# Evolutivity of the Reduced State Space and Data Assimilation Schemes Based on The Kalman Filter

Ibrahim Hoteit, Dinh-Tuan Pham

INRIA
RHÔNE-ALPES

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Evolutivity of the reduced state space and data assimilation schemes based on the Kalman filter*

Ibrahim Hoteit and Dinh-Tuan Pham

**No 4283**

Septembre 2001

THÈME 4


ISSN 0249-6399 ISRN INRIA/RR--4283--FR+ENG

# Evolutivity of the reduced state space and data assimilation schemes based on the Kalman filter

Ibrahim Hoteit and Dinh-Tuan Pham

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IDOPT

Rapport de recherche n ° 4283 — Septembre 2001 — 31 pages

**Abstract:** Brute-force implementation of the extended Kalman (EK) filter in realistic ocean models is not possible because of its prohibitive cost. Different degraded forms of the EK filter, which basically reduce the dimension of the system through some kind of projection onto a low dimensional subspace, have been proposed [4, 6, 10, 14]. The goal of this paper is to study the usefulness of the evolution in time of these reduced order spaces. This is based on the comparison, both from the theoretical and practical points of view, of the singular evolutive extended Kalman (SEEK) filter introduced by *Pham et al.* [25] and the reduced-order extended Kalman (ROEK) filter introduced by *Cane et al.* [4]. To reduce the cost of the ROEK filter, we further approximate the nonlinear dynamics of the system by a first order autoregressive stochastic model. Finally, using a twin experiment approach, the above filters are compared in assimilation experiments with altimetric data in a realistic setting of the OPA model in the tropical Pacific ocean.

**Key-words:** Data assimilation. OPA model. Reduced Kalman filtering. SEEK Filter. ROEK Filter. EOF 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.

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

# Evolutivité de l'espace d'état réduit et schémas d'assimilation de données basés sur le filtre de Kalman

**Résumé :** Une implementation directe du filtre de Kalman étendu (EK) dans les modèles océaniques réalistes est impossible en pratique vu son coût prohibitif. Plusieurs variantes sous-optimales du filtre de Kalman, basées sur des approximations qui consistent essentiellement à projeter l'état du système sur un sous-espace de dimension faible, ont été proposées pour réduire la dimension du système [4, 6, 10, 14]. Le but de ce papier est d'étudier l'utilité de faire évoluer dans le temps ces espaces d'ordre réduit. Notre démarche consiste en la comparaison, du point de vue théorique et pratique, entre le filtre de Kalman étendu singulier évolutif (SEEK) introduit par *Pham et al.* [25] et le filtre de Kalman étendu d'ordre réduit (ROEK) introduit par *Cane et al.* [4]. Pour réduire le coût du filtre ROEK, nous introduisons aussi l'utilisation d'un modèle statistique autorégressif du premier ordre pour faire évoluer l'erreur de prévision dans le sous-espace réduit. Finalement, en conduisant des expériences jumelles, nous comparons les performances de ces filtres en assimilant les données altimétriques dans une configuration réaliste du modèle OPA dans l'océan tropical Pacifique.

**Mots-clé :** Assimilation de données. Modèle OPA. Filtrage de Kalman réduit. Filtre SEEK. Filtre ROEK. Analyse EOF. Facteur d'oubli.

# 1 Introduction

*Data assimilation* refers to the methodology which aims to incorporate, in an efficient way, optimal if possible, all the available informations, namely the model and the observations, to obtain the best description of the state of a dynamical system. Roughly speaking, the observations guide the model towards a realistic trajectory, while the model provides a spatio-temporal dynamic interpolation for the observations. This technique, which was widely used in meteorology, has only recently been actively developed in oceanography, thanks to several satellite observation missions which provided a large number of measurements and to the continuous progress of computers. One can classify the assimilation methods in two principal categories: *sequential methods* based on the statistical estimation theory and *variational methods* based on the optimal control theory (see *Ghil and Manalotte-Rizzoli* [12] for a review). The present work belongs to the first category and is derived from the well known *Kalman filter*.

The Kalman filter is a statistical data assimilation scheme which provides the best estimate, in the sense of least-squares, of the state of a linear system using all observations available up to the analysis time (see [20]). This filter is easy to implement. However, its application in realistic ocean models encounters two major difficulties: *non-linearity* and *computational cost*. The first can be partially solved by linearizing the model around the state estimate, which leads to the so called extended Kalman (EK) filter [19]. The second is due to the huge dimension of the model state. Several variants of the EK filter, which essentially consist in projecting the system state, via an order-reduction operator, onto a low dimensional subspace, have been proposed to reduce the dimension of the system [4, 6, 8, 9, 10, 14]. Among these filters, of particular interest to us is the one proposed by *Cane et al.* [4] in which the state space is reduced through the projection onto a linear subspace spanned by a small set of basis function, using an empirical orthogonal function (EOF) analysis. We will refer to this filter as the reduced-order extended Kalman (ROEK) filter.

In all the above "suboptimal" Kalman filters, the reduction operator is usually considered to be constant in time, or to follow empirical laws [7]. A consequence of this assumption is that the error covariance dynamics may not be properly taken into account [7]. Recently, *Pham et al.* [25] have proposed a new variant of the Kalman filter, called singular evolutive extended Kalman (SEEK) filter, in which the reduction operator is allowed to evolve with time to follow the dynamic of the model. Otherwise, *Pham et al.* [25] have followed the same approach of EOF analysis

of *Cane et al.* [4] to initialize the reduction operator.

Our work has two man goals:

- We would like to study the usefulness of the evolution of the reduction operator by comparing the performances of the SEEK and ROEK filters. Indeed, these two filters start with the same reduction operator (obtained from an empirical orthogonal functions (EOF) analysis). However, the SEEK filter allows the reduction operator to evolve in time, while the ROEK filter replaces the evolution of the reduction operator, which now remained fixed, by the evolution of the forecast error in the reduced dynamic defined by this operator.
- *Hoteit et al.* [15] have introduced the global-local (or mixed) EOF analysis as an approach to construct a reduction operator which represents the global and the local variability of the ocean. However, the local basis functions obtained from this analysis can not be let to evolve with time as in the SEEK filter without loosing its locality property. Thus they proposed to only let the global basis functions to evolve as in the SEEK filter while keeping the local ones fixed in time. Here, we will implement and test the performance of the ROEK filter with these functions to see if the evolution of the forecast error on the reduced dynamic can compensate the "non-evolutivity" of these basis functions.

On the other hand, the ROEK filter can be as costly as the SEEK filter despite the fact that its reduction operator remains fixed. To reduce the cost of the SEEK filter, *Hoteit et al.* [17] have constructed different variants of the SEEK filter in which they simplify the evolution of the reduction operator, which is the most expensive part of the SEEK filter. To reduce the cost of the ROEK filter, we approximate the nonlinear dynamics of the system by a first order autoregressive stochastic model, which is used in the updating equation for the filter forecast error covariance matrix.

The outline of this paper is as follows. In section 2 we recall the EK filter. In section 3 we describe the ROEK and SEEK filters and we discuss their differences. In section 5 we briefly present the mixed EOF analysis. In section 4 we introduce the use of a first order autoregressive stochastic model in order to reduce the cost of the ROEK filter. Finally, using twin experiment approach, we compare the performances of all these filters in section 6 with some assimilation results based on a realistic setting of the OPA model over the tropical Pacific ocean.

## 2 The extended Kalman filter

We shall adopt the notation proposed by *Ide et al.* [18]. Consider a physical system described by

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

where $X^t(t)$ is a 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 $t_k$, the state vector is observed according to

$$Y_k^o = H_k X^t(t_k) + \varepsilon_k, \tag{2}$$

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

The sequential data assimilation consists in the estimation of the state system $X^t$ at each observation time, using only observations up to this time. In the linear case, this problem has been entirely solved by the well known Kalman filter. In the nonlinear case, one often linearizes the model around the current estimated state vector, which yields to the so called extended Kalman EK filter (see for example *Ghil and Manalotte-Rizzoli* [12] for more details). Apart from initialization, this filter works as a succession of forecasting and correction steps.

**0- Initialization stage:** To initialize the EK filter, one needs some initial state vector $X^a(t_0)$ and its initial error $P^a(t_0)$. Usually very little is known concerning the initial conditions of the system. However, the choice of $X^a(t_0)$ and $P^a(t_0)$ may not be very important thanks to the stability of the Kalman filter [19]. One often takes $X^a(t_0)$ as the average of a set of simulated state vectors and $P^a(t_0)$ as the sample covariance matrix of these vectors.

Assuming that before time $t_k$, one already has a system state estimate, often referred to as the analysis state vector $X^a(t_{k-1})$, with some analysis error covariance matrix $P^a(t_{k-1})$. The EK filter allows the construction of the next analysis vector $X^a(t_k)$ together with its error covariance matrix $P^a(t_k)$ as follows.

**1- Forecasting stage:** The model (1) is used to forecast the state at time $t_k$ as

$$X^f(t_k) = M(t_k, t_{k-1})X^a(t_{k-1}). \tag{3}$$

The forecast error covariance matrix is then given by

$$P^f(t_k) = \mathbf{M}(t_k, t_{k-1}) P^a(t_{k-1}) \mathbf{M}(t_k, t_{k-1})^T + Q_k \tag{4}$$

where $\mathbf{M}(t_k, t_{k-1})$ is the gradient of $M(t_k, t_{k-1})$ evaluated at $X^a(t_{k-1})$ and $^T$ denotes the transpose. The above formula actually is only approximate as it arises from the linearization of (1) around $X^a(t_{k-1})$:

$$X^t(t_k) \approx M(t_k, t_{k-1}) X^a(t_{k-1}) + \mathbf{M}(t_k, t_{k-1}) e^a(t_{k-1}) + \eta(t_k) \tag{5}$$

where $e^a(t_{k-1}) = X^t(t_{k-1}) - X^a(t_{k-1})$ is the analysis error.

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

$$X^a(t_k) = X^f(t_k) + G_k[Y_k^o - H_k X^f(t_k)], \tag{6}$$

where $G_k$ is the so called Kalman gain matrix which is given by

$$G_k = P^f(t_k) \mathbf{H}_k^T [\mathbf{H}_k P^f(t_k) \mathbf{H}_k^T + R_k^{-1}]^{-1} \tag{7}$$

and $\mathbf{H}_k$ is the gradient of $H_k$ evaluated at $X^f(t_k)$. The corresponding analysis error $e^a(t_k)$ has the covariance matrix

$$P^a(t_k) = P^f(t_k) - G_k \mathbf{H}_k P^f(t_k). \tag{8}$$

Brute-force implementation of the EK filter is not possible in practice for large oceanic models. Indeed, knowledge of the requisite error statistics $Q_k$ and $R_k$ is lacking and more importantly, its computational requirement are excessive since the computation of its forecast error covariance matrix $P^f$ requires $2n$ times integrations of the tangent linear model $\mathbf{M}$ to obtain the matrix $\mathbf{M}P^a\mathbf{M}^T$ in (4). Therefore, approximations is unavoidable. The ROEK and SEEK filters are two approaches to reduce the cost of the EK filter. They will be described in the next section.

# 3 The ROEK and SEEK filters

The implementation of the EK filter in the oceanic models can not be done without any massive order reduction of the dimension of the system. As proposed by *Fukumori et al.* [10], the use of a (linear) reduction operator, which relates the state

vector of the system to a small dimension reduced state vector, offers us several alternatives to reduce the cost of the EK filter. Indeed, this can be done by letting the forecast error to evolve in the reduced-order state space, after it has been "transported" in this space, and then to reconstruct this error in its original (or full) space via the pseudo-inverse of the reduction operator to recover the reduced-order analysis error in the full space. One obtains in this way the general form of most variants of the EK filter in practice (see [7]). Note that this might also be quite costly unless the dimension $r$ of the reduced space is much smaller than the dimension of the full space (of the order $10^7$).

### 3.1 The ROEK filter

Without any lost of generality, once can assume that the reduction operator $L$ is orthogonal so that its pseudo-inverse is equal to $L^T$. The full state vector $X^t$ is then related to the reduced state vector $X_r^t$ by

$$X^t = LX_r^t. \tag{9}$$

If this assumption is used in the equations of the EK filter, one obtains the algorithm of the ROEK filter which operates in two stages apart from an initialization stage as the EK filter (the reader can refer to [10] for more details).

**0- Initialization stage:** We resort to an objective analysis, based on the first observation $Y_0^o$: we take as the initial analysis state vector

$$X^a(t_0) = \bar{X} + LP_r^a(t_0)L^T\mathbf{H}_0^T R_0^{-1}[Y_0^o - H_0\bar{X}] \tag{10}$$

where

$$P_r^a(t_0) = [L^T\mathbf{H}_0^T R_0^{-1}\mathbf{H}_0 L]^{-1}, \tag{11}$$

$\bar{X}$ is the average of a sequence of state vectors and $\mathbf{H}_0$ is the gradient of $H_0$ evaluated at $\bar{X}$. The initial analysis error covariance matrix may be taken as

$$P_a(t_0) = LP_r^a(t_0)L^T.$$

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

**1- Forecast stage:** One applies the model (1) to compute the forecast state

$$X^f(t_k) = M(t_k, t_{k-1})X^a(t_{k-1}) \tag{12}$$

and let the covariance matrix of the forecast error in the reduced space $P_r^f(t_k)$ (of dimension $r \times r$) to evolve as

$$P_r^f(t_k) = [L^T\mathbf{M}(t_k, t_{k-1})L]P_r^a(t_{k-1})[L^T\mathbf{M}(t_k, t_{k-1})L]^T + L^TQ_kL. \tag{13}$$

The forecast error covariance matrix is then equal to

$$P^f(t_k) = LP_r^f(t_k)L^T. \tag{14}$$

**2- Correction stage:** The correction of the forecast state is only done in the reduced space, which therefore can be called the correction basis, according to the formula

$$X^a(t_k) = X^f(t_k) + G_k[Y_k^o - H_kX^f(t_k)] \tag{15}$$

where $G_k$ is now given by

$$G_k = LP_r^a(t_k)L^T\mathbf{H}_k^TR_k^{-1}. \tag{16}$$

The covariance matrix of the analysis error covariance matrix $P_r^a(t_k)$ is then updated from the equation

$$[P_r^a(t_k)]^{-1} = [P_r^f(t_k)]^{-1} + L^T\mathbf{H}_k^TR_k^{-1}\mathbf{H}_kL. \tag{17}$$

The analysis error covariance matrix is then given by

$$P^a(t_k) = LP_r^a(t_k)L^T. \tag{18}$$

It is important to notice that we have neglected in formula (9) the representativeness error, representing the information which has not been explained by the reduced space defined by $L$ (this error may be conveniently "inserted" into the model error [4]). To compensate (at least partially) this neglect, we adopt a simple approach which uses of a forgetting factor as in the SEEK filter (see *Pham et al.* [25]) to limit the propagation of this error in time. This approach has the advantage of avoiding the difficulty of specifying correctly the representativeness error. Moreover, it does not incur any extra computational cost: the algorithm of the ROEK filter remains

unchanged. Only the update equation of the analysis error covariance matrix in the reduced space is replaced by

$$[P_r^a(t_k)]^{-1} = \rho[P_r^f(t_k)]^{-1} + L^T\mathbf{H}_k^T R_k^{-1}\mathbf{H}_k L. \tag{19}$$

Concerning the cost of this filter, it is mostly due to the equation (13) describing the evolution of the forecast error in the reduced space. It thus depends on the dimension of the reduced space $r$ because the numerical calculation of $\mathbf{M}(t_k, t_{k-1})L$ requires $(r+1)$ integrations of the tangent linear model.

The performance of the ROEK filter highly depends on the representativity of the reduction operator $L$. A good choice of $L$ should lead to a large reduction of the dimension of the system and to a reduced state space which must well represents the variability of the model. Different forms of the operator $L$ have been proposed in practice, as the use of a coarse resolution [9], the most dominant singular modes of the tangent linear model and the most dominant eigenmodes of the analysis error covariance matrix [5], etc ..., they rely on more or less on simplifying the dynamic and the characteristics of the model (see [7] for a review).

With the same aim in view, *Cane et al.* [4] have adopted different approach: using the empirical orthogonal functions (EOF) analysis, they reduced the state space to a linear subspace spanned by a small set of basis functions, called EOFs, which nonetheless represented all of the significant structures that were predicted by the model. More than order reduction of the filter for cost reasons, the philosophy of *Cane et al.* [4] relies on the fact that one can not exactly compute the "true" error covariance matrix and hence it may not useful to work with full dynamic of the model. In numerical applications, this procedure was shown to lead to a substantial saving without any loss of accuracy compared to the full EK filter [4].

### 3.2 The SEEK filter

The reduction operator $L$ is usually considered to be constant in time, or to follow empirical laws [7]. As a consequence, the error covariance dynamics may not be properly taken into account [7]. To overcome this problem, *Pham et al.* [25] have proposed a variant of the Kalman filter, called SEEK filter, in which they allowed to the reduction operator $L$ to evolve in time, and then will be noted $L_k$, in order to better capture the dynamics of error statistics.

The SEEK filter is aimed to reduce the prohibitive cost of the EK filter in meteorology and oceanography data assimilation, due to the huge number ($n$) of the state variables. The main idea is to view the error covariance matrix as singular with a low rank $r << n$. This leads to a filter in which the errors correction is only made along certain directions parallel to a linear subspace of dimension $r$. These "correction directions" are those for which error is not sufficiently attenuated by the system dynamic.

This filter proceeds in two stages a part from an initialization stage which is identical to the one of the ROEK filter. For convenience, the reduction operator obtained from the EOF analysis will be noted here by $L_0$.

**1- Forecast 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 $L_{k-1}P_r(t_{k-1})L_{k-1}^T$ where $L_{k-1}$ and $P_r(t_{k-1})$ are of dimension $n \times r$ and $r \times r$ respectively, are available. The model (1) is used to forecast the state as

$$X^f(t_k) = M(t_k, t_{k-1})X^a(t_{k-1}). \tag{20}$$

The forecast error covariance matrix is then given by

$$P^f(t_k) = L_k P_r(t_{k-1})L_k^T + Q_k \tag{21}$$

where the reduction operator $L_k$ evolves with the dynamic of the model according to

$$L_k = \mathbf{M}(t_k, t_{k-1})L_{k-1}. \tag{22}$$

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

$$X^a(t_k) = X^f(t_k) + G_k[Y_k^o - H_k X^f(t_k)], \tag{23}$$

where $G_k$ is now given by

$$G_k = L_k P_r(t_k) L_k^T \mathbf{H}_k^T R_k^{-1}, \tag{24}$$

and, when a forgetting factor $\rho$ is used, $P_r(t_k)$ is computed from

$$[P_r(t_k)]^{-1} = \left[P_r(t_{k-1})/\rho + P_{L_k}^T Q_k P_{L_k}\right]^{-1} + L_k^T \mathbf{H}_k^T R_k^{-1} \mathbf{H}_k L_k \tag{25}$$

with $P_{L_k}$ representing the projection operator onto the reduced state space generated by $L_k$, i.e. $P_{L_k} = [L_k^T L_k]^{-1} L_k^T$. The corresponding filter error covariance matrix is then taken to be

$$P^a(t_k) = L_k P_r(t_k) L_k^T. \tag{26}$$

Since equations (21) and (26) are only included to interpret the results, the SEEK filter drastically reduces the computation cost with respect to the EK filter. Basically it requires about $r+1$ times the cost of the numerical integration of the model, which is the cost of the evolution of its correction basis $L_k$ by the equation (22). The cost of the SEEK filter is then very close to the one of the ROEK filter despite the fact that the correction basis of this former remains fixed in time.

The second remark concerns the evolution of the forecast error covariance matrix in the reduced space $P_r^f(t_k)$. It remains fixed equal to the analysis error covariance matrix in the reduced space at the previous time $P_r^a(t_{k-1})$ in the SEEK filter, while it evolves with the dynamic of the reduced state space in the ROEK filter according to the equation (13), as a compensation, in a way, for the non-evolutivity of its correction basis.

Recently, *Hoteit et al.* [15] have introduced the global-local (or mixed) EOF analysis (see section 5) as a means to construct a reduced space which represents the global and the local variability of the ocean. However, the local basis functions obtained from this analysis, called local EOFs, can not be let to evolve as in the SEEK filter without loosing its locality property. Therefore, they proposed to let only the global basis functions to evolve as in the SEEK filter while leaving the local basis functions fixed. One can then ask if the use of the ROEK filter can compensate the non-evolutivity of the local EOFs.

Since it is very difficult to find theoretical answers to these observations, we will content ourselves with implementing and comparing (in section 6) the performances of these two filters in a realistic setting of the OPA model in the tropical Pacific ocean and then trying to draw some conclusions. But prior to this, we first discuss the cost reduction of the ROEK filter and briefly recall the method of (mixed) EOF analysis.

## 4 Linearization of the reduced dynamics

As we have already said, despite the fact that the correction basis of the ROEK filter remains fixed in time, the implementation cost of this filter is as expensive as the SEEK filter. To reduce the cost of the SEEK filter, *Hoteit et al.* [17] have constructed different variants of the SEEK filter in which they simplified the evolution of the reduction operator, which is the most expensive part of the SEEK filter. To reduce the cost of the ROEK filter, we make use here of a first order stochastic autoregressive model, denoted by $AR(1)$, which describes the evolution of the reduced order state vector $X_r^t(t_k)$. This way, the computation of the gradient of the model transition operator needed for the evolution of $P_r^f(t_k)$ as described by the equation (13), is avoided.

Assume that one has constructed an $AR(1)$ model in its general form, namely (see [3])

$$X_r^t(t_k) = AX_r^t(t_{k-1}) + \mu + e_r(t_k) \tag{27}$$

where $A$ is a $r \times r$ matrix, $\mu = (I_d - A)\bar{X}_r^t$ with $\bar{X}_r^t$ the mean of $X_r^t$ and $e_r$ is a noise, which has mean zero and covariance matrix $Q_r$ and is independent of $X_r^t(t_l)$ for $l < k$. We shall make the approximation

$$e^a(t_k) = X^t(t_k) - X^a(t_k) \approx L[X_r^t(t_k) - X_r^a(t_k)] = Le_r^a(t_k) \tag{28}$$

where the subscript $_r$ refers to vectors in the reduced state space. The evolution of the analysis error in the reduced state can then be computed from the equation

$$e_r^f(t_k) = X_r^t(t_k) - X_r^f(t_k) = Ae_r^a(t_{k-1}) + e_r(t_k), \tag{29}$$

which finally gives

$$P_r^f(t_k) = AP_r^a(t_{k-1})A^T + Q_r. \tag{30}$$

Therefore, to reduce the cost of the ROEK filter, one has only to replace the equation (13) by the equation (30) in the algorithm of this filter. The resulting new filter, called ROAEK filter, can be much less costly then the ROEK filter (about $r$ times faster), but its performance particularly depends on the representativity of the reduced dynamic by the $AR(1)$ model.

- *Construction of an $AR(1)$ model*

To construct the $AR(1)$ model, one starts with a set of state vector $X^t(t_1), \ldots, X^t(t_N)$. The projection of these vectors onto the subspace spanned by $L$ provides a set of state vectors in the reduced space $X_r^t(t_1), \ldots, X_r^t(t_N)$. All one then needs to do is to compute the mean square estimators of $A, \mu$ and $Q_r$ from the sample $X_r^t(t_1), \ldots, X_r^t(t_N)$ according to the formulas

$$\tilde{A} = R_N^{-1}(1,1)R_N(0,1). \quad (31)$$

$$\tilde{\mu} = (I_d - \tilde{A})\bar{X}_r^T. \quad (32)$$

$$\tilde{Q}_r = R_N(0,0) - R_N(0,1)\tilde{A}^T. \quad (33)$$

where

$$R_N(i,j) = \frac{1}{N-1}\sum_{k=2}^{N}[X_r^t(t_{k-i}) - \bar{X}_r^t][X_r^t(t_{k-j}) - \bar{X}_r^t]^T. \quad (34)$$

The reader is referred to [3] for more details about these formula.

# 5 Construction of the reduced basis

We briefly present in this section the construction of the reduced basis via the EOF analysis. Further details can be found in *Hoteit et al.* [17].

## 5.1 (Global) EOF analysis

The "classical" (or global) EOF analysis provides the best representation of a set of state vectors $X_1, \ldots, X_N$ in $\mathbb{R}^n$ in a low-dimension (denoted by $r$) linear subspace. It consists in "compressing data" contained in these states by summarizing the correlation of their variables in a few vectors only, called EOFs. These vectors are no other than the first $r$ normalized eigenvectors of the sample covariance matrix $P$ of $X_1, \ldots, X_N$ relative to some metric $\mathcal{M}$ in the state space[1], the eigenvectors being ranked in decreasing order of their eigenvalues $\lambda_1, \ldots, \lambda_r$. With regard to the choice of $r$, one can consider the fraction of variance (or inertia) $\sum_{k=1}^{r}\lambda_k / \sum_{k=1}^{p}\lambda_k$ explained by the first $r$ EOFs, which should be close to 1.

[1]The introduction of $\mathcal{M}$ is needed in the case where the state variables are not homogeneous (as they represent different physical variable such as velocity, salinity, etc.) to obtain a distance between state vectors independent from unit of measure.

### 5.2 Local EOF analysis

The EOFs 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, *Hoteit et al.* [15] have developed a so called local EOF analysis. The basic idea consists in constructing a set of EOFs each having support a small region of the ocean. This can simply be done by independently applying the EOF analysis on different ocean subdomains. However, since the combined representation would provide no correlation between spatial points in different subdomains (if they are disjoint), these subdomains should overlap to obtain adequate representativity. To do this properly, one may consider a *partition of unity* of the ocean domain, i.e. a set 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 decomposed as

$$X(x,y,z) = \sum_{j=1}^{J} X^{(j)}(x,y,z) \quad \text{where} \quad X^{(j)}(x,y,z) = X(x,y,z)\chi^{(j)}(x,y,z) \qquad (35)$$

and $x, y$ and $z$ denote the spatial coordinates. Next, for each $j$ between 1 and $J$, one conducts separately classical EOF analysis on each local field $X^{(j)}$ to obtain a representation subspace for each ocean subdomain. It can be then easily seen that one can still reduce the "global" representation error of the original state $X$ by considering its projection onto the subspace spanned by the columns of the matrix whose columns are the (local) EOFs of all the subdomains.

### 5.3 Global-Local (Mixed) EOF analysis

The main difficulty with the local EOFs basis is that one can not make it to evolve as in the SEEK filter (see section 3.2) without loosing its locality property. *Hoteit et al.* [15] have also noticed, in their numerical experiments, that the long range variability was not well represented by such a basis. To overcome these problems, *Hoteit et al.* [15] have augmented the local EOFs with some global EOFs. The advantage of such approach is double: firstly, the resulting mixed global-local basis can represent the long range as well as the short range oceanic phenomena, and secondly, the global EOFs can be made to evolve as in the SEEK filter.

To construct the above type of basis, *Hoteit et al.* [15] have applied a local EOF analysis[2] on the residuals of the state vectors in the first global EOFs in order to

[2] not centered since these vectors represent a residual and hence should be already centered.

extract the local variability, since the representation error (or residual) of a state vector is almost due to the bad representativity of the short range variability in the first (global) EOFs.

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

To evaluate the performance of the filters described above, 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.

## 6.1 Model description

### 6.1.1 The OPA model

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 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" (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* [2]). On the lateral, diffusive and viscous operators can be either of second or of fourth order. The reader is referred to the OPA reference manual *Madec et al.* [22] for more details.

### 6.1.2 Configuration of the tropical Pacific

The model domain covers the entire tropical Pacific basin extending from $120°E$ to $70°W$ 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 $171 \times 59$ on 25 vertical levels. The model equations are

solved on an isotropic 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 [21]. The forcing fields are interpolated from the ECMWF 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). The salinity and the temperature are stem from seasonal climatologic Levitus data [21].

### 6.1.3 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

$$X^t = ( U , V , S , T )^T. \tag{36}$$

However, the observed variable, which is the sea surface height $SSH$, is a diagnostic variable computed from the barotropic velocity by a complex nonlinear algebraic equation. Thus, the observation operator $H$ which relates this state vector to the observed variables will be nonlinear and very difficult to compute. 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

$$X^t = ( U , V , S , T , SSH )^T. \tag{37}$$

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 variables is now $4 \times 171 \times 59 \times 25 + 171 \times 59 = 1\,018\,989$ instead of $4 \times 171 \times 59 \times 25 = 1\,008\,900$.

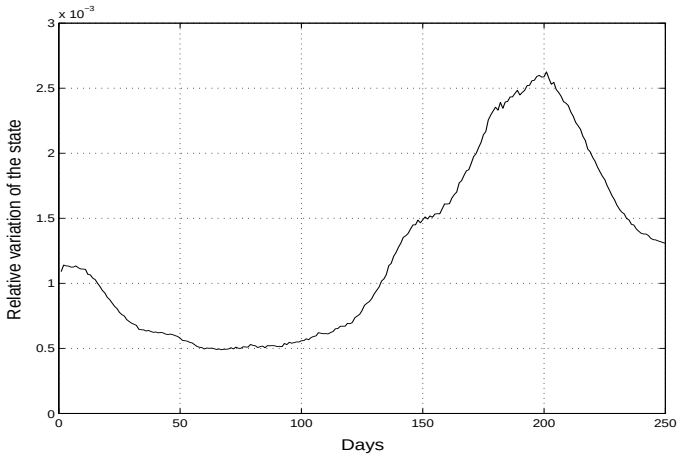


Figure 1: Relative variation of the state vector in the assimilation period.

## 6.2 Data and filters validation

Twin experiments are used to assess the performances and the capabilities of our filters. Therefore, a reference experiment is performed and the reference state $X^t$ retained to be later 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 March $1^{st}$ 1991 to November $10^{st}$ 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 occurs between July and September (see Figure 1).

The performance of the 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

$$RRMS(t_k) = \frac{\|X^t(t_k) - X^a(t_k)\|}{\|X^t(t_k) - \bar{X}\|}, \tag{38}$$

where $\bar{X}$ is the mean state of the sample $H_S$ and $\|\cdot\|$ denotes the Euclidien norm. Thus 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.

### 6.3 Results of the different EOF analysis

In the present study, the data for the assimilation experiments is again simulated but in an unrelated way with the above simulation. 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 (37) are not of the same nature, we shall in fact apply a multivariate EOF analysis (local and global). We define a metric $\mathcal{M}$ in the state space to make the distance between state vectors independent from unit of measure. We choose $\mathcal{M}$ as the diagonal matrix with diagonal elements being the inverse of the variances of each state variables, namely $U$, $V$, $S$, $T$ and $SSH$, average over the grid points.

- *Global EOF analysis*

Figure 2 plots the number of (global) EOFs and the percentage of inertia contained in the sample $H_S$ they explain. From this result, we have chosen to retain $r = 30$ global EOFs in our assimilation experiments, as this achieve 85% of the inertia of the sample and this percentage is not much increased for higher value of $r$.

- *Mixed EOF analysis*

To compute the global-local basis, we have applied the local EOF analysis on the residues of the ocean states in the subspace spanned by the first 5 global EOFs which explained almost 50% of the total global inertia. Concerning the local analysis, the partition of unity functions was considered in tensorial form (i.e. $\chi^{(j)}(x,y,z) = \chi_X^{(j_1)}(x)\chi_Y^{(j_2)}(y)$). Note that we do not limit the correlation length in the vertical direction to let the surface informations propagate to the ocean bottom. The unity functions are plotted in Figure 3. The single one-dimensional tensorial function in the meridional $\chi_Y^{(1)}(y)$ direction was then taken constant equal to 1. The tropical

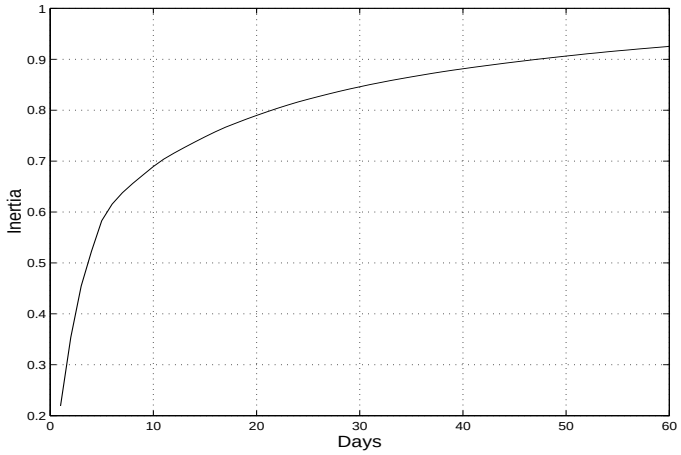


Figure 2: Percentage of inertia versus the number of retained EOFs.

Pacific domain has then been subdivided into three zonal subdomains to limit the spatial correlation length of the ocean variables in this direction. After applying separately (classical) EOF analysis on the residuals of the state vectors, we have retained 6, 6 and 7 EOFs in order to explain 40% of the inertia of the residues in the first, second and the third subdomain respectively. The dimension of reduced state space spanned by the mixed EOFs is thus equal to $5 + 6 + 6 + 7 = 24$.

### 6.4 Assimilation results

We first present the results of the SEEK filter and compare them to those of the ROEK filter. Next, we test the ROEK filter with the mixed EOFs to see if the the evolution the forecast error in the reduced state space compensates the "non-evolutivity" of the local EOFs. Finally, we discuss the assimilation results of the ROAEK filter and we study its performance with respect to the ROEK filter.

- *The SEEK filter*

The SEEK filter has been implemented with a fixed forgetting factor set to $\rho = 0.8$. It can be seen from Figure 4 that the SEEK filter performs well both in the upper layers 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

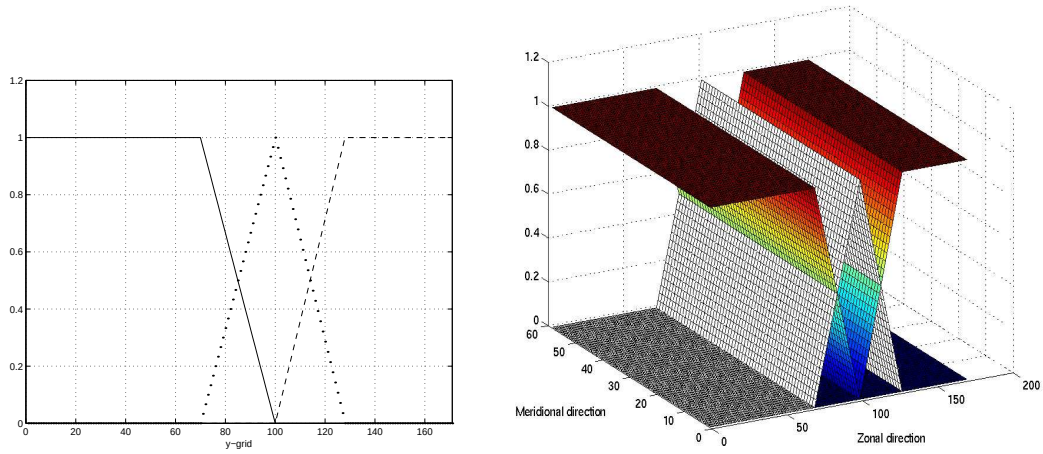


Figure 3: One-dimensional tensorial functions $(\chi_X^{(j)}(x), j = 1, ..., 3)$ and two-dimensional partition of unity functions $(\chi^{(j)}(x, y) = \chi_X^{(j)}(x), j = 1, ..., 3)$.

this period. One may think that the meridional velocity $V$ is not sufficiently well-assimilated because the assimilation error is only reduced by less than a 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 march $1^{st}$ 1991, 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 filter computed in all the layers are quite similar to that computed on each layer. Therefore, in the sequel we will only present results in all layers, to save space.

• *Comparison between the ROEK and SEEK filters*

We have conducted an experiment with the ROEK filter in the same setup as before to show the usefulness of the evolution of the reduced state space with the dynamic of the model carried out in the SEEK filter over the evolution of the forecast error in the reduced state space. The assimilation results of this experiment, plotted in

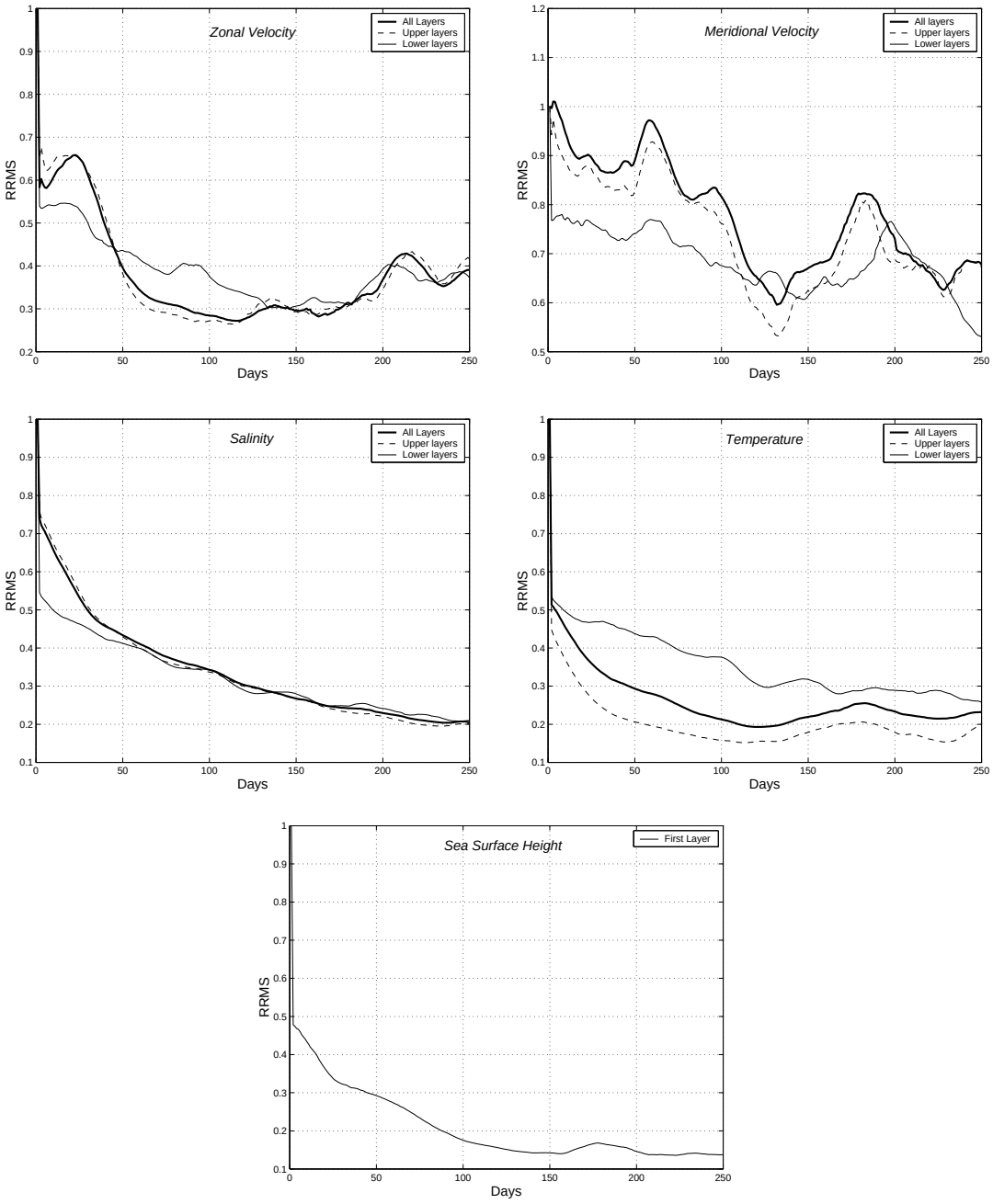


Figure 4: Evolution in time of the $RRMS$ for the SEEK filter on the whole model domain, on the (mean of the 5) upper and the (mean of the 5) lower layers.

Figure 5 and Figure 6-7, show that, despite a good behavior during the stable period, the ROEK filter is less stable than the SEEK filter during model instabilities. This may be explained by the evolution of the correction basis of the SEEK filter to follow the dynamic of the model. The good performance of the ROEK filter in the stable period enable us to presume that this filter can be as effective as the SEEK filter if its reduced state space space represents well the variability (of all the modes) of the model. On the other hand, the performance of this filter have been seriously degraded during the unstable period, which may be attributed to the fact that its correction basis, obtained from the EOF analysis, does not sufficiently represents the variability of the model. The evolution of the correction basis thus seems to be necessary during the model unstable periods, to follow the new modes of the model.

- *The ROEK and SEEK filters with the mixed EOFs*

As we have said before, one can not let the local part of the mixed EOFs to evolve as in the SEEK filter without loosing its locality property. Therefore, in a first experiment, we let this part fixed in time while letting the global part (which contains 5 EOFs) to evolve with the evolution equation of the SEEK filter to follow the dynamic of the model. We have also implemented the ROEK filter with the mixed EOFs to compare its performance to those obtained with the above (semi-evolutive) SEEK filter. Indeed, our aim is to see if the evolution of the forecast error in the reduced state space compensate the non-evolutivity of the local EOFs. Figure 8 shows the assimilation results of these experiments. One can see that the SEEK filter performs better than the ROEK filter (also it is 6 times faster since only 6 vectors evolve with time), particularly in the unstable period of the model. This suggests that the evolution, even partially, of the correction basis is beneficial during this period. However, on can notice that the use of the mixed EOFs appreciably enhances the performances of the ROEK filter in the unstable period with regard to the classical (global) EOFs.

- *The ROAEK filter*

Finally, we have tested the performance of the ROAEK filter and we have compared its results to those obtained with the ROEK filter. For this experiment, we have used the classical EOFs of dimension 30. Note that we have not used a forgetting factor in this filter because the error covariance matrix $Q_r$ of the $AR(1)$ model play

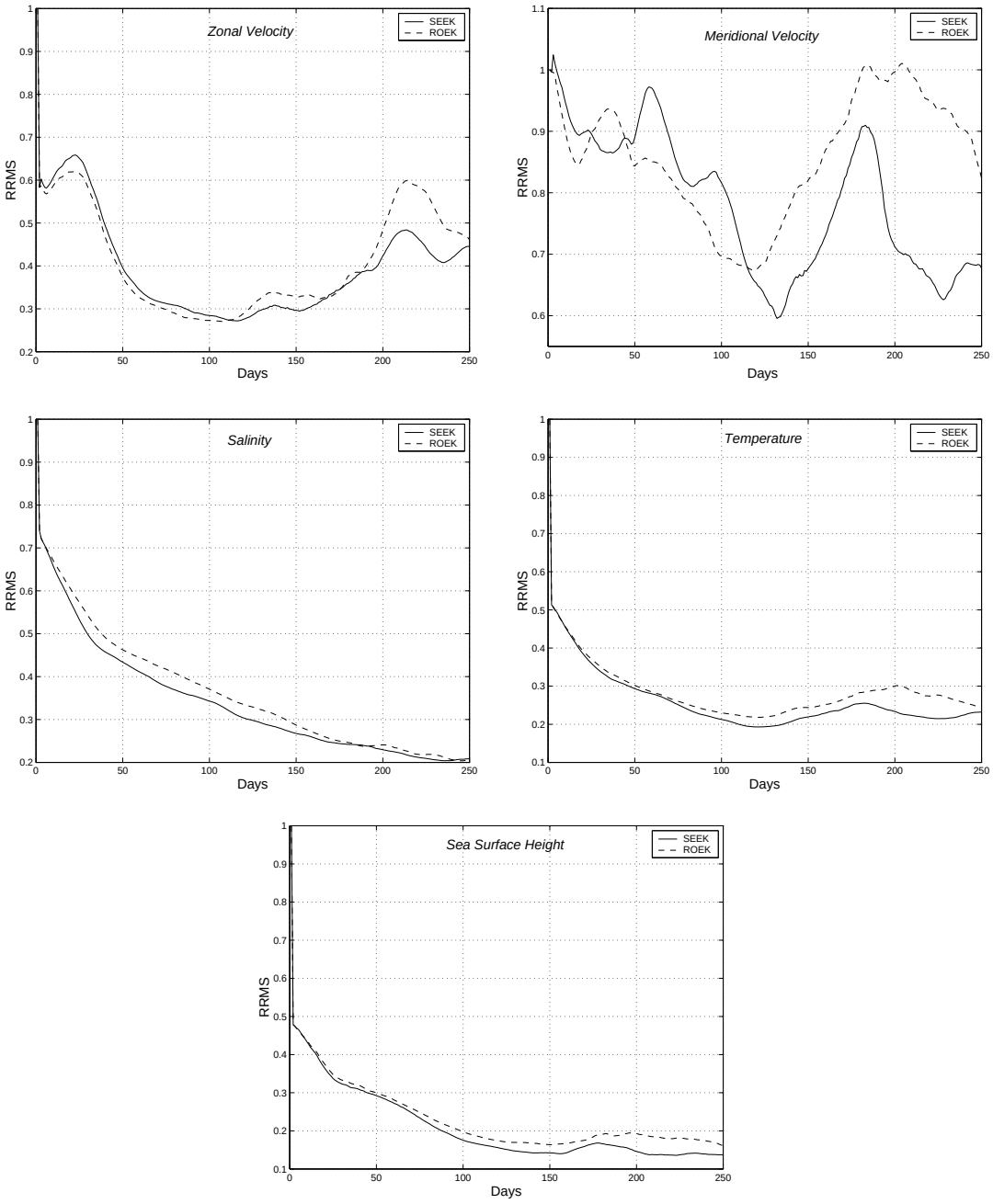


Figure 5: Evolution in time of the $RRMS$ for the SEEK and ROEK filters.

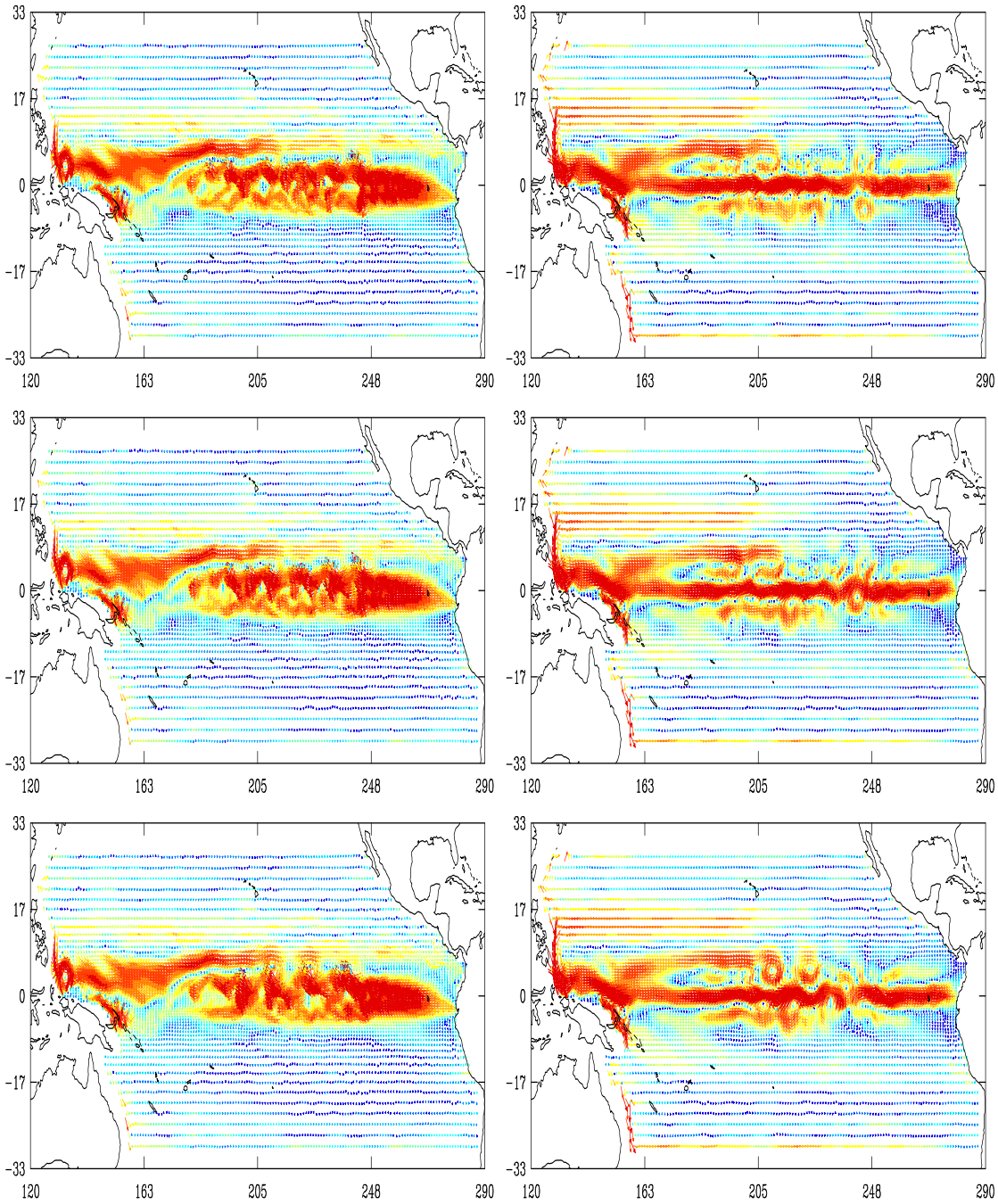


Figure 6: Maps of ocean velocity on Sept $1^{st}$ 90 in the uppermost (left) and the $17^{th}$ (right) layers: from the SEEK filter (top); from the reference (middle); and from the ROEK filter (bottom).

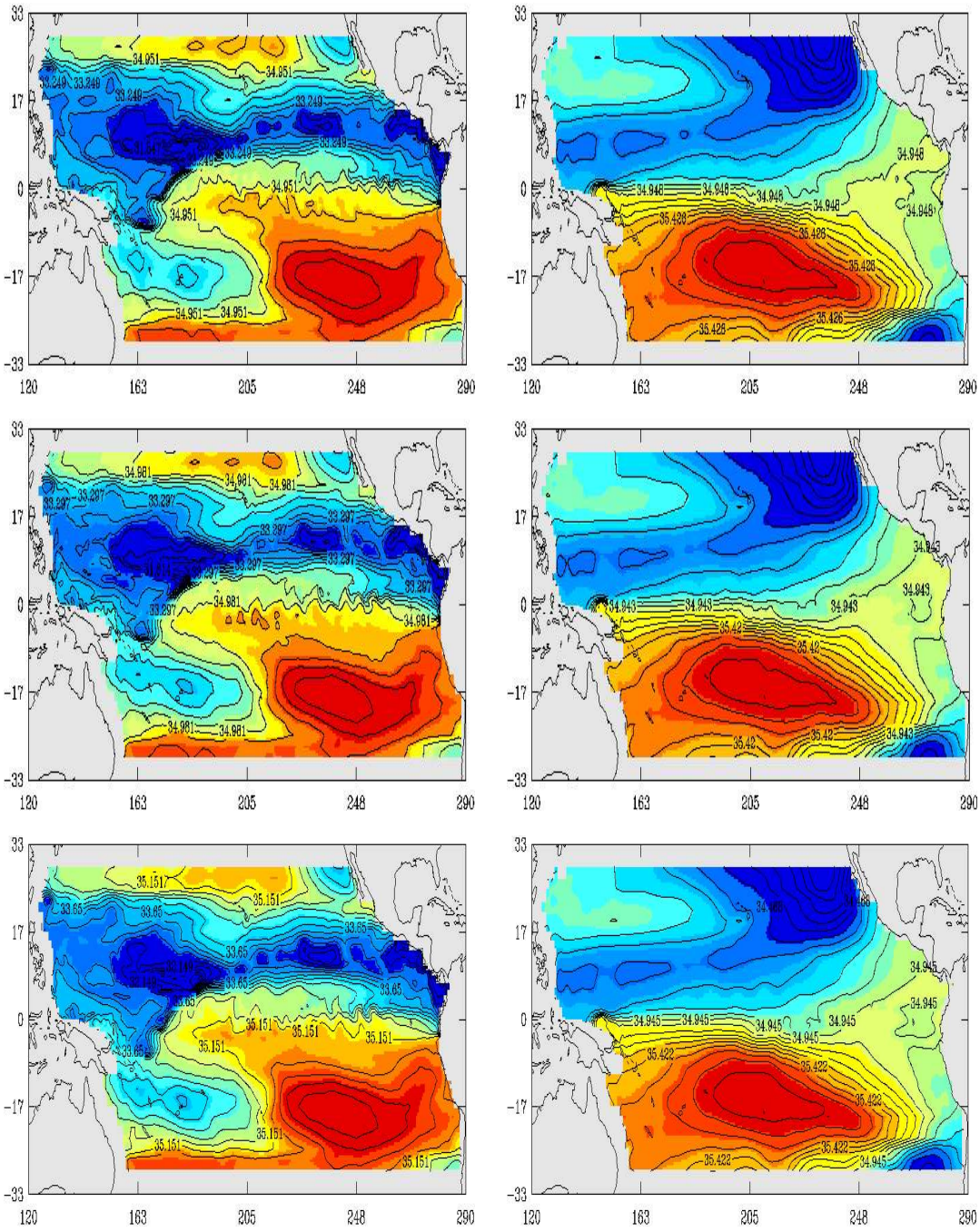


Figure 7: Maps of sea salinity on Sept $1^{st}$ 90 in the uppermost (left) and the $17^{th}$ (right) layers: from the SEEK filter (top); from the reference (middle); and from the ROEK filter (bottom).

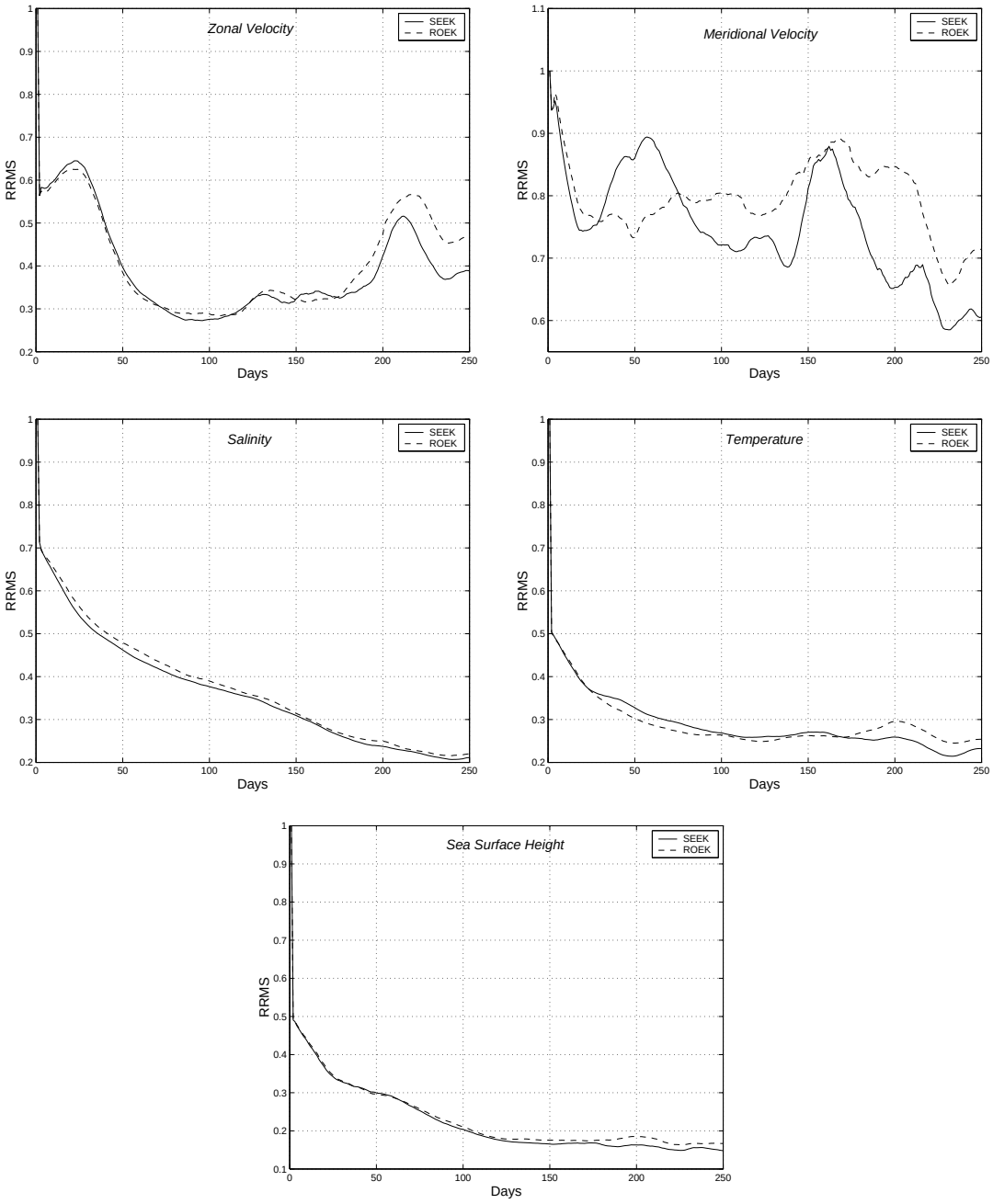


Figure 8: Evolution in time of the $RRMS$ for the ROEK and the SEEK filters with the mixed EOFs.

a similar role to this factor.

To construct the $AR(1)$ model, we have first projected the state vectors of the sample $H_S$ on the subspace generated by the EOFs. Next, we have used the sample consisting of the projected state vectors to construct the model according to the formulas described in the section 4.

FIG. 9 plots the results of this experiment and compare them to those obtained with the ROEK filter. With regard to its cost (30 times faster than the ROEK filter), one can say that the ROAEK filter behaves fairly well. However, its performances was not sufficiently good with respect to the ROEK filter. After looking more closely into the results, we have noticed that the bad performance of the ROAEK filter is due to the bad representativity of the $AR(1)$ model, all the more as so as the coefficients of the error covariance matrix $Q_r$ of this model have been very large.

# 7 Discussion

The extended Kalman (EK) filter is one of the major tool to assimilate data into ocean models. However, its implementation in realistic ocean models is not possible because of its prohibitive cost. As proposed by *Fukumori et al.* [10], the use of a reduction operator, which relates the state vector of the system to a small dimension reduced-order state vector, offers us several alternatives to reduce the cost of the EK filter. In this paper, we have been interested by the comparison between the ROEK filter proposed by *Cane et al.* [4] and the SEEK filter proposed by *Pham et al.* [25]. Our principal aim was to show the usefulness of the evolution of the reduced-order state space carried out in the SEEK filter. The results of the twin experiments which we have conducted in a realistic setting of the OPA model in the tropical Pacific ocean seems to confirm this. More precisely, we have noticed that the evolution of the reduced state space is beneficial during the mode instabilities, where the reduced space spanned by the EOFs may not well represents the variability of the model.

In the other hand, *Hoteit et al.* [15] have introduced the notion of global-local (or mixed) EOF analysis to construct a set of EOFs functions which represents the global variability of the model as well as the local variability. However, the local part EOFs can not be let to evolve with dynamic of the model as in the SEEK filter without loosing its locality property. Here, we have implemented the ROEK filter

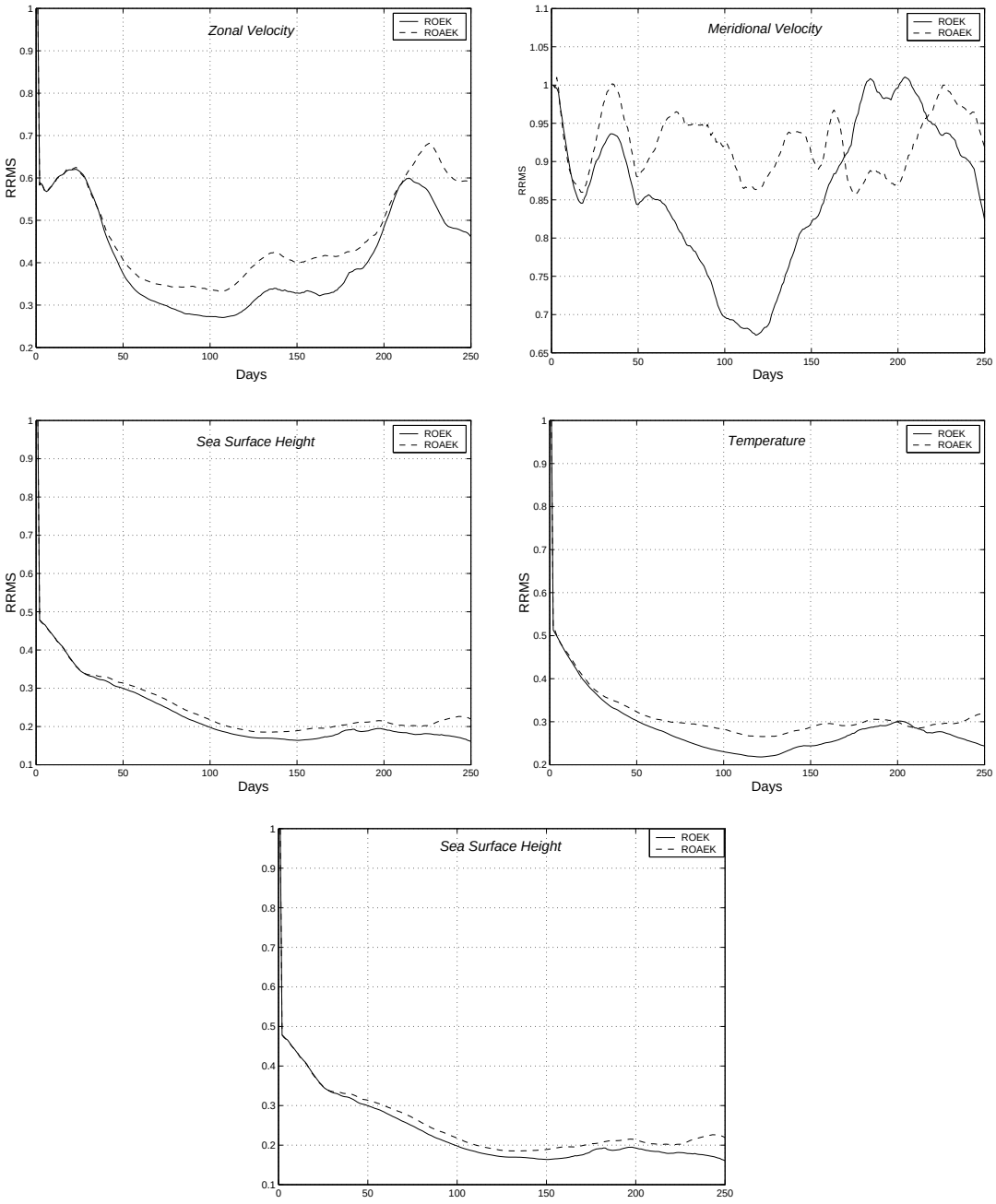


Figure 9: Evolution in time of the $RRMS$ for the ROEK and ROAEK filters.

with these EOFs to see if the evolution of the forecast error in the reduced state space compensate the non-evolutive character of the local EOFs. Therefore, two experiments have been conducted with the ROEK filter, using the mixed EOFs and with the SEEK filter in which only the global elements of the mixed EOFs was let to evolve while the local elements was kept fixed. The results of our experiments clearly show that the (semi-evolutive) SEEK filter performed better than the ROEK filter despite the fact that only the global elements of the mixed EOFs evolve with the dynamic of the model (which moreover means that this this (semi-evolutive) SEEK filter was much less costly than the ROEK filter).

Finally, a new filter, called ROAEK, derived from the ROEK filter has been developed. Our motivation was essentially to reduce the cost of the ROEK filter which remains expensive for an operational oceanography. Our approach was to use a first order autoregressive model $AR(1)$ to let the forecast error to evolve in the reduced state space, which is the most expensive part of the ROEK filter. The assimilation results of our experiments obtained by this filter enable us to presume that this filter can be as effective as the ROEK filter if the $AR(1)$ represents sufficiently well the variability of the reduced-order state space.

In twin experiments, the ROEK and the SEEK filters was found to be fairly effective in assimilating of surface-only pseudo-altimeter data. Further works will consider a more realistic situations, like the addition of the model error or the use of a more realistic observations (according to satellite tracks and real data from satellite). However, these preliminary twin experiments applications were a necessary steps before realistic applications and provide us with encouraging 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] 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.

[3] Brillinger D.R. (1981): Time series, data analysis and theory. New-York: *Rinehart & winston.*

[4] Cane M.A., A. Kaplan, R.N. Miller, B. Tang, E.C. Hackert and A.J. Busalacchi (1996): Mapping tropical Pacific sea level: data assimilation via a reduced state Kalman filter. *J. Geophys. Res.*, **vol.101**, no.C10, 599-617.

[5] Cohn S.E. and R. Tolding (1996): Approximate Kalman filters for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63-75.

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

[7] De Mey P. (1997): *Data Assimilation at the Oceanic Mesoscale: A Review.* J. Met. Soc. Japan, 75(1B), 415-427.

[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. (1995): Assimilation of Topex sea level measurements with a reduced-gravity shallow water model of the tropical Pacific ocean. *J. Geophys. Res.*, **100**(C12), 25027-25039.

[10] 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.

[11] 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.

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

[13] Gourdeau L., S. Arnault, Y. Menard and J. Merle (1992): Geosat sea level assimilation in a tropical Atlantic model using Kalman filter. *Oceanologica Acta*, **15**, 567-574.

[14] Hoang H.S., P. De Mey, O. Tallagrand and R. Baraille (1997): Adaptive filtering: Application to satellite data assimilation in oceanography. *J. Dynamicas of Atmospheres and Oceans*, **27**, 257-281.

[15] Hoteit I., D.T. Pham and J. Blum (2000): A semi-evolutive filter with partially local correction basis for data assimilation in oceanography. *Technical report*, **RR-3975**, INRIA. Submitted to *Mon. Wea. Rev.* Available on http://www.inria.fr/RRRT/RR-3975.

[16] Hoteit I., D.T. Pham and J. Blum (2000): A semi-evolutive partially local filter for data assimilation. To appear in *Marine Pollution Bulletin.*

[17] Hoteit I., D.T. Pham and J. Blum (2000): Efficient reduced Kalman filtering and application to altimetric data assimilation in tropical pacific. *Technical report*, **RR-3937**, INRIA. Submitted to *J. Mar. Sys.*. Available on http://www.inria.fr/RRRT/RR-3937.

[18] 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.

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

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

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

[22] 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.

[23] Miller R.N. and M. Cane (1989): A Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.*, **19**, 773-790.

[24] 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.

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

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

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