Blum

No 3975

Juillet 2000

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--3975--FR+ENG

INRIA
RHÔNE-ALPES

A Semi-Evolutive Filter with Partially Local Correction Basis for Data Assimilation in Oceanography

Ibrahim Hoteit, Dinh-Tuan Pham and Jacques Blum

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IDOPT

Rapport de recherche n ° 3975 — Juillet 2000 — 36 pages

Abstract: The singular evolutive extended Kalman (SEEK) filter has been proposed recently by *Pham et al.* [21] for data assimilation into numerical oceanic models. This filter has been applied in different realistic ocean frameworks and have provided satisfactory results [20, 21, 24]. However, the SEEK filter remains expensive in real operational assimilation. To reduce cost and obtain a better representativity, we introduce the idea "local correction basis". In this local analysis, the basis vectors support a small region of the model domain and vanish elsewhere. Such basis however can not be made to evolve according to the model without destroying its locality property. Therefore we shall keep this basis fixed and we argument it by a few global basis vectors which evolve. The resulting semi-evolutive partially local filter is much less costly to implement than the SEEK filter and yet can yield better results. In a first application, validation twin experiments are conducted in a realistic setting of the OPA model over the tropical pacific ocean.

Key-words: Data assimilation. Reduced Kalman filtering. SEEK Filter. EOFs analysis. Forgetting factor.

(Résumé : tsvp)

This work was carried out within the framework of the IDOPT project which is a joint project between INRIA, CNRS, University Joseph Fourier and INPG. The authors would like to thank Dr. Eric Blayo for interesting discussions during part of this work. Thanks are also given to Laurent Debreu and Laurent Parent for assistance with the numerical model.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Un Filtre Semi-Evolutif avec une Base de Correction Partiellement Locale pour l'Assimilation de Données en Océanographie

Résumé : Le filtre de Kalman étendu singulier évolutif (SEEK) a été proposé par *Pham et al.* [21] pour l'assimilation de données dans les modèles océaniques. Ce filtre a été implémenté et testé avec succès dans plusieurs situations réalistes. Cependant ce filtre reste cher pour une océanographie opérationnelle. Pour réduire son coût et avoir plus de représentativité, on introduit l'idée "base de correction locale". Dans cette analyse, les vecteurs de la base ont pour support une petite région du domaine du modèle. Une telle base cependant ne peut pas évoluer avec le modèle sans perdre son caractère local. On propose alors de garder cette base fixe et on lui rajoute quelques vecteurs de base globaux qui évoluent. Le filtre semi-évolutif à base de correction partiellement locale qui en résulte est bien moins coûteux que le filtre SEEK mais donne des meilleurs résultats. En première application, des experiences jumelles ont été conduites dans une configuration réaliste du modèle OPA dans l'océan pacifique tropical.

Mots-clé : Assimilation de données. Filtrage de Kalman réduit. Filtre SEEK. Analyse EOFs. Facteur d'oubli.

1 Introduction

Developing operational data assimilation schemes in oceanography is one of the major challenge for the coming year. Such schemes should be operated in conjunction with realistic ocean models, with reasonable quality and at an acceptable cost as in meteorology. Increasing interest in these assimilation schemes is emerging for the purposes of improving short and mid term weather prediction, developing climate predictions or for the specific objectives of the navies in various countries. Since the occurrence satellites launching, considerable progress has been made in applying concepts and techniques from statistical estimation and optimal control theories to the problem of oceanographical data assimilation (see for example *Ghil and Manalotte-Rizzoli* [11] for a review).

The Kalman filter is a statistical data assimilation scheme which provides the best estimate, in the sens of least-squares, of the state of a linear model giving the previous observations [16]. However, its application into oceanic models encounters two major difficulties: non lineartiy and computational cost. The first can be partially resolved by linearising the model around the state estimate, which leads to the so-called extended Kalman (EK) filter [15]. The second is due to the huge dimension of the model state. Several approximations, that essentially consist in projecting the system state onto a low dimensional sub-space, have been proposed to reduce the dimension of the system [5, 6, 7, 9, 12].

The singular evolutive extended Kalman filter (SEEK) introduced by *Pham et al.* [21] is an approach to reduce the cost of the EK filter. It essentially consists in approximating the error covariance matrix by a singular matrix which leads to not making corrections in the directions where the errors are naturally strongly attenuated by the system. These "directions of correction" evolve in time according to the model dynamics and as usually initialized through the empirical orthogonal function (EOFs) analysis. This filter has been applied in different realistic ocean frameworks and quite satisfactory results have been obtained [20, 21, 24].

Nevertheless, the SEEK filter remains expensive in real operational assimilation since the evolution equation of its correction basis requires model integration for each correction basis vectors. To reduce the cost of this filter, *Hoteit et al.* [13] have proposed different degraded forms of the SEEK filter in which they simplify the evolution of its correction basis. But this does not address a built-in weakness of the SEEK filter (in fact it may even exacerbate it) in that the correction basis

may not contain enough vectors to capture the (usually short) range variability of the model dynamics. Increasing the number of basis vector would of course increase cost and, as our experiences show, can only increase marginally the representativity of the basis. Indeed, the percentage of explained inertia in an EOFs analysis (see section 2) quickly level out as the number of basis vector grows. Further, the last EOFs usually cannot be accurately evaluated, due to statistical fluctuation and the use of a not long enough historical run. More importantly in real applications, the model can be subjected to errors and hence an EOFs analysis based on a run of the model would be subjected to errors too. We believe that such error affect mostly the last EOFs because they are the most numerically unstable.

For the above reasons, we shall introduce a kind of "local EOFs analysis" which would hopefully provide better representativity. The local EOFs have support a small region of the ocean and vanish elsewhere and therefore reduce the computation cost thus allow more basis vector for a given cost. This local analysis limits the spatial correlation length of the ocean variables which is consistent with the idea that such correlation should vanish for far away spatial locations. Further, it allows choosing the number of the EOFs differently in each ocean sub-domain in order to maximize representativity. There is however a difficulty in the above "local basis" approach. Such basis cannot be made to evolve with the model without destroying its locality property. Therefore we shall keep this basis fixed and we will augment it by a few global basis vectors which evolve. The resulting semi-evolutive partially local filter Kalman (SEPLEK) filter, is much less costly to implement than the SEEK and yet can yield better results.

The paper is organized as follows. Section 2 discusses different approaches to obtain better representativity from the EOFs analysis. The SEPLEK filter is described in section 3. In section 4, we present an adaptive scheme to tune the value of the forgetting factor to overcome model instabilities. Finally, the performance of the new filter is illustrated in section 5 with some simulation results performed under a realistic setting of the OPA model in the tropical pacific ocean.

2 Empirical orthogonal functions (EOFs) analysis

The "classical" EOFs analysis is an efficient method to find an optimal representation of the "global" ocean variability. It can be viewed as a method of "compressing data" contained in ocean states by summarizing the correlation of their variables in a few

vectors only, called EOFs. However, a too small number of EOFs will be able to capture generally only the long range spatial correlation of the ocean variables. To capture the short range correlation without requiring an excessive number of EOFs, we shall develop a so called local EOFs analysis.

2.1 Classical (Global) EOFs analysis

This analysis provides a representation as accurately as possible of an ensemble of state vectors $X_1, \ldots, X_N$ in $\mathbb{R}^p$ in a low-dimension (noted r) subspace. Indeed, for a vector X, let $\tilde{X}$ denotes its orthogonal projection onto a subspace spanned by an $\mathcal{M}-$orthogonal basis $L = \{\phi_k\}_{k=1,\ldots,r}$ where $\mathcal{M}$ is some metric (to be chosen) in the state space, and the constant function:

$$\tilde{X} = \bar{X} + \sum_{k=1}^{r} c_k \phi_k = \bar{X} + LL^T \mathcal{M}(X - \bar{X}) \tag{1}$$

where $\bar{X}$ is the barycenter of $X_1, \ldots, X_N$ i.e.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \tag{2}$$

and

$$c_k = < X - \bar{X}, \phi_k >_{\mathcal{M}} = \phi_k^T \mathcal{M}(X - \bar{X}). \tag{3}$$

Then the EOFs analysis aims at minimizing the mean squares projection error

$$e^2 = \frac{1}{N} \sum_{i=1}^{N} \|X_i - \tilde{X}_i\|_{\mathcal{M}}^2 \tag{4}$$

with respect to all choices of the basis. Here the introduction of a metric $\mathcal{M}$ is needed in the case where the state variables are not homogeneous (as they represent different physical variable such as velocity, salinity, temperature ...) to obtain a distance between state vectors independent from unit of measure.

The solution L of the above minimization problem is given by the first r normalized eigenvectors of the matrix P relative to $\mathcal{M}$, where P is the sample covariance matrix of $X_1, \ldots, X_N$ ranked in decreasing order according to their eigenvalues $\lambda_1, \ldots, \lambda_r$. Moreover, one can verify that the EOFs analysis also provides the

best low rank r approximation of P (in the sens of least squares) by

$$P \approx L\Lambda L^T \tag{5}$$

where Λ is a diagonal matrix containing $\lambda_1, \ldots, \lambda_N$ on its diagonal. With regard to the choice of r, it has been shown that the fraction of variance (or inertia) explained by the first r EOFs is given by

$$\mathcal{I}(\phi_1, \ldots, \phi_r) = \frac{\sum_{k=1}^{r} \lambda_k}{\sum_{k=1}^{p} \lambda_k}. \tag{6}$$

which can be used as a guide for choosing r ($\mathcal{I}$ should be close to 1). The reader is referred to [19, 22] for more details.

In our case, we are interested in representing the variability of the state model around its mean and thus we use a long historical sequence of model states $X_1, \ldots, X_N$ which can be extracted from a model run. The matrix P is given by

$$P = \frac{1}{N}\mathcal{X}\mathcal{X}^T \qquad \text{with} \qquad \mathcal{X} = [X_1 - \bar{X} \cdots X_N - \bar{X}]. \tag{7}$$

Such matrix P contains a bulk of information on the system variability when N is sufficiently large.

2.2 Local EOFs analysis

The classical EOFs analysis summarizes the correlations between all ocean variables. The correlations which results often have long range if a few EOFs are retained. However, one can expect that the ocean variables have a limited spatial correlation length since two variables evaluated at ocean locations too far away would not, in principle, correlated. Further, the EOFs analysis does not distinguish calm zones from turbulent zones because it treats the ocean as a single homogeneous domain. The resulting basis L contains therefore common information on these zones and this can reduce its capacity to well-capture rather different local variabilities.

To remedy to these deficiencies, we propose to force the EOFs to be local without losing too much the optimal characteristic of the EOFs analysis. Our basic idea consists in constructing a set of EOFs having support a small region of the ocean. This can simply be done by independently applying the EOFs analysis on different ocean sub-domains. Such local analysis will limit the spatial correlation length in

the relevant domain. Since the combined representation would provide no correlation between spatial points in different sub-domains, to obtain better representativity, we will use overlapping sub-domains. Such an approach is very flexible as we have at our disposal a wide range of options to choose: (i) the number of sub-domains, (ii) the extent of their overlaps and possibly the size and shape of these sub-domains as well, and (iii) the number of the local basis elements in each sub-domains. Such flexibility would allow us to construct a basis best adapted to the problem at hand. Another point is that calculations on local basis functions are much less costly, therefore one can afford more local basis elements (on the whole) without increasing (even possible decreasing) the computation cost.

• *Construction of the local EOFs basis:*

To construct the local EOFs basis, one should first define the ocean sub-domains on each of a separate EOFs analysis will be applied. As said before, these sub-domains should overlap. To do this properly, one consider a *partition of unity* of the ocean domain, i.e. an ensemble of positive functions $\{\chi^{(j)}, j = 1, \ldots, J\}$ defined in the ocean domain whose sum is identically equal to 1. Therefore any ocean state vector X can be written as

$$X(x,y,z) = \sum_{j=1}^{J} X(x,y,z)\chi^{(j)}(x,y,z) = \sum_{j=1}^{J} X^{(j)}(x,y,z) \tag{8}$$

where

$$X^{(j)}(x,y,z) = X(x,y,z)\chi^{(j)}(x,y,z) \tag{9}$$

and x, y and z design the spatial coordinates. Next, for each j between 1 and J, one carries out separately an EOFs analysis on each local field $X^{(j)}$ to obtain a basis for each ocean sub-domain.

Now, using equation (1), one obtains a representation formula for each X_k according to

$$X^{(j)}(x,y,z) = \bar{X}^{(j)}(x,y,z) + \sum_{l=1}^{r^{(j)}} c_l^{(j)} \phi_l^{(j)}(x,y,z) + e^{(j)} \tag{10}$$

where the $\phi_l^{(j)}(x,y,z)$ design the EOFs resulting from the j^{th} local EOFs analysis, $\bar{X}^{(j)}$ is the mean vector, $r^{(j)}$ is the number of retained EOFs and the coefficients

$c_l^{(j)}$ are constants. Thus by taking the sum of the last equation on the index j, one obtains a representation formula of the global vector X in the form

$$X(x,y,z) = \bar{X}(x,y,z) + \sum_{j=1}^{J} \sum_{l=1}^{r^{(j)}} c_l^{(j)} \phi_l^{(j)}(x,y,z) + e. \tag{11}$$

In this formula, the representation error e is equal to the sum of all local representation errors $e^{(j)}$, which clearly vanishes outside the support of the function $\chi^{(j)}(x,y,z)$. However, one can still reduce this error by fitting the coefficients $c_l^{(j)}$ in order to obtain the minimum of the squared errors sum. Indeed, if one writes equation (11) in the matrix form

$$X = \bar{X} + \mathcal{C}^T B + e \tag{12}$$

where B is the matrix whose $r = \sum_{j=1}^{J} r^{(j)}$ columns are the local EOFs $\phi_l^{(j)}$ and $\mathcal{C}$ is the column vector containing the coefficients $c_l^{(j)}$, then it can be easily seen that one can still reduce the representation error e in (11) by considering the vector $\mathcal{C}^*$ which minimizes the norm of $(X - \bar{X}) - \mathcal{C}^T B$, which is no other than the orthogonal projection of X on the subspace spanned by the columns of B. Therefore, one takes as in section 2.1, the set of row vector of B as a representation basis, called local EOFs basis, of the variability of the system state.

The number $r^{(j)}$ of the EOFs in each sub-domain can be chosen as in the classical EOFs analysis. Thus the value of $r^{(j)}$ varies according to the variability of each sub-domain. But this may served as a first guess, one can still readjust it for example by increasing the value in the sub-domains of strong variability.

- *Choice of the ocean sub-domains:*

The size of every sub-domain is obviously characterized by the support of the partition of unity functions $\chi^{(j)}$. In practice such functions should support a small region of the ocean and vanish elsewhere. For example, in the case of a rectangular oceanic domain, one can take these functions in tensorial form i.e.

$$\chi^{(j)}(x,y,z) = \chi_X^{(j_1)}(x) \chi_Y^{(j_2)}(y) \tag{13}$$

such as j_1 varies from 1 to J_1, j_2 from 1 to J_2, $j = j_1 + J(j_2 - 1)$ varies from 1 to $J = J_1 J_2$ and the functions $\chi_X^{(j_1)}$ and $\chi_Y^{(j_2)}$ have also a small support with sum identical to 1. Moreover, since the ocean vector field is manifestly continuous, it is a good idea to take these functions continuous too, which would require them to have overlapping support. Note that one should not limit the correlation length in the vertical direction to let the surface informations propagate to the ocean bottom.

The main difficulty with the local EOFs is that we have not find a way to make it to evolve with the model (see section 3) without loosing its locality property. Even if we decide to use it to initialize the SEEK filter and abandon the locality property by allowing the basis function to evolve to track the model dynamics, this can be very costly because of the large dimension of the basis. We have also noticed in our experiments that the long range variability was not well represented by the local basis. For these reasons, we will introduce some global basis elements, which results in a so-called mixed global-local basis. An advantage of this approach is that the global basis can be made to evolve to track the system dynamics as in the SEEK filter.

2.3 Global-Local (Mixed) EOFs analysis

The oceanic phenomena can have both long range and short range variability. The last type of variability is not generally well represented in the first few global EOFs. The residues of state vectors in the sub-space generated by the global EOFs analysis thus contain mostly informations about the short range motions. Therefore, by applying a local EOFs analysis on these residues we can obtain a local EOFs basis which summarizes the local variability. By combining the local and the global basis we will possess a representation basis capable to represent the long range as well as the short range oceanic phenomena. Moreover, as it has been said, this global part of this basis can be made to evolve with model dynamic (see section 3).

To construct such basis, one first computes a global EOFs basis L from a sample of ocean states $X_1, \ldots, X_N$ as usual. The reconstruction formula of these states in the sub-space generated by L is then given by (see section 2.1)

$$X = \bar{X} + LL^T \mathcal{M}(X - \bar{X}) + e \tag{14}$$

where e is the representation error (or residue) of state vector X in the sub-space generated by L. Next, by applying a local EOFs analysis (not centered) on the

representation residues $e_1, \ldots, e_N$ of the state vectors set $X_1, \ldots, X_N$, one obtains a reconstruction formula for the residues according to

$$e = BB^T\mathcal{M}e + e^{'}. \tag{15}$$

where the matrix B contains the first local EOFs and $e^{'}$ is the representation error of the residue e in the sub-space generated by B. Note that we do not center the vector e, as it represents a residue and thus should be already centered. Now, by replacing e by its value in (14), one gets

$$\begin{aligned} X &= \bar{X} + LL^T\mathcal{M}(X-\bar{X}) + BB^T\mathcal{M}e + e^{'} \\ &= \bar{X} + LL^T\mathcal{M}(X-\bar{X}) + BB^T\mathcal{M}(X-\bar{X}) - BB^T\mathcal{M}LL^T\mathcal{M}(X-\bar{X}) + e^{'} \\ &= \bar{X} + L^B(L^B)^T\mathcal{M}(X-\bar{X}) - BB^T\mathcal{M}LL^T\mathcal{M}(X-\bar{X}) + e^{'} \end{aligned} \tag{16}$$

with

$$L^B = [\, L \vdots B \,]. \tag{17}$$

But since e is $\mathcal{M}-$orthogonal to L, B is too, hence we have $B^T\mathcal{M}L = 0$. Therefore, one obtains a new reconstruction formula for a state vector X namely

$$X = \bar{X} + L^B(L^B)^T\mathcal{M}(X-\bar{X}) + e^{'} \tag{18}$$

which can be much more accurate than the one obtained from a simple EOFs analysis (1). Finally, as in section 2.1, one takes the matrix L^B, called mixed EOFs basis, as a representation basis of the variability of the system state.

Note that one can also start by applying a local EOFs analysis and then computes a global basis from the residues of the state vectors in the sub-space generated by the local basis. However, we have noticed in our experiments that the representation basis obtained from the first method seems to have better reprentativity.

3 The semi-evolutive partially local filter

We shall adopt the notation proposed by *Ide et al.* [14]. Consider a physical system described by

$$X^t(t_k) = M(t_{k-1}, t_k)X^t(t_{k-1}) + \eta(t_k) \tag{19}$$

where $X^t(t)$ denotes the vector representing the true state at time t, $M(s,t)$ is an operator describing the system transition from time s to time t and $\eta(t)$ is the system noise vector. At each time t_k, one observes

$$Y_k^o = H_k X^t(t_k) + \varepsilon_k \tag{20}$$

where H_k is the observational operator and ε_k is the observational noise. The noises $\eta(t_k)$ and ε_k are assumed to be independent random vectors with mean zero and covariance matrices Q_k and R_k, respectively.

Direct application of the extended Kalman (EK) filter to data assimilation in meteorology and oceanography is not possible due to the huge dimension (n) of the considered system. To overcome this problem, *Pham et al.* [21] proposed a suboptimal EK filter very close to the original EK, called SEEK filter, in which the error covariance matrix was assumed to be singular with a low rank $r << n$. This leads to a filter in which the errors correction is made only along certain directions parallel to a linear subspace of dimension r. They are the directions for which error is not sufficiently attenuated by the system dynamic. Indeed, the ocean is fundamentally a forced dissipative dynamical system which possesses an attractor. This means that the trajectories of the system are drawn towards a small dimension subspace of the phase space. Since the errors in the directions perpendicular to the attractor will be naturally attenuated one do not need to correct them in these directions. For more details one can consult *Pham et al.* [21].

Although the cost reduction of the EK filter is very important, the SEEK filter remains expensive in real operational assimilation since the evolution of its correction basis L_k requires $r+1$ times the model run. To reduce the cost of this filter, one should simplify the evolution of its correction basis. *Brasseur et al.* [4] proposed to keep the initial EOFs basis fixed in time as they have noticed in there numerical experiments that most of the estimation error was reduced immediately after the first correction i.e. while the evolution of the EOFs basis was not yet effective yet. This approach leads to a very low cost filter, called singular fixed extended Kalman (SFEK) filter. This filter provides an interesting cheap way to test the relevance of the EOFs basis.

With the same aim in view, *Hoteit et al.* [13] proposed different degraded forms of the SEEK filter and its interpolated variant, which are less costly and yet perform reasonably well. Among them, the singular semi-evolutive extended Kalman

(SSEEK) filter can be most easily adapted to be used with the present global-local approach. This filter consists essentially in letting only evolve a few basis vectors of the correction basis while keeping the others fixed. Moreover, the basis vectors which do not evolve are those which contribute the least to the error representation as this would minimize the effect of keeping them fixed.

Here, our idea is to let the global part of the mixed EOFs basis evolve according to the model while leaving the local part fixed in time, as we can not let it to evolve without losing its locality property. However, we can not use the SSEEK filter directly as described in *Hoteit et al.* [13] because this will destroy the partition of the mixed EOFs basis as global-local parts. Thus we simply use the evolution equation of the correction basis of the SEEK filter to let the global part of the mixed EOFs basis evolve while keeping the local part fixed in time as in the SFEK filter. This results in a filter called semi-evolutive partially local filter Kalman (SEPLEK) filter. Note that we have only constructed this version associated with the SEEK filter but not that associated with the (its interpolated version) SEIK filter, due to some technical difficulties.

The SEPLEK filter proceeds in two stages after an initialization stage just as the SEEK filter. It may be desirable, for numerical stability reasons, to normalize the evolutive basis vectors at each filtering steps. This amounts to dividing these vectors by suitable constants. Therefore, in the representation of the forecast error covariance matrix in the form $L_k U_{k-1} L_k^T$, one should then multiply the rows and the columns of U_{k-1} of the same index as that of the evolutive basis vectors by the same constants. The filter operates, as usual, in two stages, excluding the initialization, which can be summarized as follows.

- *Initialization stage*: To initialize the filter, we make a long historical run of the model. Then we perform a mixed global-local EOFs analysis which yields a mixed EOFs basis L_0^B factorized as

$$L_0^B = [\, L \vdots B \,] \tag{21}$$

  with $(r+s)$ columns, r and s denoting the dimensions of the global and local bases respectively. However, such an analysis does not readily provides a rank $(r+s)$ error covariance matrix for stat vector. For this purpose, we shall resort to an objective analysis, based on the first observation Y_0^o: we take as the

initial analysis state vector

$$X^a(t_0) = \bar{X} + L_0^B[(L_0^B)^T \mathbf{H}_0^T R_0^{-1} \mathbf{H}_0 L_0^B]^{-1} (L_0^B)^T \mathbf{H}_0^T R_0^{-1} (Y_0^o - H_0 \bar{X}) \quad (22)$$

where $\bar{X}$ is the average of the state vectors (from the historical run) and $\mathbf{H}_0$ is the gradient of H_0 evaluated at $\bar{X}$. The initial analysis error covariance matrix may be taken as

$$P^a(t_0) = L_0^B U_0 (L_0^B)^T \quad (23)$$

where

$$U_0 = [(L_0^B)^T \mathbf{H}_0^T R_0^{-1} \mathbf{H}_0 L_0^B]^{-1}. \quad (24)$$

Note that we have used the first observation for initialization, the algorithm actually starts with the next observation.

1- Prediction stage: At time t_{k-1}, an estimate $X^a(t_{k-1})$ of the system state and its corresponding error covariance matrix $P^a(t_{k-1})$, in the factorized form

$$P^a(t_{k-1}) = L_{k-1}^B U_{k-1} (L_{k-1}^B)^T \quad (25)$$

where the mixed correction basis $L_{k-1}^B = [\, L_{k-1} \vdots B \,]$ and the matrix U_{k-1} are of dimension $n \times (r+s)$ and $(r+s) \times (r+s)$ respectively, are available. The model (19) is used to forecast the state as

$$X^f(t_k) = M(t_{k-1}, t_k) X^a(t_{k-1}). \quad (26)$$

The corresponding forecast error covariance matrix is approximated by

$$P^f(t_k) = L_k^B U_{k-1} (L_k^B)^T + Q_k \quad (27)$$

where the mixed correction basis is taken equal to

$$L_k^B = [\, L_k \vdots B \,]. \quad (28)$$

The global part of the correction basis L_k evolves as in the SEEK filter (see *Pham et al.* [21]), i.e.

$$L_k = \mathbf{M}(t_k, t_{k-1}) L_{k-1} \quad (29)$$

where $\mathbf{M}(t_k, t_{k-1})$ is the gradient of $M(t_{k-1}, t_k)$ evaluated at $X^a(t_{k-1})$.

Finally, we shall normalize the evolutive basis vectors such that they have norm equal to a constant c, to be chosen. This amounts to replacing the j^{th} column L_k^j of L_k by the vector $(c/\|L_k^j\|_{\mathcal{M}})L_k^j$ and then dividing the j^{th} column and the j^{th} row of U_{k-1} by the constant $c/\|L_k^j\|_{\mathcal{M}}$.

2- Correction stage: The new observation Y_k^o at time t_k is used to correct the forecast according to

$$X^a(t_k) = X^f(t_k) + G_k[Y_k^o - H_k X^f(t_k)], \tag{30}$$

where G_k is the gain matrix given by

$$G_k = L_k^B U_k (L_k^B)^T \mathbf{H}_k^T R_k^{-1}, \tag{31}$$

$\mathbf{H}_k$ the gradient of H_k evaluated at $X^f(t_k)$ and U_k computed from

$$U_k^{-1} = \left[U_{k-1} + P_{L_k^B} Q_k P_{L_k^B}^T\right]^{-1} + (L_k^B)^T \mathbf{H}_k^T R_k^{-1} \mathbf{H}_k L_k^B, \tag{32}$$

with $P_{L_k^B}$ represents the projection operator onto the correction basis, i.e.

$$P_{L_k^B} = [(L_k^B)^T L_k^B]^{-1} (L_k^B)^T. \tag{33}$$

The corresponding filter error covariance matrix is then equal to

$$P^a(t_k) = L_k^B U_k (L_k^B)^T. \tag{34}$$

The SEPLEK filter can be much less costly than the SEEK filter since the dimension of the global part of the mixed basis can be much smaller ($3 - 5$ for example) than the dimension of the SEEK correction basis ($30 - 50$ for example).

Concerning the choice of the normalizing constant c, we take the mean square of the norms of the EOFs (which are equal to the corresponding eigenvalues), which seems to be a good choice.

4 Algorithm with variable forgetting factor

There are three reasons behind the use of the forgetting factor in the SEEK filter. Firstly, it limits the effective filter memory length by discarding old data. This will attenuate the error propagation and enable the SEEK filter to follow system changes. Secondly, it sets up the gain matrix to avoid the "blow up" phenomena (see *Aström* [2]). Thirdly, it does not require any extra cost for its implementation: the filter equations remain unchanged except for the emergence of the forgetting factor ρ in the time propagation error covariance equation (see *Pham* [21]). Specifically, the updating equation (32) for U_k now changes to

$$U_k^{-1} = \left[\rho U_{k-1} + P_{L^B} Q_k P_{L^B}^T\right]^{-1} + (\mathbf{H} L_k^B)^T R_k^{-1} \mathbf{H}_k L_k^B. \tag{35}$$

With $\rho = 1$, all data have the same weight, but with $\rho < 1$, recent data are exponentially more weighted than old data.

However, the use of a too small forgetting factor when the system evolution is stable would degraded the filer performance especially when there is little information in the measurements. To maximize the benefit of the forgetting factor, we propose to use a variable one (*Sorenson and Sacks* [23]): such factor should be close to 1 when the model is stable and much less then 1 when the model is unstable. Indeed old data should be forgotten more in the last case to adapt the filter to the new model's mode.

To detect the periods of model instability, one can track the filter's state by computing a short term and a long term averages of the prediction error norm, denoted by s_k and l_k respectively. If $s_k \leq l_k$, one may assume that steady conditions have been achieved and consider that the model is in a stable period. In this case, the forgetting factor is set close to unity. If $s_k > l_k$ this is an indication that the model may be in an unstable period, therefore it is better to use a forgetting factor strictly less than 1. In short, ρ_k can be adaptively tuned as follows.

$$\rho_k = \begin{cases} \rho_1^\star \lesssim 1 & \text{if} \quad cs_k < l_k, \\ \rho_2^\star < \rho_1^\star & \text{if} \quad cs_k \geq l_k \end{cases} \tag{36}$$

where c is a tuning constant. Finally, the estimates of s_k and l_k are computed recursively as follows.

$$s_k = \alpha s_{k-1} + (1-\alpha)\|Y_k^o - H_k X^f(t_k)\|, \tag{37}$$

$$l_k = \beta l_{k-1} + (1-\beta)\|Y_k^o - H_k X^f(t_k)\| \tag{38}$$

where α and β are constants chosen such that $\beta \lesssim 1$ and $\alpha < \beta$.

5 Application to altimetric data assimilation in the OPA model of the tropical Pacific

In order to evaluate the performance of our filters, we have implemented them in a realistic setting of the OPA model in the tropical Pacific ocean, under the assumption of a perfect model ($Q_k = 0$). The assimilation is based on the pseudo-observations which are extracted from twin experiments. The SEEK filter is used as a reference to compare the performance of these new filters. The configuration and the characteristics of the model used in our experiments are presented in the next section.

5.1 OPA model in tropical Pacific

In this section we briefly present the model basics and a description of our configuration.

5.1.1 Model description

The OPA model (OPA for Océan PArallélisé) is a primitive equation ocean general circulation model which has been developed at the LODYC laboratory (Laboratoire d'Océanographie DYnamique et de Climatologie) to study large scale ocean circulation. It solves the Navier-Stokes equations which express the momentum balance, the hydrostatic equilibrium, the incompressibility, the heat and salt balance and a non-linear realistic equation of state plus the rigid lid assumption and some hypothesis made from scale considerations. The system equations is written in curvilinear z-coordinates and discretized using the centered second order finite difference approximation on a three dimension generalized "C-grid Arakawa". In this scheme, the scalar variables are computed in the center of the cells and the vector variable in the center of cell faces (see *Arakawa* [1] for details). Time stepping is achieved by two time differencing schemes: a basic leap-frog scheme associated to an Asselin filter for the non-diffusive processes and a forward scheme for diffusive terms. The sub-grid scale physics are a tracer diffusive operators of second order on the vertical, the eddy coefficients being computed from a turbulent closure model (see *Blanke and Delecluse* [3]). On the lateral, diffusive and vicious operators can be either of second or of fourth order. The reader is referred to the OPA reference manual *Madec et al.* [18] for more details.

5.1.2 Model configuration

The model domain covers the entire tropical Pacific basin extending from $120°E$ to $70°O$ and from $33°S$ to $33°N$ and the level depth varies from 0 at the sea surface to $4000m$. Two buffer zones are included between 20° and 33° in the north and south of the domain, for the connection with the sub-tropical gyres. The number of horizontal grid points is 171×59 on 25 vertical levels. The model equations are solved on an isotropical horizontal grid with a zonal resolution 1° and a meridional resolution maximum at the equator of 0.5° and goes down to 2° to north and south boundaries. The vertical resolution is approximatively $10m$ from the sea surface to $120m$ depth then decreases to $1000m$ at the sea bottom. The time step is one hour.

The bathymetry is relatively coarse. It was obtained from Levitus data's mask [17]. The forcing fields are interpolated from the ECMWF (European center for medium-range weather forecasts) reanalysis with monthly variability. It is composed of wind stress and heat, temperature and fresh water fluxes. Zero fluxes of heat and salt and non-slip conditions are applied at solid boundaries.

A second order horizontal friction and diffusion scheme for momentum and tracers is chosen with a coefficient of $2000m^2/s$ in the strip $10°N - 10°S$ and increase up to $10000m^2/s$ at the north and south basins boundaries. The static instabilities are resolved in the turbulent closure scheme.

The model starts from rest (i.e. with zero velocity field) and S and T are stem from seasonal climatologic Levitus data [17].

5.2 Experiments design

5.2.1 The state vector

The state vector is the set of prognostic model variables that must be initialized independently. Since the prognostic variables of the OPA model are the zonal U and meridional V velocities, the salinity S and the temperature T, one should consider the state vector

$$X^t = (\ U\ ,\ V\ ,\ S\ ,\ T\ )^T. \tag{39}$$

However, the observed variable, which is the sea surface height SSH, is a diagnostic variable computed from the barotrope velocity by a complex nonlinear algebraic

equation. Thus, the observation operator H which relies the observed variable to this state vector will be nonlinear. Moreover, in order to determine H one should inverse this equation and this can be very costly. To avoid these difficulties, we will adopt a pseudo-state vector in our experiments which contains the true state vector augmented by the SSH, namely

$$X^t = (\ U\ ,\ V\ ,\ S\ ,\ T\ ,\ SSH\ )^T. \tag{40}$$

In this case, H will always be linear of the form $(0 \vdots I_d)$. Of course, the dimension of the state vector will increase but the increase is insignificant since the SSH is computed only on the sea surface. More precisely, the number of state variable is now $4 \times 171 \times 59 \times 25 + 171 \times 59 = 1\,018\,989$ instead of $4 \times 171 \times 59 \times 25 = 1\,008\,900$.

5.2.2 Filters initialization and EOFs analysis

Following the strategy explained in *Pham et al.* [21], the choice of the initial state estimator flow field and the corresponding error covariance matrix is made through a simulation of the model itself. In the present study, the data for the assimilation experiments is again simulated but in an unrelated way with the above simulation.

Thus, in a first experiment, the model has been spun up for 7 years from 1980 to 1986 with the aim to reach a statistically steady state of mesoscale turbulence. Next, another integration of 4 years is carried out from 1987 to 1990 to generate a historical sequence H_S of model realization. A sequence of 480 state vectors was retained by storing 1 state vector every 3 days to reduce the calculation since successive states are quite similar. Because the state variables in (40) are not of the same nature, we shall in fact apply a multivariate EOFs analysis (local and global). We define a metric $\mathcal{M}$ in the state space to make distance between state vectors independent from unit of measure. We choose $\mathcal{M}$ as the diagonal matrix with diagonal elements being the spatial variances of each state variables, namely U, V, S, T and SSH, average over the grid points.

- *Global EOFs analysis*

Figure 1 plots the number of EOFs and the percentage of variability (or inertia) contained in the sample H_S they explain. From this result, we have chosen to retain $r = 30$ global EOFs in all assimilation experiments, as this achieve 85% of the inertia of the sample and this percentage is not much increased for higher value of r.

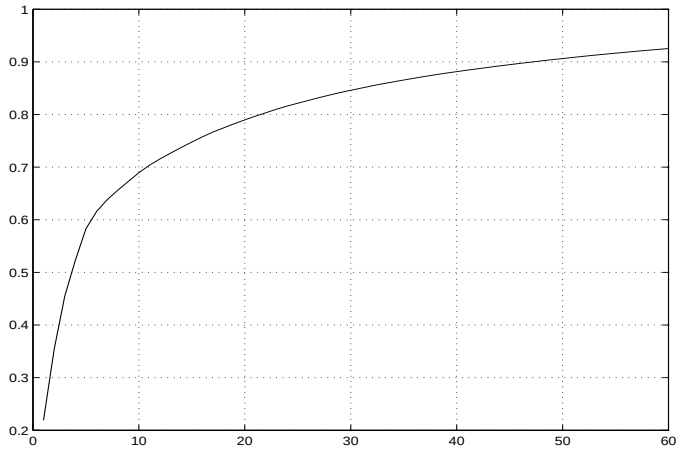

Figure 1: Percentage of inertia versus the number of retained EOFs.

• *Local EOFs analysis*

Since the domain is rectangular with a large width with respect to its height, we have subdivided the ocean domain into three zonal sub-domains to limit the spatial correlation length of the ocean variables in this direction. We have actually considered subdividing the ocean domain also along the meridional, but the results are not as good. The reason may be that the height of the domain is quite short and further it includes two buffer zones in the north and the south.

The choice of the partition of unity functions is shown in Figure 2. The single one-dimensional tensorial function in the meridional $\chi_Y^{(1)}(y)$ direction is then taken constant equal to 1. After applying separately EOFs analysis on the state vectors of these sub-domains, we have retained 18, 30 and 23 EOFs to attain 85% of the inertia in the first, second and third sub-domains respectively. Thus, the dimension of the local EOFs basis is equal to $18 + 30 + 23 = 71$.

• *Mixed EOFs analysis*

We have applied the same previous local EOFs analysis on the residues of the ocean states in the sub-space spanned by the first five global EOFs which contain

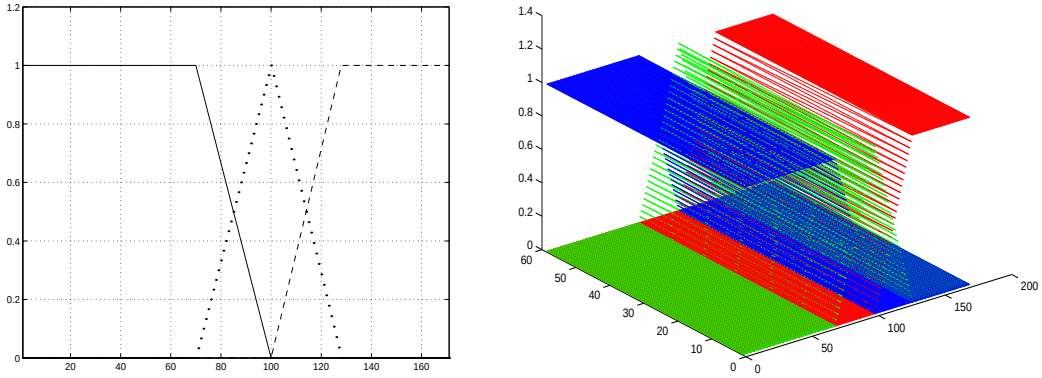

Figure 2: One-dimensional tensorial functions $(\chi_X^{(j)}(x), j = 1, ..., 3)$ and two-dimensional partition of unity functions $(\chi^{(j)}(x, y) = \chi_X^{(j)}(x), j = 1, ..., 3)$.

almost 50% of the total global inertia. Results of these analysis shown that one should consider 19, 19 and 20 EOFs in order to explain 65% of the inertia of the residues in the first, second and the third sub-domain respectively. Therefore, the dimension of the EOFs mixed basis is equal to $5 + 19 + 19 + 20 = 63$.

5.2.3 Data and filters validation

Twin experiments are used to assess the performances and the capabilities of our filters. A reference experiment is performed and the reference X^t retained to be latter compared with the fields produced during the assimilation experiments. More precisely, a sequence of 250 state vectors was retained every $24h$ during the period of Mars 1^{st} 1991 to October 10^{st} 1991.

The assimilation experiments are performed using the pseudo-measurements which are extracted from the reference experiment. The SSH are assumed to be observed at every grid points of the model surface with a nominal accuracy of $3cm$. The observation error is simulated by adding randomly generated Gaussian noise to the synthetic observations of SSH. Note that in the assimilation interval, a period of very strong model instability occur between July and September (see Figure 3).

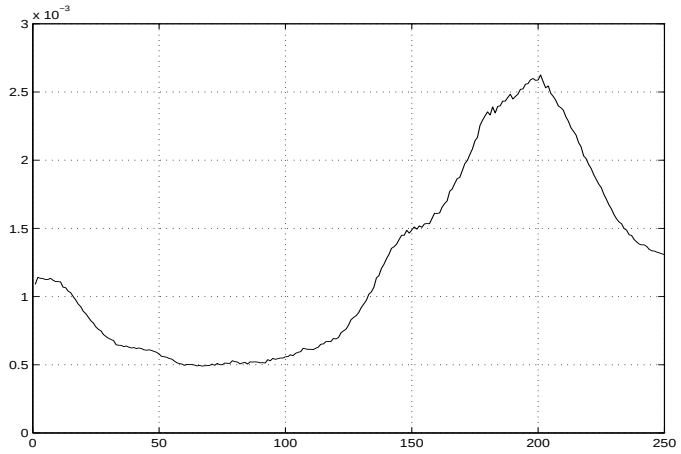

Figure 3: Relative variation of the state vector.

Finally, the performance of all our filters is evaluated by comparing the relative root mean square ($RRMS$) error for each state variable, in each layer or in the whole domain of the ocean model. The $RRMS$ is defined as

$$RRMS(t_k) = \frac{\|X^t(t_k) - X^a(t_k)\|}{\|X^t(t_k) - \bar{X}\|}, \tag{41}$$

where $\bar{X}$ is the mean state of the sample H_S and $\|\cdot\|$ denotes the Euclidien norm. Note that the error is relative to the free-run error since the denominator represents the error when there is no observation and the analysis vector is simply taken as the mean state vector.

5.3 Results of assimilation experiments

The results of the assimilation experiments will be described as follows. Firstly, we present the results of the SEEK filter. Secondly, we test the relevance of the different types of EOFs analysis by comparing their performances with the SFEK filter. Thirdly, we conduct experiment with the SEPLEK filter and compare its results with respect to the SEEK filter and the SFEK filter (with the mixed EOFs basis). Finally, we study the influence of the variable forgetting factor on the SEPLEK filter and the SFEK filter (with the mixed EOFs basis).

• *The SEEK filter*

We first present the assimilation results of the SEEK filter with a fixed forgetting factor equal to $\rho = 0.8$. It can be seen from Figure 4 that the SEEK filter performs well both in the upper layers just and in the lower layers. Although the performances of the SEEK appears to degrade somewhat in the presence of instabilities, it still behave satisfactory during this period too. One may think that the meridional velocity V is not sufficiently well-assimilated because the assimilation error is only reduced by almost to half. But it is worthwhile to point out that, since the velocity field of the tropical Pacific ocean is particularly zonal, the meridional velocity fields are generally, and especially the referenced field in our experiment on Mars 1^{st} 1991, is well-approached by the average of the meridional velocity. Since this average serve as our initial analysis, the initial error is already low and therefore it would be hard to reduce it much further.

We have presented the results of our experiments for the SEEK filter in both the upper and lower layers for completeness. But we have noticed that, for our new filters, the difference between their $RRMS$ and that of the SEEK in the upper lower layer are quite similar. Therefore, in the sequel we will only present results in all layers, to save space.

• *Study of the representativeness of the different EOFs basis*

A simple and low-cost way to test the relevance of the different EOFs analysis presented in section 2 is to use the EOFs basis obtained from these analyses as a fixed correction basis for the SFEK filter and examine its performance. Therefore we have conducted 3 experiments using the SFEK filter with the different correction basis obtained from the global, local and the mixed EOFs analysis. The forgetting factor was chosen fixed equal to 0.8. Results are plotted in Figure 5, and show that the local EOFs basis is much more representative, concerning the variables U, V and SSH, than the global EOFs basis. However the assimilation results with the variables S and T, which may be thought as being controlled by phenomena of long-range variability, are not as good as those obtained from the global basis. Concerning the mixed EOFs basis, the SFEK filter performs very well with the variables U, V and SSH just like the local EOFs basis and moreover it considerably improves the assimilation of S and T with respect to the last basis.

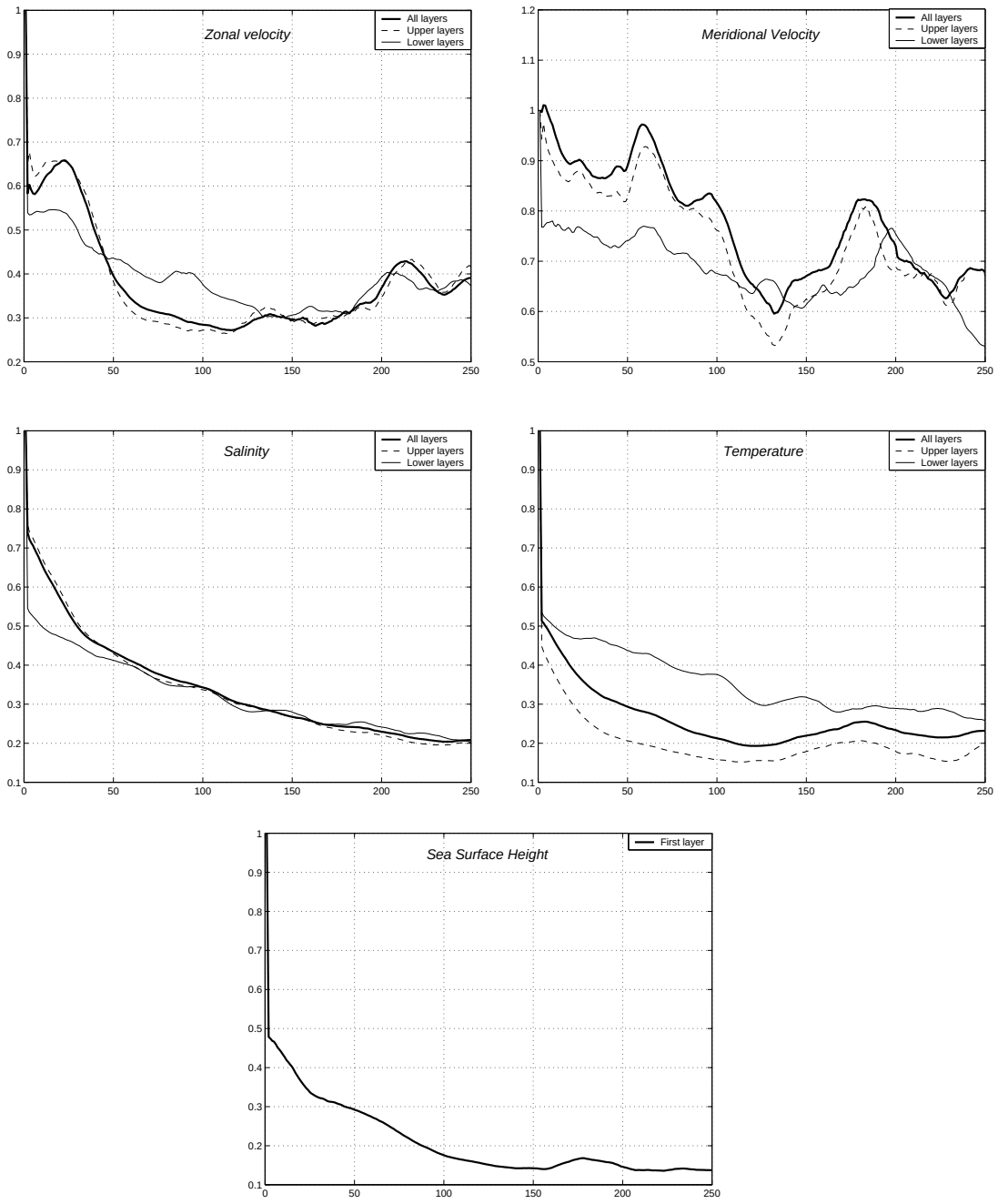

Figure 4: Evolution in time of the $RRMS$ for the SEEK filter on the whole model grid, on the (mean of the 5) upper and the (mean of the 5) lower layers.

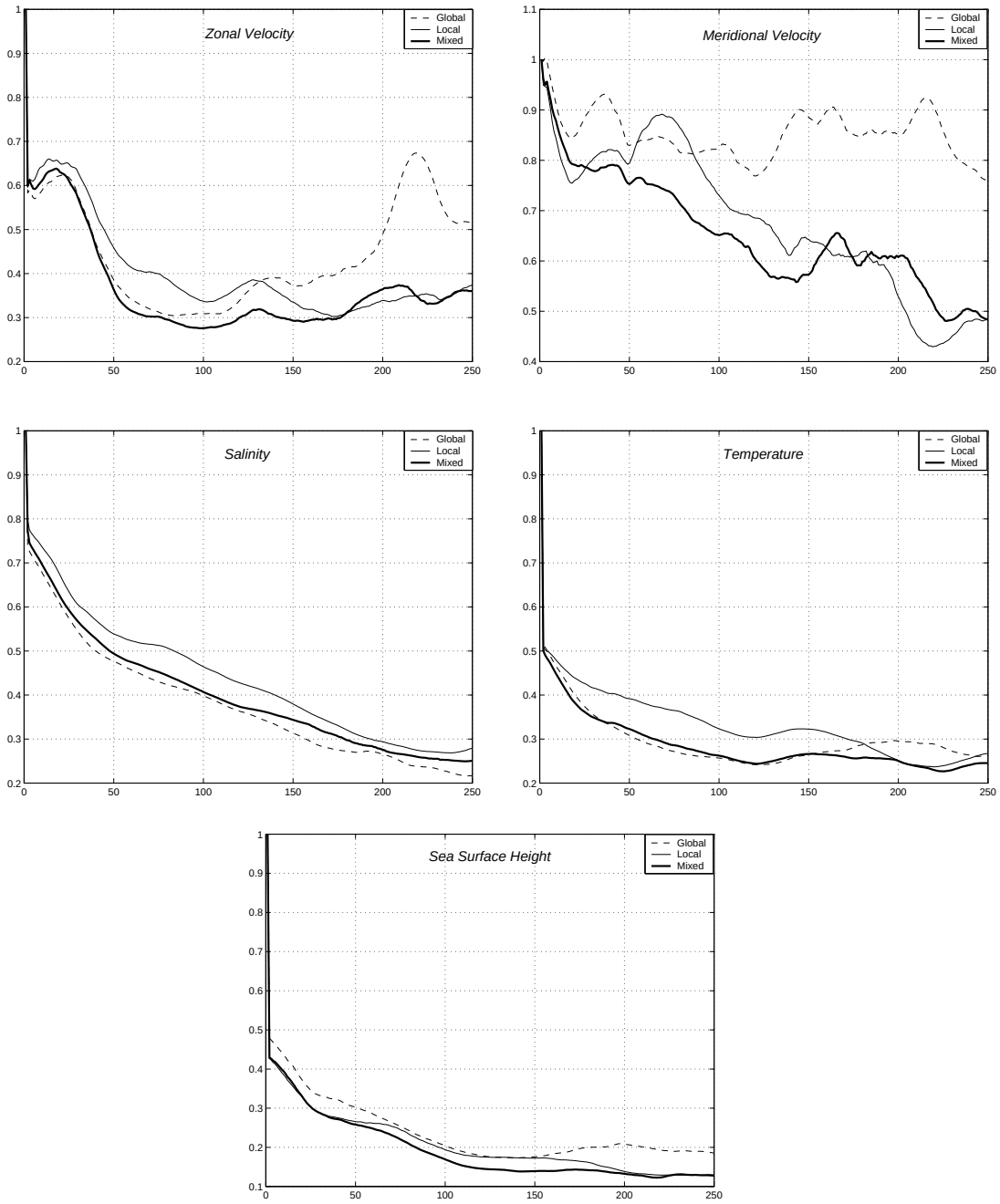

Figure 5: Evolution in time of the $RRMS$ for the SFEK with the global, local and the mixed EOFs basis.

One can also see that the use of a mixed EOFs analysis particular enhances the filter performance in the assimilation of the variable S and T. Thus suggests that the mixed EOFs analysis is able to capture the ocean variability of both the short and the long range phenomena.

Finally it is important to mention here that the divergence of the velocity on the vertical obtained in the OPA model should be equal to zero. But when the local EOFs analysis is applied on the state vectors, the basis vectors do not fulfill this condition on the common border of the ocean sub-domains. To correct this, we have projected the analysis state $X^a(t_k)$ on a zero-divergence subspace before using the model to forecast the state at the next step and this at every filtering step. Despite the loss of precision on $X^a(t_k)$ because of this projection, the local basis was shown to have very good performance.

- *The SEPLEK filter*

Here, we have used the SEPLEK filter with a fixed forgetting factor $\rho = 0.8$. The global part of the mixed basis was made to evolve as in the SEEK filter while the local part is fixed. We shall compare its performance with regard to the SEEK filter. Since the dimensions of the (global) EOFs basis and the global part of the mixed basis were taken equal to 30 and 5 respectively, the SFEK the SEPLEK are respectively almost 30 and 6 times faster than the SEEK filter.

It can be seen from Figure 6 and Figure 9-12 that the SEPLEK performs very well. Its assimilation results of the velocity components U and V and the SSH are even shown to be much better than those obtained with the SEEK filter during the unstable period. Concerning the salinity S and the temperature T, the SEPLEK filter performs almost just as well as the SEEK filter. One can also notice the good influence of the evolution of the global basis part of the mixed correction basis by comparing the results of the SEPLEK filter with those obtained with the SFEK filter when the mixed EOFs basis is used as a fixed correction basis.

- *Testing the variable forgetting factor with the SFEK and SEPLEK filters*

The use of a variable forgetting factor was shown to enhance the performance of the SEEK filter and its variants (see *Hoteit et al.* [13]). Here, we have used a

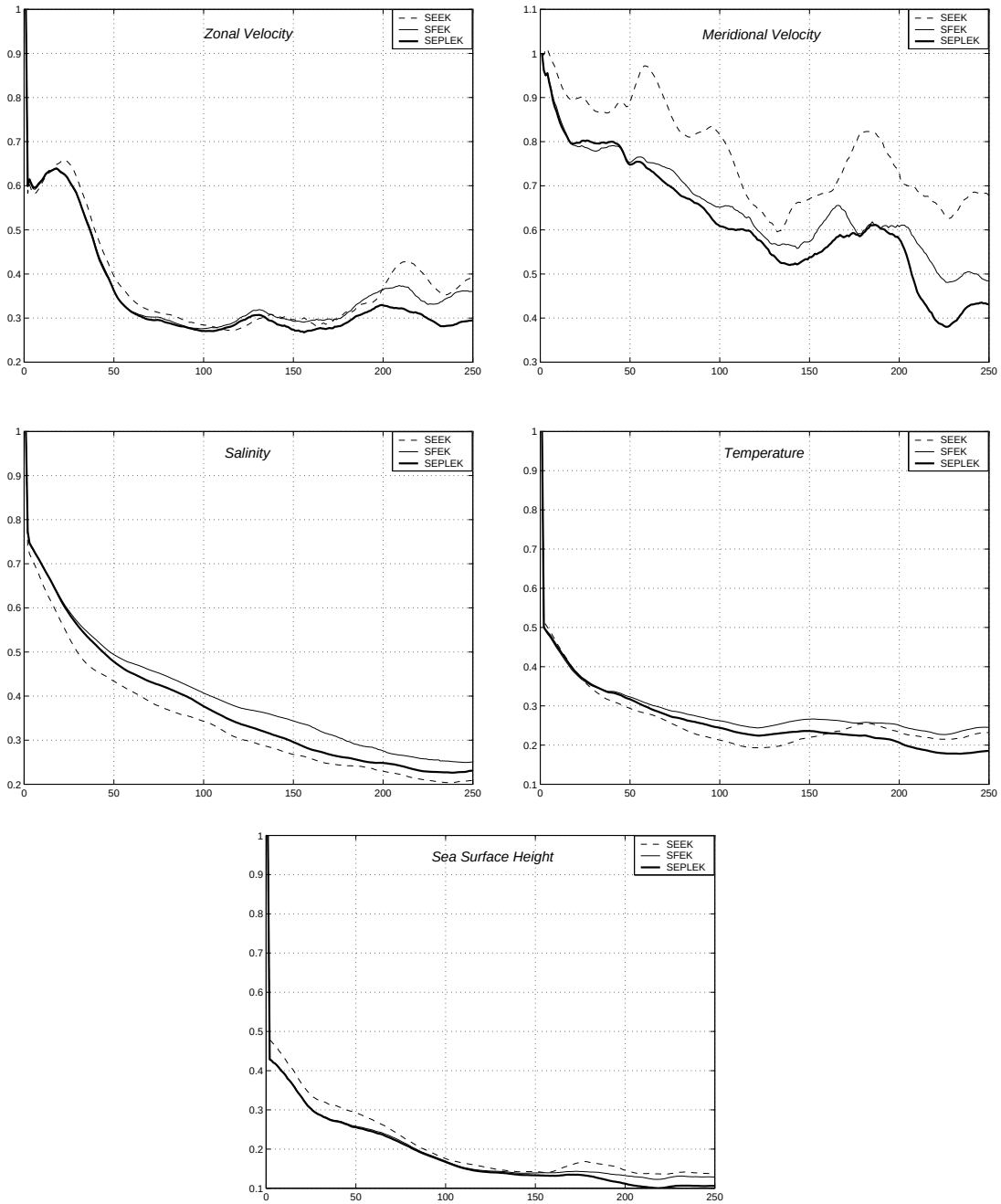

Figure 6: Evolution in time concerning the $RRMS$ for the SFEK (with the mixed EOFs basis), SEPLEK and SEEK filters with a forgetting factor $0,8$.

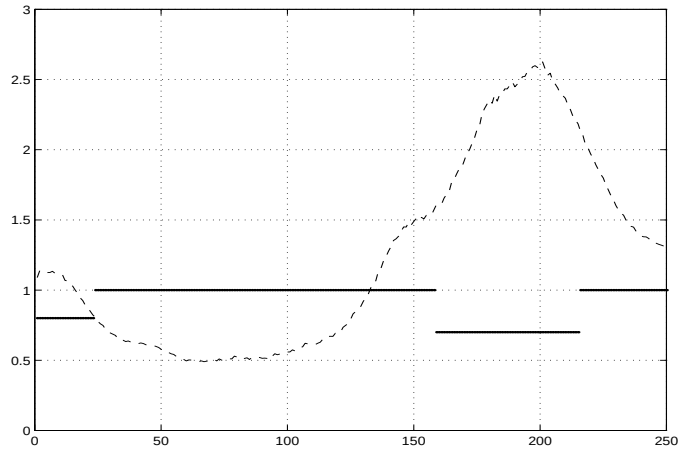

Figure 7: Evolution of the forgetting factor (solid line) and the relative variation of the state vector (dotted line).

variable forgetting factor with SFEK (with the mixed EOFs basis) and SEPLEK filters. Such forgetting factor takes one of the two values 1 or 0.8 according to the relative magnitudes of the short-term and the long-term prediction error s_k and l_k prediction error. The initial values s_0 and l_0 were taken as $\|Y_0^o - H_k X^f(t_0)\|^2$ to make sure that ρ takes the value 0.6 during early assimilation period. The values of the constants α and β were chosen as 0.85 and 0.8 respectively.

Figure 8 shows the $RRMS$ error for these experiments compared with those of the SEPLEK filter with a fixed forgetting factor equal to 0.8. These results show the efficiency of our adaptive tuning scheme of the forgetting factor and, as can be seen form Figure 7, for the detection of the unstable periods.

6 Conclusions

A new data assimilation scheme, called SEPLEK, derived from the SEEK filter has been developed and discussed. The motivation for developing such filter was to reduce the cost of the SEEK filter and possibly obtain a better representativity for its correction basis. Our approach was to construct a mixed EOFs basis composed of global and local EOFs and then to let only evolve the global part. To compute the

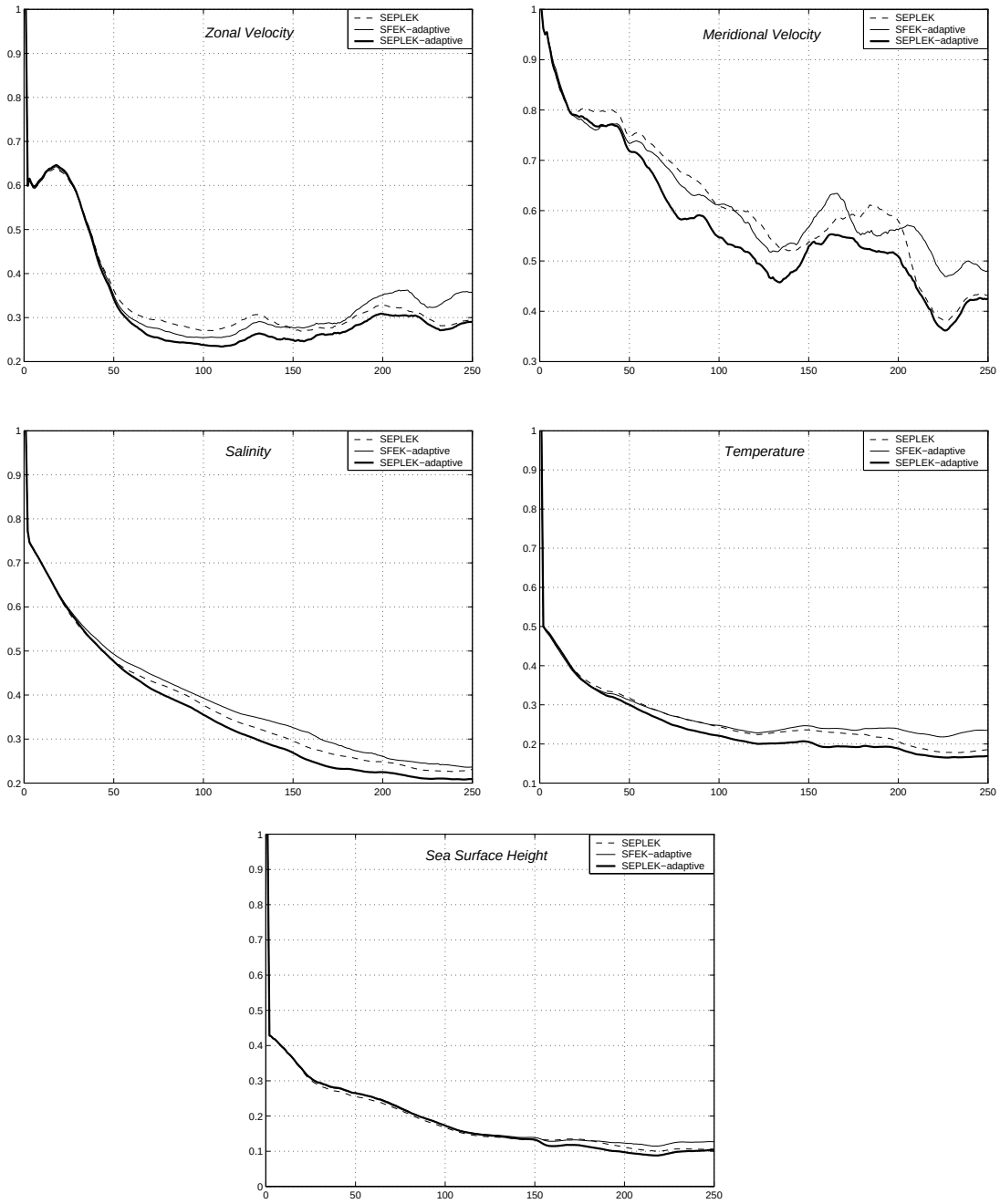

Figure 8: Evolution in time of the $RRMS$ for the SFEK and SEPLEK filters with a variable forgetting factor and for the SEPLEK filter with a constant forgetting factor.

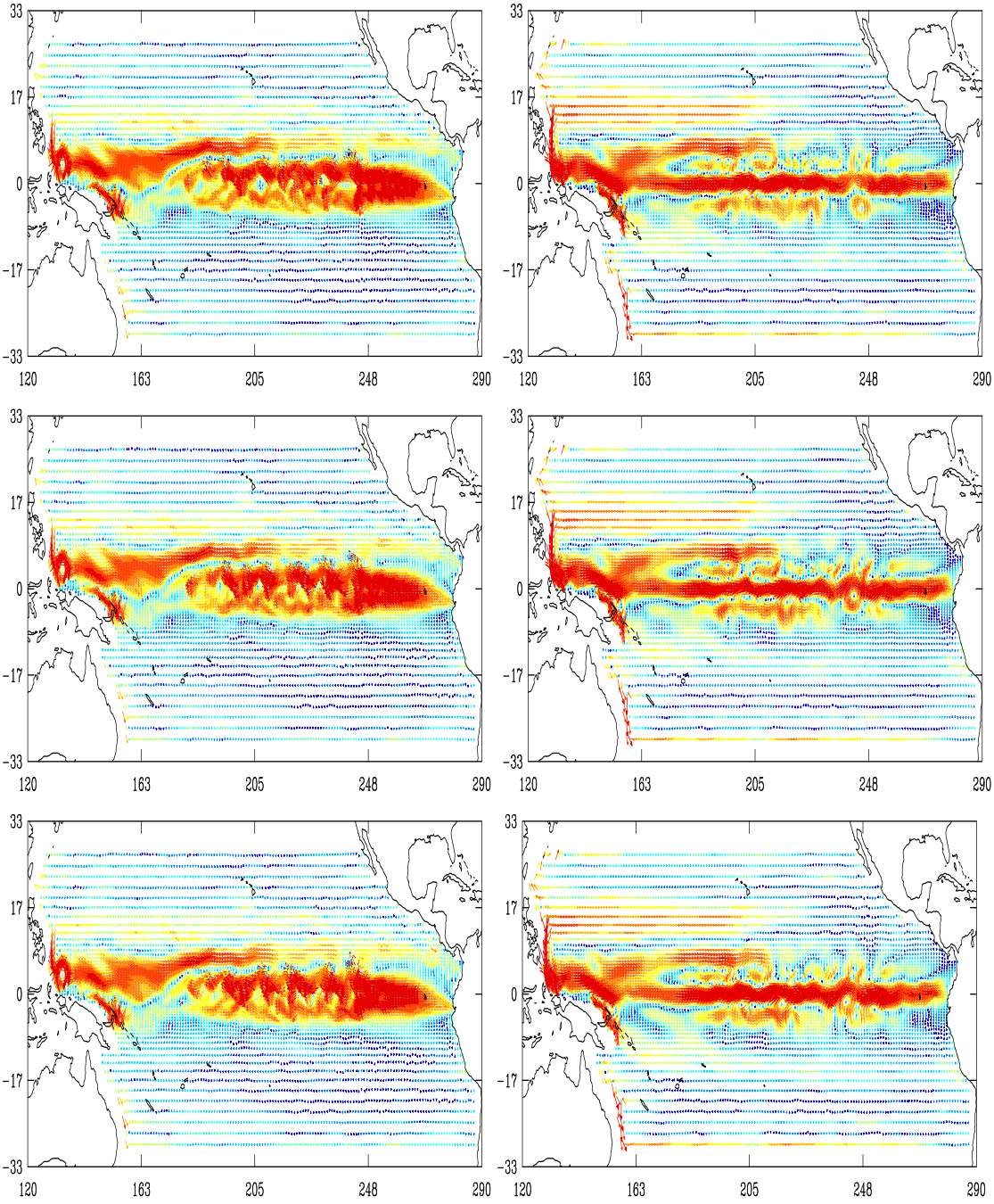

Figure 9: Maps of ocean velocity on Sept 1^{st} 90 in the first (left) and the 17^{th} (right) layers: from the SEEK filter (top); reference (middle); from the SEPLEK filter (bottom).

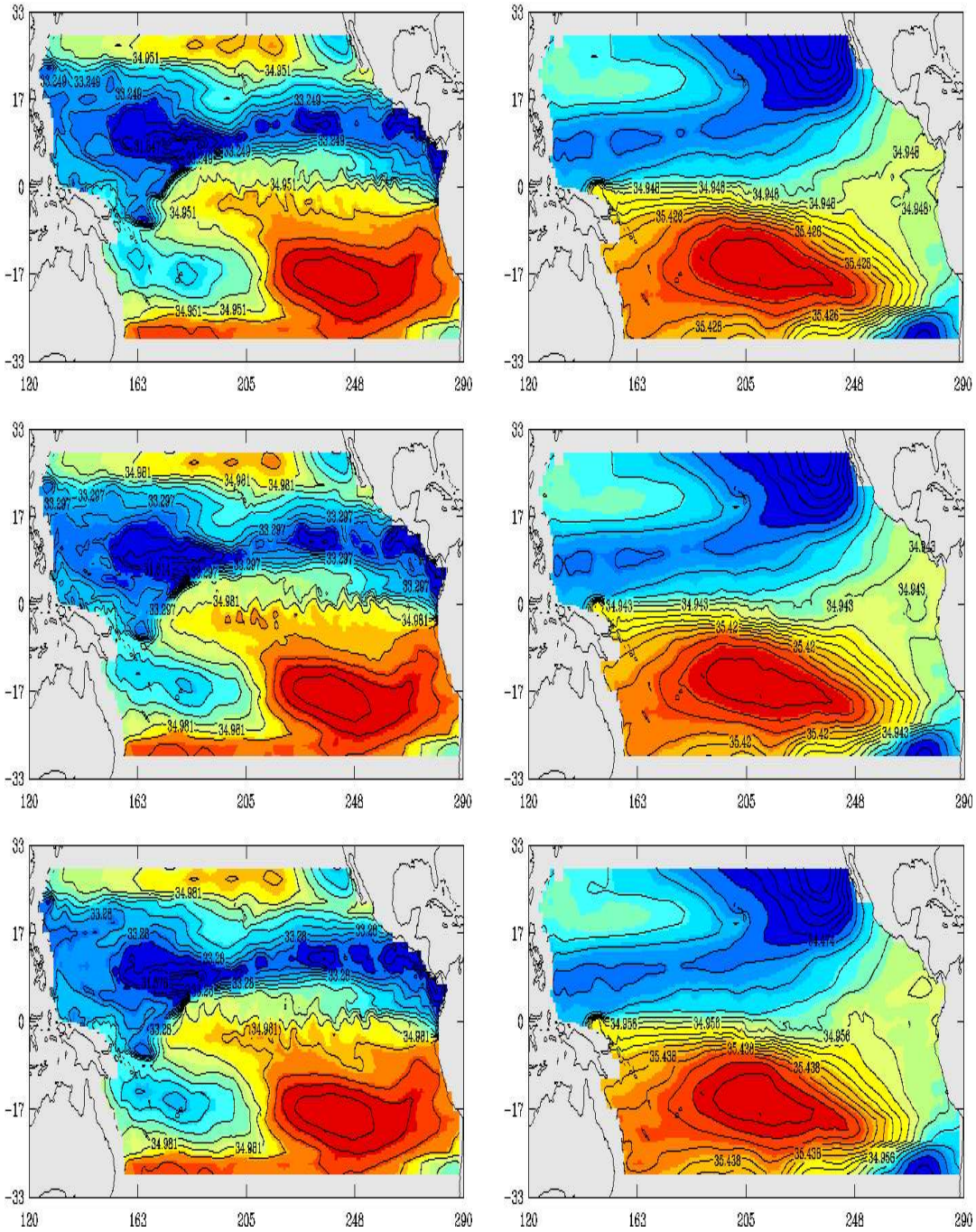

Figure 10: Maps of sea salinity on Sept 1^{st} 90 in the first (left) and the 17^{th} (right) layers: from the SEEK filter (top); reference (middle); from the SEPLEK filter (bottom).

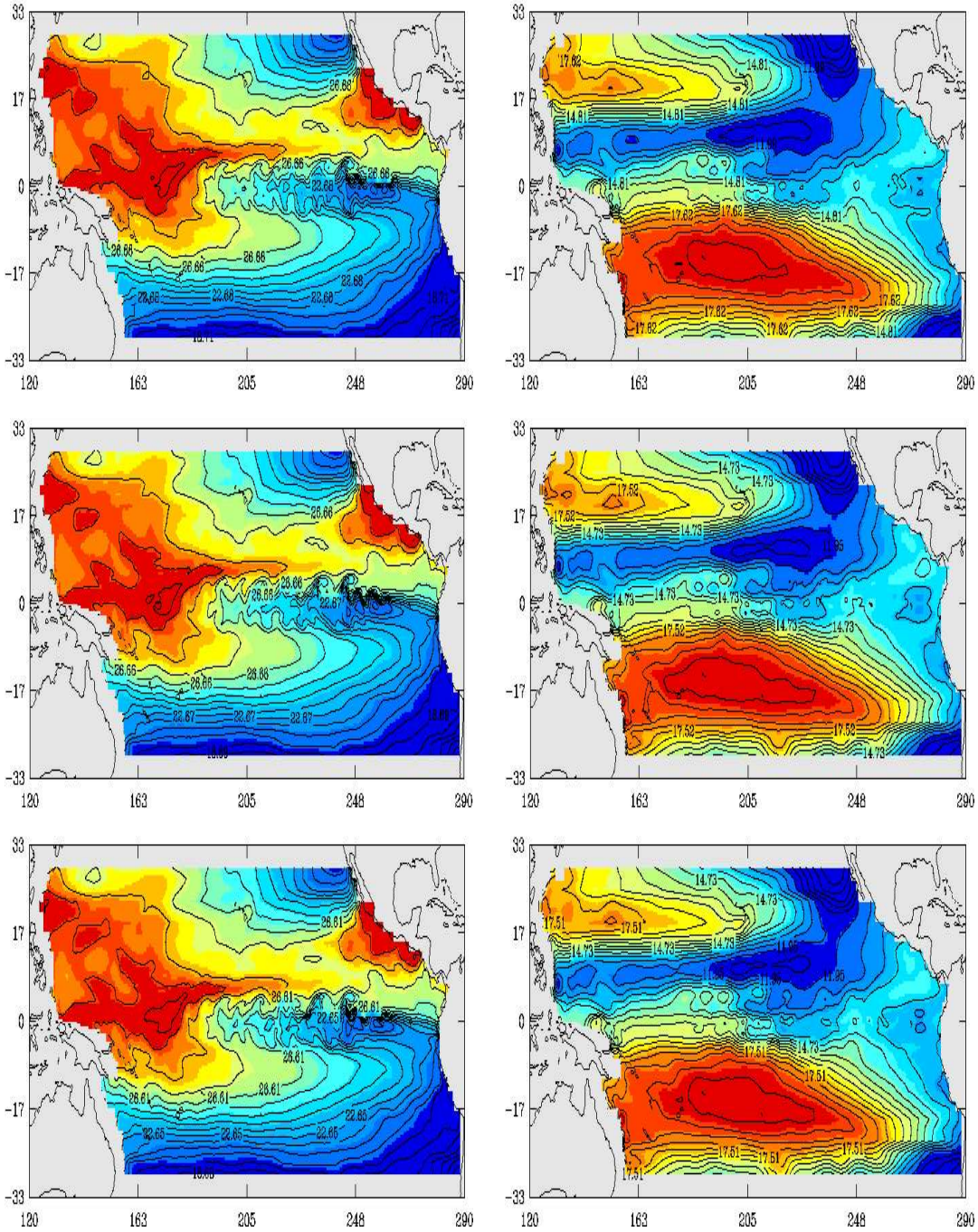

Figure 11: Maps of sea temperature on Sept 1^{st} 90 in the first (left) and the 17^{th} (right) layers: from the SEEK filter (top); reference (middle); from the SEPLEK filter (bottom).

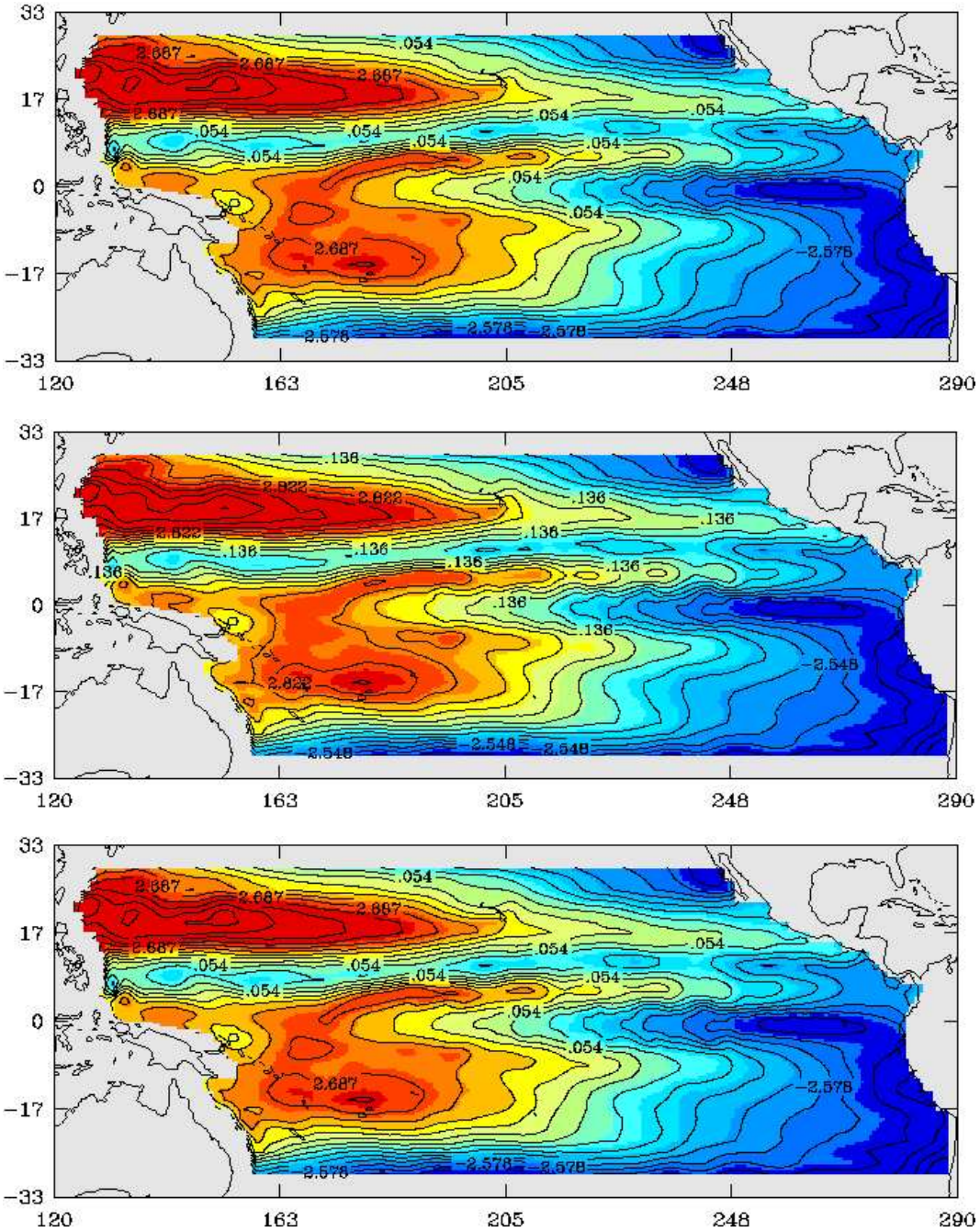

Figure 12: Maps of sea surface pressure on Sept 1^{st} 90: from the SEEK filter (top); reference (middle); from the SEPLEK filter (bottom).

local EOFs, we have introduced the use of a local EOFs analysis on the residues of the ocean states in the sub-space generated by the global part. In this new EOFs analysis, the EOFs supported a small sub-domain of the ocean and vanished elsewhere. Indeed, our aim was on the one hand to limit the spatial correlation length of the ocean variables and on the other hand to construct a set of EOFs basis suited for every sub-domains according to its variability. A series of twin experiments was conducted with the OPA model in the tropical Pacific. Our aim was first to test the relevance of the representativeness of the different type of the EOFs basis representativeness and second to assess the feasibility of the SEPLEK filter by evaluating its performance with regard to the SEEK filter. Our main conclusions are as follows.

1- The local EOFs basis represents much better the variability of the velocity and the sea surface height than the classical EOFs basis but not of the salinity and the temperature which seen to be essentially of global variability. However, finding a way to let the local basis evolve with the model dynamic remains an open problem.

2- The mixed EOFs basis is shown to perform very well even if we keep it fixed. It especially enhances the filter performance concerning the salinity and the temperature, with respect to the local EOFs basis. When its global part is made to evolve as in the SEEK filter, it provides a dynamically evolutive filter which performs even better than the SEEK filter, but with a much lower cost. One can also note the good influence of the evolution of the global part of this basis.

3- By tracking the prevision errors one can obtain informations about the filter state and then adapt the filter parameters to the present situation. In particular, the adaptive tuning of the forgetting factor considerably enhances the performances of the SEEK filter and its variants.

In twin experiments, our SEPLEK filter combined with the mixed EOFs basis was found to be very effective in assimilating of surface-only pseudo-altimeter data. Further works will consider more realistic situations, like the addition of the model error or the use of more realistic observations (according to satellite tracks and real data from satellite). However, these preliminary twin experiments applications were necessary steps before realistic applications and provide us with encouraging results as regard to that purpose.

References

[1] Arakawa A. (1972): Design of the UCLA general circulation model. Numerical integration of weather and climate. *Dept. of Meteorology*, University of California, **Rep. 7**, 116 pp.

[2] Astrom K.J. and B. Wittenmark (1989): Adaptive control. *Addison-Wesley publishing company*.

[3] Blanke B. and P. Delecluse (1993): Variability of the tropical Atlantic ocean simulated by a general circulation model with two different mixed layer physics. *J. Phys. Oceanogr.*, **23**, 1363-1388.

[4] Brasseur P., J. Ballabrera-Poy and J. Verron (1999): Assimilation of altimetric observations in a primitive equation model of the gulf stream using a singular evolutive extended Kalman filter. *J. Mar. Systems*, **22**(4), 269-294.

[5] Cane M.A., A. Kaplan, R.N. Miller, B. Tang, E.C. Hackert and A.J. Busalacchi (1995): Mapping tropical Pacific sea level: data assimilation via a reduced state Kalman filter. *J. Geophys. Res.*, **vol.101**, no.C10, 599-617.

[6] Dee D.P. (1990): Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.*, **vol.117**, 365-384.

[7] Evensen G. (1994): Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophysical Research*, **vol.99**, no.C5, 10143-10162.

[8] Evensen G. (1992): Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophysical Research*, **vol.97**, no.C11, 17905-17924.

[9] Fukumori I. and P. Malanotte-Rizzoli (1995): An approximate Kalman filter for ocean data assimilation: an example with an idealized gulf stream model. *J. Geophys. Res.*, **100**(C4), 6777-6793.

[10] Gauthier P., P. Courtier and P. Moll (1993): Assimilation of simulated wind Lidar data with a Kalman filter. *Mon. Wea. Rev*, **vol.121**, 1803-1820.

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

[12] Hoang H.S., P. De Mey, O. Tallagrand and R. Baraille (1995): Assimilation of altimeter data in multilayer quasi-geostrophic model by simple nonlinear adaptive filter. *Proc. Internat. Symposium Assimilation Obser. Meteo. Oceano.*, 521-526, Tokyo, Japan.

[13] Hoteit I., D.T. Pham and J. Blum (2000): Efficient reduced Kalman filtering and application to altimetric data assimilation in tropical pacific. *Rapport de recherche*, **RR-3937**, INRIA. Submitted to *J. Mar. Syst.*.

[14] Ide K., A.F. Bennett, P. Courtier, M. Ghil and A.C. Lorenc (1995): Unified notation for data assimilation: operational, sequential and variational. Submitted to the *J. Met. Soc.*, Japan.

[15] Jazwinski A.H. (1970): Stochastic processes and filtering theory. *Academic Presse*, New York.

[16] Kalman R.E. (1960): A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D, J. Basic Eng.*, **82D**, 35-45.

[17] Levitus S. (1982): Climatological atlas of the world ocean. *Geophysical fluid dynamics laboratory*, Princeton.

[18] Madec G., P. Delecluse, M. Imbard and C. Levy (1997): Ocean General Circulation Model Reference Manual. *Technical report*, University Pierre and Marie Curie, Paris VI.

[19] Morrisson D.F. (1978): Multivariate statistical methods. *McGraw-Hill International Editions*, 415pp.

[20] Pham D.T., J. Verron and L. Gourdeau (1998): Singular evolutive Kalman filter for data assimilation in oceanography. *C. R. Acad. sci. Paris*, **vol.326**, 255-260.

[21] Pham D.T., J. Verron and M.C. Roubaud (1997): Singular evolutive Kalman filter with EOFs initialization for data assimilation in oceanography. *J. Mar. Syst.*, **vol.** 16, 323-340.

[22] Preisendorfer R. (1988): Principal Component Analysis in meteorology and oceanography. *Elsevier Sci. Publ.*, **17**, 425 pp.

[23] Sorenson H.W. and J. E. Sacks (1971): Recursive fading memory filtering. *Information Sciences*, **3**, 101-119.

[24] Verron J., L. Gourdeau, D.T. Pham, R. Murtugudde and A.J. Busalacchi (1998): An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical pacific: method and validation. *J. Geophys. Res.*, **vol. 104**, C3, 5441-5458.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399