A SINGULAR EVOLUTIVE EXTENDED KALMAN FILTER FOR DATA ASSIMILATION IN OCEANOGRAPHY

Dinh Tuan Pham[*], *Jacques Verron*[**] *and Marie Christine Roubaud*[*]

[*]Laboratoire de Modélisation et Calcul, URA 397 CNRS, UJF, B.P. 53X, 38041 Grenoble Cédex, France

[**] Laboratoire des Écoulements Géophysiques et Industriels, UMR 5519 CNRS, UJF, INPG, BP 53X, 38041 Grenoble Cédex, France

(This is the preprint a paper submitted to J. of Marine Systems. It has also appear as "Rapport Technique RT 162 LMC/IMAG (September. 1996)

Abstract

In this work, we propose a modified form of the extended Kalman filter for assimilating oceanic data into numerical models. Its development consists essentially in approximating the error covariance matrix by a singular low rank matrix, which amounts in practice to making no correction in those directions for which the error is attenuated by the system. This not only reduce the implementation cost to an acceptable level but may also improve the filter stability as well. The "directions of correction" of the filter evolve with time according to the model evolution, which is the most original feature of this filter, distinguishing it from other sequential assimilation methods based on the projection onto a fixed basis of functions. A method for initializing the filter based on the empirical orthogonal functions is also described. An example of assimilation based on the quasigeostrophic model for a square ocean domain with a certain wind stress forcing pattern, is given. Although this is only a simple test case designed to assess the feasibility of the method, the results are very encouraging.

1. Introduction

This work was motivated by the problem of data assimilation in oceanography (or in geophysical sciences generally). In general, data assimilation is seen as the means of obtaining a consistent picture of a geophysical system optimally blending all the information available on this system. This information may consist of data of all types, of varying accuracy and geographical distribution, and also of information derived from the theoretical knowledge already possessed on the system expressed in terms of physical laws, either deterministic or statistical, and therefore via various forms of models.

Several methods exist for assimilating observations into numerical models, most of them originally developed in meteorology. Two main approaches are usually to be seen in the assimilation techniques, following either the optimal control theory or the statistical

estimation theory [see for example the review by *Ghil and Malanotte-Rizzoli*, 1991]. In the present study, we are interested only in the second approach for which the Kalman filtering theory is the primary framework. But the application of Kalman filtering encounters enormous difficulties due to the huge dimension of the state vector of the system under consideration. A further major difficulty is caused by to the nonlinear nature of the system. Here we shall content ourselves with using the so called extended Kalman filter (EKF). This is a linearized form of the ordinary Kalman filter (KF) adapted to deal with nonlinear systems, since the later was designed only for a linear system. We shall assume that the linearized system is close enough to the original system for the results in the linear case to be transferable to this case.

In this paper, we shall concentrate on the issue of dimension and propose a method to partially overcome this difficulty. To this end, a brief description of extended Kalman filtering together with some discussion of problems arising in its use in geophysical data assimilation will be first given in section 2. Following on from this, an algorithm based on the use of singular low rank error covariance matrix, will be introduced. This filter, called the singular evolutive extended Kalman (SEEK) filter, not only solves the practical problem of reducing the computational cost to an acceptable level, but in addition possesses a certain theoretical foundation. Indeed, as it is pointed out in section 2, it would be utopic to look for an "optimal" filter in such a high dimensional context, so our aim may simply be confined to finding a stable filter which reduces the propagation of error from one step to the next. Our SEEK filter is designed to achieve this. The filter operates in a similar way to certain other recently proposed reduced order Kalman filters (see references in section 3), by applying correction in certain directions only. However, the original feature of this particular filter is that the directions in question depend on the system dynamics, are precisely those for which the system does not attenuate the error. They therefore evolve in time, as the state of the system changes. Our filter has the built-in capability to track such evolution, hence the name "evolutive". In this regard, the method possesses a certain robustness with respect to initialization of possibly poor quality. However, we also provide a method for finding a good initial singular low rank error covariance matrix, based on the EOFs (empirical orthogonal functions) also known as PCs (principal components). The overall method is illustrated by an example in section 4. In this simple academic test case performed with a simple quasigeostrophic model, an assessment of the feasibility of the method only is sought. Real applications are currently being investigated in the North Atlantic and in the Tropical Pacific with more complex models and these will be reported later.

2. The extended Kalman filter (EKF)

For consistency with the existing literature on the subject, the notation proposed by *Ide et al.* [1995] will be adopted as far as possible. Consider a physical system described by

the equation

$$\mathbf{x}^t(t_i) = M(t_{i-1}, t_i)\mathbf{x}^t(t_{i-1}) + \eta(t_i). \tag{2.1}$$

Here $\mathbf{x}^t$ is a vector representing its true state, $M(t_{i-1}, t_i)$ is an operator describing the system transition from time t_{i-1} to t_i. The actual fluid system is supposed to differ from that described by the operator M by an error of $\eta(t_i)$ which is assumed to be a white noise process with zero mean and covariance matrix $\mathbf{Q}(t_i)$. The transition operator $M(s,t)$ is usually obtained from the integration of a certain partial differential system, typically a numerical model. At each t_i, one observes a vector

$$\mathbf{y}_i^o = H_i\mathbf{x}^t(t_i) + \epsilon_i, \tag{2.2}$$

where H_i is the observational operator. The observational error ϵ_i is assumed to be a white noise process with zero mean and covariance matrix $\mathbf{R}_i$, independent of the model error $\eta(t_i)$.

The filtering problem, which is mathematically equivalent to sequential data assimilation, consists of estimating $\mathbf{x}^t(t_i)$ at each time t_{i-1}, given the observations up to this time. Since *Kalman* [1960], this problem has been the source for numerous developments in various fields of science and in particular that of atmospheric and oceanic flow [*Ghil*, 1986; *Ghil and Malanotte-Rizzoli*, 1991]. However, the Kalman filter was designed for a linear system only. For a nonlinear system, one often resorts to linearization of the operators M and H around the currently available state estimate, yielding the so-called extended Kalman filter (EKF) [*Ghil et al.*, 1982]. The application of EKF to nonlinear models has been the subject of various investigations [*Miller et al.*, 1994; *Evensen*, 1993, 1994]. It may however result in instability in the covariance evolution equation [*Evensen*, 1992; *Gauthier at al.*, 1993].

An attractive feature of the EKF is that it is recursive. Computation is done "on line" as soon as new observations are available. Assuming that at a time t_i, one already has a system state estimate, often referred to as the analysis state vector $\mathbf{x}^a(t_{i-1})$, with some analysis error covariance matrix $\mathbf{P}^a(t_{i-1})$, the EKF permits the construction of the next analysis vector $\mathbf{x}^a(t_i)$ together with its error covariance matrix $\mathbf{P}^a(t_i)$.

Classically, this construction is described in two stages

(i) The forecasting stage: the model equation of (2.1) is used to forecast the state at time t_i using the previous analysis vector: $\mathbf{x}^f(t_i) = M(t_i, t_{i-1})\mathbf{x}^a(t_{i-1})$. The forecast error covariance matrix is

$$\mathbf{P}^f(t_i) = \mathbf{M}(t_{i-1}, t_i)\mathbf{P}^a(t_{i-1})\mathbf{M}(t_{i-1}, t_i)^T + \mathbf{Q}(t_i) \tag{2.3}$$

where $\mathbf{M}(t_{i-1}, t_i)$ denotes the gradient of $M(t_i, t_{i-1})$ *evaluated at* $\mathbf{x}^a(t_{i-1})$ and ^{T} denotes the transpose. At this stage we introduce the linearization of (2.1) around $\mathbf{x}^a(t_{i-1})$:

$$\mathbf{x}^t(t_i) \approx M(t_i, t_{i-1})\mathbf{x}^a(t_{i-1}) + \mathbf{M}(t_{i-1}, t_i)[\mathbf{x}^t(t_{i-1}) - \mathbf{x}^a(t_{i-1})] + \eta(t_i). \tag{2.1'}$$

so that the formula (2.3) is only approximate.

(ii) The analysis (or correction) stage: The new observation $\mathbf{y}_i^o$ is used to correct the forecast state vector $\mathbf{x}^f(t_i)$. The correction will be taken to be proportional to the observational residual (also called innovation) vector: the new analysis state vector is of the form $\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i[\mathbf{y}^o(t_i) - H_i\mathbf{x}^f(t_i)]$ where $\mathbf{K}_i$ is a certain matrix called the gain. To determine the optimal gain, one also linearizes the observation equation (2.2) around $\mathbf{x}^f(t_i)$ and it can then be shown that the optimal gain is given by

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i]^{-1} \tag{2.4}$$

where $\mathbf{H}_i$ is the gradient of H_i evaluated at $\mathbf{x}^f(t_i)$. The corresponding analysis error covariance matrix is

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{P}^f(t_i)\mathbf{H}_i^T[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i]^{-1}\mathbf{H}_i\mathbf{P}^f(t_i).$$

For the pure mathematical analysis, the whole algorithm can be described by two equations:

- the Riccati equation which updates the error covariance matrix

$$\begin{aligned}\mathbf{P}_i^a = \mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T + \mathbf{Q}_i - (\mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T + \mathbf{Q}_i)\mathbf{H}_i^T \\ [\mathbf{H}_i(\mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T + \mathbf{Q}_i)\mathbf{H}_i^T + \mathbf{R}_i]^{-1}\mathbf{H}_i(\mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T + \mathbf{Q}_i)\end{aligned} \tag{2.5}$$

putting $\mathbf{P}_i^a = \mathbf{P}^a(t_i)$, $\mathbf{Q}_i = \mathbf{Q}(t_i)$ and $\mathbf{M}_{i-1,i} = \mathbf{M}(t_{i-1}, t_i)$ for short,

- and the filter equation

$$\mathbf{x}^a(t_i) = M(t_{i-1}, t_i)\mathbf{x}^a(t_{i-1}) + \mathbf{K}_i[\mathbf{y}_i^o - H_i\mathbf{x}^a(t_{i-1})] \tag{2.6}$$

where $\mathbf{K}_i$ is given by (2.3) and (2.4). Calculations based on (2.5) show that (if $\mathbf{R}_i$ is invertible) it can be also computed as (if $\mathbf{R}_i$ is invertible)

$$\mathbf{K}_i = \mathbf{P}_i^a\mathbf{H}_i^T\mathbf{R}_i^{-1}. \tag{2.7}$$

The above development is very attractive in theory but it possesses the practical difficulty that everything need to be known: the system transition $M(t_{i-1}, t_i)$, the observation operator H_i, the covariance matrices $\mathbf{R}_i$, $\mathbf{Q}_i$ and the initial error covariance matrix $\mathbf{P}_0^a$. The operator H_i is indeed known and one can reasonably assume $M(t_{i-1}, t_i)$ to be known because it comes from physical laws in spite of possibly containing (physical) parameters and parameterizations tuned or known only imperfectly (as for example the forcing functions or the horizontal subgridscale frictional effects). The observation error covariance matrices $\mathbf{R}_i$

may be harder to assess. One may have good knowledge of the instrumental error variances in situations such as altimetric observations from the satellite Topex/Poseidon over the ocean, for which the error estimates have become fairly solidly established, being based on now extensive verification and validation phases (see for example the two special issues of the Journal of Geophysical Research of Dec. 1994 vol. 99, C12 and Dec. 1995, vol. 100, C12). However, it is not always clear how the correlations of the observation error can be obtained. One may (and usually does) assume that they are zero but the possibility that neighboring observations may be correlated cannot be entirely ruled out. The greatest difficulty lies in obtaining $\mathbf{Q}_i$ and $\mathbf{P}_0$. The covariance matrix of the dynamic noise $\eta(t_i)$ can only be obtained through accurate knowledge of the statistical behavior of the state process $\mathbf{x}^t(t_i)$, but how can this be acquired without actually observing it? As for the initial error covariance matrix $\mathbf{P}_0$, the fact is that very little is usually known concerning the initial state of the system, and so even less about the degree of the error when a given $\mathbf{x}^a(t_0)$ is chosen. Note however that the choice of $\mathbf{P}_0^a$ may be of lesser importance, in view of the stability of the Kalman filter, but this consideration is also related to the issue of dimension as will be discussed later.

From the above discussion, we may conclude that while it is justifiable to assume that M and H are known, there is a real possibility that the matrices $\mathbf{R}_i$ and especially $\mathbf{Q}_i$ are imperfectly specified. (The *Fukumori*'s [1995] discussion with regard to this question of the model and data error estimates is particularly valuable.) The KF is optimal, but only if the parameters $\mathbf{R}_i$ and $\mathbf{Q}_i$ are correct. If any one of them is erroneous, it is no longer so. Thus, any KF is in practice somewhat suboptimal. As far as the EKF is concerned, there is also the crucial additional issue of the linearization error.

For low dimensional systems, one may hopefully be able to choose the matrices $\mathbf{R}_i$ and $\mathbf{Q}_i$ sufficiently close to the true ones for the filter still to be close to optimal. The situation is different for very large systems encountered in meteorology and oceanography. The dimension N of such system is of the order of 10^5 or 10^6 or even higher. Apart from the practical problem that the computational cost of the EKF in such a situation being prohibitive (it requires an $O(N^3)$ number of floating point operations), there is the difficulty that too many filter parameters need to be specified for $\mathbf{R}_i$ and especially $\mathbf{Q}_i$. Even if the $\mathbf{Q}_i$ are the same (constant in time), one still has to specify a matrix with $N(N+1)/2$ independent elements which, for $N = 10^5$, comes out to about five billion. It is questionable whether one can have even a crude knowledge on such a huge number of parameters. This point is discussed at length in [*Cane et al.*, 1995]. Here we note, only briefly, that a statistical estimate based on a sample of length T has a convergence rate of $1/\sqrt{T}$ and in practice T would be only a tiny fraction of N. Therefore if one tries to estimate such a huge number of parameters, there will be inevitably large errors in many. Thus, it is doubtful whether $\mathbf{Q}_i$ can ever be specified with sufficient accuracy.

The same could be said for $\mathbf{P}_0^a$. However, the Kalman filtering contains an interesting

result suggesting in particular that the choice of $\mathbf{P}_0^a$ may not be so crucial. This result applies to the linear autonomous case, i.e. that where $M(t_{i-1}, t_i)$ and H_i are linear and their gradients and the $\mathbf{Q}_i$, $\mathbf{R}_i$ are independent of i. Here $M(t_{i-1}, t_i)$ is non linear but we may consider the situation where it is nearly linear its gradient depending on time only through the difference $t_i - t_{i-1}$, assumed to be constant. In such a case, $\mathbf{M}_{i-1,i}$, being almost independent of the point $\mathbf{x}^a(t_{i-1})$ of its evaluation and with $\mathbf{x}^a$ usually evolving slowly, it will be nearly constant in time. Such a situation, with H_i linear and $\mathbf{H}_i$, $\mathbf{Q}_i$, $\mathbf{R}_i$ independent of i (as is often the case), will be referred to as the "near linear autonomous" case. But, the theory of the Riccati equation for linear autonomous (or near autonomous) systems states that under the "controllability" (or stabilizability) and "observability" (or detectability) conditions,"*any solution of (2.5) will converge to the same limit regardless of the initial* $\mathbf{P}_0^a$" [*Hagger and Horowitz*, 1976]. Thus even if $\mathbf{P}_0^a$ is poorly specified, one still obtains a good approximation for $\mathbf{P}_i^a$, hence for $\mathbf{K}_i$, *provided* i *is large enough*. But here again, the huge dimension of the system could make this statement *inapplicable*: convergence may be so slow that $\mathbf{P}_i^a$ never approach its limit (bearing in mind that our system can only be considered "autonomous" over a fairly short time). Furthermore, the "controllability" (or stabilizability) condition, taken from the control theory with the model error $\eta(t_i)$ playing the role of control, requires that the state of the system may be brought to any point in a neighborhood of its starting point by some realization of $\eta(t_i)$. While this may be true in absolute term, there will be a large part of the phase space that the system will be very unlikely to enter because it is an *attractor* (we will return to this feature later). Thus one is situated at the limit of validity of the "controllability" condition.

The above considerations show that it may be too ambitious to demand optimality, even to an approximate degree. One may have to content oneself with a more modest aim: stability. By this we mean that the filtering error should remain bounded. Consider a filter algorithm defined by the equation (2.6) with a certain gain matrix $\mathbf{K}_i$ *not necessarily satisfying* (2.5), (2.7) (the filter not necessarily being an EKF). Then from (2.1') the filter error is propagated according to

$$\mathbf{x}^a(t_i) - \mathbf{x}^t(t_i) \approx (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{M}_{i-1,i}[\mathbf{x}^a(t_{i-1}) - \mathbf{x}^t(t_{i-1})] - \mathbf{K}_i\epsilon_i - (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\eta(t_i) \qquad (2.8)$$

This formula shows that the filter error behavior depends essentially on the sequence of matrices $(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{M}_{i-1,i}$. To simplify, consider the "near linear autonomous" case where $\mathbf{M}_{i-1,i}$ may be regarded as a constant equal to $\mathbf{M}$ (over a certain period of time) and $\mathbf{H}_i = \mathbf{H}$ and let us assume that $\mathbf{K}_i$ is chosen constant and equal to $\mathbf{K}$. Then what matters are the eigenvalues of $\mathbf{M} - \mathbf{KHM}$. If one such eigenvalue has modulus greater than 1, then the error will diverge. On the other hand, if the matrix $\mathbf{M} - \mathbf{KHM}$ is stable, in that all its eigenvalues have moduli less than 1, then it will remain bounded. Thus, our aim would be to make the matrix $(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{M}_{i-1,i}$ stable. Referring to (2.8), this means that one aims to reduce the

error propagation of the filter, so that it is attenuated at each step. We will show in the next section that this can be achieved by making corrections only in the directions in which the errors are not attenuated by the system. Moreover, this results in a low cost filter with little sacrifice to the global performance. It is worth pointing out that while the full KF is stable if the controllability and observability conditions are met, this stability could be too weak, i.e. the error attenuation at each step is too small, to be practically useful. This is because, as it has been explained above, one may be at the limit of validity of the controllability condition. By concentrating the correction on "critical" directions, it is hoped that stability can be improved.

It should be noted that the stability property is related to the system dynamics $\mathbf{M}_{i-1,i}$, the observation operator H_i and the choice of the filter gain $\mathbf{K}_i$, but does not depend on the system and observational error covariance matrices $\mathbf{Q}_i$ and $\mathbf{R}_i$ (provided that linearization errors are neglected). Thus by focusing on this property one avoids the need to specify these matrices with a high degree of accuracy.

3. Singular evolutive extended Kalman (SEEK) filter

Although the physical systems we are dealing with are of very high dimension, their "effective degrees of freedom" might not be so high. Indeed, the ocean is basically a forced and dissipated dynamical system possessing an attractor, meaning that asymptotically the trajectories of the state vector occupy only a small part of the phase space, the so called "attractor". Clearly the description of this attractor implies a much lower degree of freedom than the description of the whole system. In addition, in the case of the ocean (and of the atmosphere) the attractor is "strange" (i.e. even a remote vector state converges towards the attractor while at the same time even a slight perturbation of the state vector can develop into a marked divergence in the phase space) which affects to the predictability properties of the oceanic system. This has some implications on our filter, discussed later. There has been little research into this problem, especially in the case of the ocean, but this provides solid grounds for thinking in terms of reduced state dimension in order to describe system statistics.

Practical constraints on actual implementation of the KF or EKF leave no choice but to reduce the size of the system in some way. Schematically several strategies are possible. Firstly, one can simply reduce the dimension of the model state vector to make it compatible with the computational constraint [*Miller and Cane*, 1989; *Gaspar and Wunsch*, 1989; *Gourdeau et al.*, 1992; *Fukumori et al.*, 1993; *Benveniste et al.*, 1994]. Quite often, one also assumes a certain asymptotic property of constancy for the error covariance matrix [*Gourdeau et al.*, 1992; *Fu et al.*, 1993; *Benveniste et al.*, 1994; *Fukumori et Malanotte-Rizzoli*, 1995, *Fukumori*, 1995). Secondly, one might consider reducing the KF working space until the

computation becomes feasible (without necessarily reducing the size of the model !). *Fukumori* [1995], for example, undersamples the computational grid used for the filter description. Several studies are based on some ways of projecting the filter evolution on a reduced dimension basis [*Evensen*, 1994; *Cane et al.*, 1995, *Hoang et al.*, 1995]. Another approach consists in relying on physical considerations which will naturally decrease the degree of freedom of the system: long wave approximation in the tropics [*Cane and Patton*, 1984] and geostrophy [*Dee*, 1991], being examples, amongst others.

Here we adopt a somewhat different approach while retaining the idea of reducing the KF working space. Indeed, on physical grounds, it would seems legitimate to assume at the very least, the feasibility of representing the covariance matrix of the state vector to a good approximation by a low rank matrix. Likewise for the error covariance matrix since this should be smaller than the former. Our idea consists in using a low rank matrix $\mathbf{P}_i^a$ in the EKF algorithm while keeping the rest of the algorithm essentially unchanged. It will be seen below that in the absence of dynamic noise ($\mathbf{Q}_i = \mathbf{0}$), the matrices $\mathbf{P}_i^a$ in the Riccati equation all have the same rank as $\mathbf{P}_0^a$. Therefore it is only necessary to initialize the filtering process with a low rank $\mathbf{P}_0^a$.

In a first stage, we will develop our filter in the context of no dynamic noise, which in fact corresponds to the framework of the strong constraint variational approach. We will actually use this only as a working assumption, in order to develop the filter more easily. It will be later shown that our filter possess good stability property in the near linear autonomous case. Furthermore, it operates on the effective and readily understood principle of making no correction in the directions where the error propagation is attenuated by the system dynamics. One should bear in mind that, as our system is a strange attractor, there will be an appreciable number of directions for which the errors will be amplified. These directions are those in which corrections will be made in priority, as can be seen in section 3.1.2. A further point to be noted, as mentioned at the end of section 2, is that the above stability *does not depend on the specification of* $\mathbf{Q}_i$ *and* $\mathbf{R}_i$, hence our filter can continue to be used in the case where dynamic noise exists, without losing this property. Such noise only entails some loss of performance. But as was pointed out previously, optimality is an elusive aim. If $\mathbf{Q}_i \neq \mathbf{0}$ but is imperfectly specified, there is also a performance loss, and it may very well be no worse to simply put $\mathbf{Q}_i = \mathbf{0}$ if the noise is not too high.

When $\mathbf{Q}_i \neq \mathbf{0}$, the matrix $\mathbf{P}_i^a$ in the EKF will not be of low rank even if one initializes the filter with a low rank $\mathbf{P}_0^a$. However, one may expect $\mathbf{P}_i^a$ to be represented to a close approximation by a low rank matrix. The problem is that this approximation would depend on $\mathbf{Q}_i$ (which could be grossly mis-specified) and there is no "cheap" method for obtaining it. We will therefore construct our filter on the principle upon which the previous one operates, i.e. making no corrections in those directions for which the error propagation is attenuated by the system dynamics. Unlike the previous filter which is an EKF for $\mathbf{Q}_i = \mathbf{0}$, this one is

not EKF except for a very particular form of $\mathbf{Q}_i$. However, it should have good performance since it makes correction in the directions for which the error is likely to be high. More importantly still, it is designed to possess the stability property.

We also consider a further version, identical to the first one but with a "forgetting" factor, designed to take into account the linearization errors and possibly some dynamic noise as well. Its appeal is its simplicity and its low cost (as for the first version).

Because of its stability, our filter possess the interesting property of being able to correct itself in the case of a bad initialization: the sequence $\mathbf{P}_i^a$ is in the long run little affected by the initial $\mathbf{P}_0^a$. Nevertheless, in high dimensional systems, the choice of $\mathbf{P}_0^a$ remains important: the time taken by the filter to correct the effect of bad initialization could prove long, which would then hamper its capacity to track the evolution of the system dynamics. Taking this into consideration, we provide a initialization method based on the EOFs (empirical orthogonal functions).

3.1. No dynamic noise version

3.1.1. Filter description

When $\mathbf{Q} = \mathbf{0}$, the Riccati equation (2.5) can be simplified to:

$$\mathbf{P}_i^a = \mathbf{M}_{i-1,i}[\mathbf{P}_{i-1}^a - \mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T\mathbf{H}_i^T(\mathbf{H}_i\mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a\mathbf{M}_{i-1,i}^T\mathbf{H}_i^T + \mathbf{R}_i)^{-1}\mathbf{H}_i\mathbf{M}_{i-1,i}\mathbf{P}_{i-1}^a]\mathbf{M}_{i-1,i}$$

Let us factorize $\mathbf{P}_0^a$ into $\mathbf{L}_0\mathbf{U}_0\mathbf{L}_0^T$ where $\mathbf{U}_0$ is a positive definite matrix with dimension equal to the rank of $\mathbf{P}_0^a$ (this being low in practical applications). Then:

$$\mathbf{P}_1^a = \mathbf{L}_1[\mathbf{U}_0 - \mathbf{U}_0\mathbf{L}_1^T\mathbf{H}_1^T(\mathbf{H}_1\mathbf{L}_1\mathbf{U}_0\mathbf{L}_1^T\mathbf{H}_1^T + \mathbf{R}_1)^{-1}\mathbf{H}_1\mathbf{L}_1\mathbf{U}_0]\mathbf{L}_1^T,$$

where $\mathbf{L}_1 = \mathbf{M}_{0,1}\mathbf{L}_0$. Repeating this computation, one obtains

$$\begin{aligned}
\mathbf{L}_i &= \mathbf{M}_{i-1,i}\mathbf{L}_{i-1}, \\
\mathbf{P}_i^a &= \mathbf{L}_i\mathbf{U}_i\mathbf{L}_i^T, \\
\mathbf{U}_i &= \mathbf{U}_{i-1} - \mathbf{U}_{i-1}\mathbf{L}_i^T\mathbf{H}_i^T(\mathbf{H}_i\mathbf{L}_i\mathbf{U}_{i-1}\mathbf{L}_i^T\mathbf{H}_i^T + \mathbf{R}_i)^{-1}\mathbf{H}_i\mathbf{L}_i\mathbf{U}_{i-1}. \qquad (3.1)
\end{aligned}$$

The above equations show that the rank of $\mathbf{P}_i^a$ can only decrease with i. In fact it remains unchanged (if the $\mathbf{R}_i$ are invertible), since it can be shown that for invertible $\mathbf{U}_{i-1}$, the right hand side of (3.1) is also invertible with inverse given by

$$\mathbf{U}_i^{-1} = \mathbf{U}_{i-1}^{-1} + \mathbf{L}_i^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{L}_i. \qquad (3.2)$$

Thus $\mathbf{U}_i$ is invertible for all i and

$$\mathbf{U}_i^{-1} = \mathbf{U}_0^{-1} + \sum_{j=1}^{i} \mathbf{L}_j^T \mathbf{H}_j^T \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{L}_j. \tag{3.3}$$

The filter gain $\mathbf{K}_i$ is computed from (2.7), (3.1), (3.2) and it defines the filter equation (2.6). Since the columns of $\mathbf{K}_i$ are a linear combination of those of $\mathbf{L}_i$, corrections *take place only in the directions parallel to the space spanned by these columns.*

3.1.2. Stability consideration and interpretation

We consider here the near linear autonomous case where $\mathbf{M}_{i-1,i}$, $\mathbf{H}_i$, $\mathbf{Q}_i$, $\mathbf{R}_i$ can all be considered constant* (for a certain duration) and we thus drop the subscript i where appropriate. We can then recognize the iteration $\mathbf{L}_i = \mathbf{M}\mathbf{L}_{i-1}$ as being the power method for computing the eigenvectors of $\mathbf{M}$. Indeed, when the $\mathbf{L}_i$ reduce to column vectors, it is well known that they tend to align with the first eigenvector of $\mathbf{M}$, the eigenvectors being ordered according to decreasing moduli of their eigenvalues. In general when the $\mathbf{L}_i$ have r columns, the linear space spanned by them tends to align with that spanned by the first r eigenvectors of $\mathbf{M}$.

To be explicit, consider the invariant subspaces associated with the first r eigenvalues of $\mathbf{M}$ and those remaining (which are assumed to have have strictly smaller moduli than the former). Let a basis of these spaces be given by the columns of $\mathbf{V}$ and $\mathbf{V}_\ominus$, respectively, so that $\mathbf{M}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$, $\mathbf{M}\mathbf{V}_\ominus = \mathbf{V}_\ominus\mathbf{\Lambda}_\ominus$ with $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\ominus$ having as eigenvalues the first r eigenvalues of $\mathbf{M}$ and the remaining ones, respectively. One can express $\mathbf{L}_0$ as $\mathbf{V}\mathbf{N} + \mathbf{V}_\ominus\mathbf{N}_\ominus$ for some matrices $\mathbf{N}$ and $\mathbf{N}_\ominus$. Assuming that $\mathbf{N}$ (a square matrix) is invertible, it can be absorbed into $\mathbf{V}$ by replacing $\mathbf{V}\mathbf{N}$ by $\mathbf{V}$ and $\mathbf{\Lambda}$ by $\mathbf{N}^{-1}\mathbf{\Lambda}\mathbf{N}$. Then it can be seen that

$$\mathbf{L}_i\mathbf{\Lambda}^{-i} = \mathbf{V} + \mathbf{V}_\ominus\mathbf{\Lambda}_\ominus^i\mathbf{N}_\ominus\mathbf{\Lambda}^{-i} \to \mathbf{V} \qquad \text{as } i \to \infty.$$

Let r^* be the number of eigenvalues of $\mathbf{M}$ of modulus (strictly) greater than 1. In the case where $r \leq r^*$, all the eigenvalues of the matrix $\mathbf{\Lambda}^{-1}$ has modulus less than 1. Hence by (3.3) and the above result

$$(\mathbf{\Lambda}^T)^{-i}\mathbf{U}_i^{-1}\mathbf{\Lambda}^{-i} \to \sum_{j=0}^{\infty}(\mathbf{\Lambda}^T)^{-j}\mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{V}\mathbf{\Lambda}^{-j} = \mathbf{\Pi}^{-1} \qquad \text{as } i \to \infty,$$

This argument does not applies to the case $r > r^*$ since then the matrix $\mathbf{\Lambda}^{-1}$ has an eigenvalue of modulus greater tan or equal to 1. Therefore we change our definitions of $\mathbf{V}$, $\mathbf{V}_\ominus$ and $\mathbf{\Lambda}$

* In applications, $\mathbf{H}_i$ may vary in time in a cyclical manner, as, for example, is in the case in oceanographic satellite observations, but by considering a complete cycle and enlarging the state vector accordingly one is brought back to the situation where it is constant.

somewhat. We now take $\mathbf{V}$ and $\mathbf{V}_\ominus$ to be matrices of which the columns are a basis of the invariant spaces associated with the r^* first eigenvectors and with the remaining ones, respectively, and define $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\ominus$ as above. Assume that $\mathbf{M}$ admits no eigenvalue of unit modulus so that $\mathbf{\Lambda}_\ominus$ has all its eigenvalues of modulus less than 1. Again expressing $\mathbf{L}_0$ as $\mathbf{VN} + \mathbf{V}_\ominus\mathbf{N}_\ominus$ and assuming that the matrix $\mathbf{N}$ is of full rank, there then exists a $r \times r$ invertible matrix $\mathbf{T}$ such that the first r^* columns of $\mathbf{NT}$ form an invertible matrix while the remaining ones are zero. As previously, this invertible matrix can be absorbed into $\mathbf{V}$, so that the matrices formed by the first r^* rows of $\mathbf{L}_0\mathbf{T}$ and by those remaining can be written respectively as $\mathbf{V} + \mathbf{V}_\ominus\tilde{\mathbf{N}}$ and $\mathbf{V}_\ominus\tilde{\mathbf{N}}_\ominus$, for some matrices $\tilde{\mathbf{N}}$ and $\mathbf{N}_\ominus$. By a similar (but much more complicated) calculation to the one above, it can be shown that

$$\mathbf{L}_i\mathbf{T}\begin{bmatrix}\mathbf{\Lambda}^{-i} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \to [\mathbf{V}\ \mathbf{0}] \qquad \text{as } i \to \infty$$

and

$$\begin{bmatrix}\mathbf{\Lambda}^{-i} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}^T \mathbf{T}^T\mathbf{U}_i^{-1}\mathbf{T}\begin{bmatrix}\mathbf{\Lambda}^{-i} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \to \begin{bmatrix}\mathbf{\Pi}^{-1} & \mathbf{0}\\ \mathbf{0} & *\end{bmatrix}, \qquad \text{as } i \to \infty,$$

$\mathbf{\Pi}$ being as before and $*$ denoting an unspecified positive definite matrix.

Thus in all cases, $\mathbf{P}_i^a \to \mathbf{V\Pi V}^T$ as $i \to \infty$. This limit is of rank $\min(r, r^*)$ and is independent of the initial $\mathbf{P}_0^a$ as long as it is of rank of at least r^* and certain generic assumptions (such as $\mathbf{N}$ of full rank ...) are met.

It follows from this and (2.7) that $\mathbf{K}_i$ tends to $\mathbf{K} = \mathbf{V\Pi V}^T\mathbf{H}^T\mathbf{R}^{-1}$. It can be verified that the matrix $\mathbf{M - KHM}$ is indeed stable, *provided that* $r \geq r^*$. Indeed

$$\begin{aligned}(\mathbf{M - KHM})\mathbf{V} &= \mathbf{V\Lambda - V\Pi V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV\Lambda}\\ &= \mathbf{V\Pi}(\mathbf{\Pi}^{-1} - \mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV})\mathbf{\Lambda}\\ &= \mathbf{V\Pi}(\mathbf{\Lambda}^T)^{-1}\mathbf{\Pi}^{-1}.\end{aligned}$$

The last equality following from the fact that $\mathbf{\Pi}^{-1}$ satisfies the Lyapounov equation: $\mathbf{\Pi}^{-1} = (\mathbf{\Lambda}^T)^{-1}\mathbf{\Pi}^{-1}\mathbf{\Lambda}^{-1} + \mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV}$. This yields

$$(\mathbf{M - KHM})[\mathbf{V}\ \mathbf{V}_\ominus] = [\mathbf{V}\ \mathbf{V}_\ominus]\begin{bmatrix}\mathbf{\Pi\Lambda}^{-1T}\mathbf{\Pi}^{-1} & \mathbf{\Pi V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV}_\ominus\mathbf{\Lambda}_\ominus\\ \mathbf{0} & \mathbf{\Lambda}_\ominus\end{bmatrix}. \qquad (3.4)$$

This formula shows that the eigenvalues of $\mathbf{M - KHM}$ are the same as those of $\mathbf{\Lambda}^{-1}$ and $\mathbf{\Lambda}_\ominus$. But by definition $\mathbf{\Lambda}$ has as eigenvalues the first $\min(r, r^*)$ eigenvalues of $\mathbf{M}$ and $\mathbf{\Lambda}_\ominus$ the remaining ones. Thus $\mathbf{M - KHM}$ is stable if and only if $r \geq r^*$.

An interpretation of the above filter is as follows. Suppose that we have committed an error $\mathbf{e}$ in the analysis state vector $\mathbf{x}^a(t_i)$. If no further correction is made, then at time t_{i+k} the error will have become $\mathbf{M}^k\mathbf{e}$ (assuming that no dynamic noise enter the system). Thus if the error is parallel to the space spanned by the columns of $\mathbf{V}_\ominus$, that is $\mathbf{e} = \mathbf{V}_\ominus\mathbf{f}$

for a certain $\mathbf{f}$, then it dies out automatically, since $\mathbf{M}^k\mathbf{V}_\ominus = \mathbf{V}_\ominus\mathbf{\Lambda}_\ominus^k$ and all the eigenvalues of the matrix $\mathbf{\Lambda}_\ominus$ are of modulus less than 1. On the other hand, if the error is parallel to the space spanned by the columns of $\mathbf{V}$, then it will diverge since $\mathbf{M}^k\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^k$ and all the eigenvalues of the matrix $\mathbf{\Lambda}$ have modulus greater than 1. It is therfore intuitively clear that all efforts should be concentrated on correcting the latter type of error and there is no need to correct the former type. Our filter operates precisely on this principle. To see this, consider (for simplicity) the limiting case where $\mathbf{K}_i$ has converged to $\mathbf{K}$. By (2.8), a filter error will be transmitted through the multiplication with $\mathbf{M} - \mathbf{KHM}$. If the error is of the form $\mathbf{Vf}$, it is transformed into $\mathbf{V\Pi\Lambda}^{-1T}\mathbf{\Pi}^{-1}\mathbf{f}$ by (3.4), is hence attenuated and moreover remains of the same form. If the error is of the form $\mathbf{V}_\ominus\mathbf{f}$, it is simply attenuated by the system dynamics, the correction even creating an additional error parallel to the space spanned by the columns of $\mathbf{V}$ (because $\mathbf{K}$ contains the factor $\mathbf{V}$). But this error will be attenuated next time. Finally, the observational errors ϵ_i enter the system only through $\mathbf{K}$, are hence parallel to the space spanned by the columns of $\mathbf{V}$ and will thus be also attenuated.

3.2. General version

3.2.1. Filter description

We have just seen the logic of the previous filter which was to make no error correction in those directions for which the error will be attenuated by the system itself. Thus in the absence of dynamic noise, the error in these directions will gradually decay and die out so that whether or not correction in such directions is made, it will make no difference. The situation is different in the case where dynamic noise is present, since this continually enters the system and its cumulative effect cannot be eliminated by any attenuation. Thus an appropriate correction in any direction is still beneficial and our filter will inevitably be non optimal, except in the absence of dynamic noise in those directions where no correction is made. Nevertheless, one can argue that the error will be very large in those directions where it is amplified by the system, hence it is of prime importance and greater benefit to make correction in those directions first. Correction in the directions for which the error is greatly attenuated by the system would result in only marginal benefit and, in view of the computational cost and the need to specify correctly the dynamic noise covariance matrix, it may not be worthwhile. Thus we propose *not to perform such correction*.

Proceeding on this principe, we consider as before the near linear autonomous case so that $\mathbf{M}_{i-1,i}$, $\mathbf{Q}_i$, $\mathbf{H}_i$ and $\mathbf{R}_i$ can be regarded as independent of i (over a certain duration) and we shall drop the latter. Let $\mathbf{V}$ be a matrix whose columns form a basis of the invariant space associated with the first r eigenvectors of $\mathbf{M}$, where r is chosen such that the remaining eigenvalues have small modulus, less than 1 in any case. We consider the change of basis

$$\mathbf{x} = [\mathbf{V}\ \mathbf{V}_\ominus]\begin{bmatrix}\mathbf{z}\\ \mathbf{z}_\ominus\end{bmatrix}, \qquad \begin{bmatrix}\mathbf{z}\\ \mathbf{z}_\ominus\end{bmatrix} = \begin{bmatrix}\mathbf{W}\\ \mathbf{W}_\ominus\end{bmatrix}\mathbf{x},$$

where $\mathbf{V}_\ominus$ is a matrix with columns forming a basis of the orthogonal complement to the space spanned by the columns of $\mathbf{V}$ and $\mathbf{W} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$, $\mathbf{W}_\ominus = (\mathbf{V}_\ominus^T\mathbf{V}_\ominus)^{-1}\mathbf{V}_\ominus^T$, so that the first matrix in the last right hand side is the inverse of that in the preceeding right hand side. This change of basis will be applied simultaneously to $\mathbf{x}^a$, $\mathbf{x}^f$ and $\mathbf{x}^t$. Thus, $\mathbf{x}^f(t_i) = \mathbf{V}\mathbf{z}^f(t_i) + \mathbf{V}_\ominus\mathbf{z}^f_\ominus(t_i)$ where $\mathbf{z}^f(t_i) = \mathbf{W}\mathbf{x}^f(t_i)$ and $\mathbf{z}^f_\ominus(t_i) = \mathbf{W}_\ominus\mathbf{x}^f(t_i)$. The idea is to perform no correction on $\mathbf{z}^f_\ominus(t_i)$, i.e. to take as correction equations:

$$\mathbf{z}^a(t_i) = \mathbf{z}^f(t_i) + \tilde{\mathbf{K}}_i[\mathbf{y}_i^o - H_i\mathbf{x}^f(t_i)], \qquad \mathbf{z}^a_\ominus(t_i) = \mathbf{z}^f_\ominus(t_i),$$

where $\tilde{\mathbf{K}}_i$ is a gain matrix to be defined. Since $\mathbf{x}^a(t_i) = \mathbf{V}\mathbf{z}^a(t_i) + \mathbf{V}_\ominus\mathbf{z}^a_\ominus(t_i)$, the above equations reduce to (2.6) with $\mathbf{K}_i = \mathbf{V}\tilde{\mathbf{K}}_i$.

Note that

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{W}_\ominus \end{bmatrix} \mathbf{M}[\mathbf{V}\ \mathbf{V}_\ominus] = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{W}\mathbf{M}\mathbf{V}_\ominus \\ \mathbf{0} & \mathbf{\Lambda}_\ominus \end{bmatrix}, \qquad \mathbf{\Lambda}_\ominus = \mathbf{W}_\ominus\mathbf{M}\mathbf{V}_\ominus,$$

which shows that the first r eigenvalues of $\mathbf{M}$ and the remaining ones are those of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\ominus$, respectively. From the above equation, (2.1') can be written as

$$\begin{aligned} \mathbf{z}^t(t_i) &\approx \mathbf{z}^f(t_i) + \mathbf{\Lambda}[\mathbf{z}^t(t_{i-1}) - \mathbf{z}^a(t_{i-1})] + \mathbf{W}\mathbf{M}\mathbf{V}_\ominus[\mathbf{z}^t_\ominus(t_{i-1}) - \mathbf{z}^a_\ominus(t_{i-1})] + \mathbf{W}\eta(t_i) \quad (3.5) \\ \mathbf{z}^t_\ominus(t_i) &\approx \mathbf{z}^f_\ominus(t_i) + \mathbf{\Lambda}_\ominus[\mathbf{z}^t_\ominus(t_{i-1}) - \mathbf{z}^a_\ominus(t_{i-1})] + \mathbf{W}_\ominus\eta(t_i). \end{aligned}$$

This equation together with $\mathbf{z}^a_\ominus(t_i) = \mathbf{z}^f_\ominus(t_i)$ yields the error propagation equation for $\mathbf{z}^t_\ominus$: $\mathbf{z}^a_\ominus(t_i) - \mathbf{z}^t_\ominus(t_i) = \mathbf{\Lambda}_\ominus[\mathbf{z}^a_\ominus(t_{i-1}) - \mathbf{z}^t_\ominus(t_{i-1})]$. Because of the attenuation effect of $\mathbf{\Lambda}_\ominus$, the error $\mathbf{z}^a_\ominus - \mathbf{z}^t_\ominus$ will be small (but will not die out, unless $\mathbf{W}_\ominus\eta(t_i) = \mathbf{0}$). Ignoring this error, equation (3.5) has the same form as (2.1') in the EKF, with $\mathbf{z}$ playing the role of $\mathbf{x}$, $\mathbf{\Lambda}$ the role of $\mathbf{M}$ and $\mathbf{W}\eta$ the role of η. Note that the linearized observation equation (2.2) can be written as

$$\mathbf{y}_i^o \approx H_i\mathbf{x}^f(t_i) + \mathbf{H}\mathbf{V}[\mathbf{z}^t(t_i) - \mathbf{z}^f(t_i)] + \mathbf{H}\mathbf{V}_\ominus[\mathbf{z}^t_\ominus(t_i) - \mathbf{z}^f_\ominus(t_i)] + \epsilon_i,$$

which can be considered to be the "observation equation for $\mathbf{z}^t$" if one regards $\mathbf{z}^t_\ominus$ as equal to $\mathbf{z}^f_\ominus$ (that is ignoring the error $\mathbf{z}^t_\ominus - \mathbf{z}^f_\ominus$). These considerations suggest our taking as $\tilde{\mathbf{K}}_i$ the Kalman gain in the filtering problem for $\mathbf{z}$, i.e. $\tilde{\mathbf{K}}_i$ is given by (2.5), (2.7) with $\mathbf{M}_{i-1,i}$ replaced by $\mathbf{\Lambda}$, $\mathbf{H}_i$ by $\mathbf{H}\mathbf{V}$, $\mathbf{R}_i$ by $\mathbf{R}$ and $\mathbf{Q}_i$ by $\mathbf{W}\mathbf{Q}\mathbf{W}^T$. Explicitly, according to (2.5) and (2.7), $\tilde{\mathbf{K}}_i = \mathbf{\Pi}_i\mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}$, where

$$\begin{aligned} \mathbf{\Pi}_i = \mathbf{\Lambda}\mathbf{\Pi}_{i-1}\mathbf{\Lambda}^T + \mathbf{W}\mathbf{Q}\mathbf{W}^T - (\mathbf{\Lambda}\mathbf{\Pi}_{i-1}\mathbf{\Lambda}^T + \mathbf{W}\mathbf{Q}\mathbf{W}^T)\mathbf{H}^T\mathbf{V}^T \\ [\mathbf{H}\mathbf{V}(\mathbf{\Lambda}\mathbf{\Pi}_{i-1}\mathbf{\Lambda}^T + \mathbf{W}\mathbf{Q}\mathbf{W}^T)\mathbf{V}^T\mathbf{H}^T + \mathbf{R}]^{-1}\mathbf{H}\mathbf{V}(\mathbf{\Lambda}\mathbf{\Pi}_{i-1}\mathbf{\Lambda}^T + \mathbf{W}\mathbf{Q}\mathbf{W}^T) \end{aligned}$$

To compute $\mathbf{\Pi}_i$, one needs $\mathbf{V}$ and $\mathbf{\Lambda}$, which can be obtained from $\mathbf{M}$. But our assumption $\mathbf{M}_{i-1,i} = \mathbf{M}$ was only postulated in order to simplify the arguments. In fact

$\mathbf{M}_{i-1,i}$ evolves with time, albeit slowly. To take this into account, we draw our inspiration from the computations in section 3.1.2. Define the sequences

$$\mathbf{V}_i = \mathbf{M}_{i-1,i}\mathbf{V}_{i-1}\mathbf{\Lambda}_i^{-1}, \quad \mathbf{\Lambda}_i = \mathbf{W}_{i-1}\mathbf{M}_{i-1,i}\mathbf{V}_{i-1}, \quad \mathbf{W}_{i-1} = (\mathbf{V}_{i-1}^T\mathbf{V}_{i-1})^{-1}\mathbf{V}_{i-1}^T \tag{3.6}$$

starting from some $\mathbf{V}_0$. Then, if all the $\mathbf{M}_{i-1,i}$ are equal to $\mathbf{M}$, it can be shown that $\mathbf{V}_i$ converges generally to a limit $\mathbf{V}$ with columns spanning the invariant space associated with the first r eigenvectors of $\mathbf{M}$. The $\mathbf{\Lambda}_i$ also converge to $\mathbf{\Lambda}$, satisfying $\mathbf{MV} = \mathbf{V\Lambda}$. Thus, we shall take as our filter gain that given by (2.7) with $\mathbf{P}_i^a = \mathbf{V}_i\mathbf{\Pi}_i\mathbf{V}_i^T$ and $\mathbf{\Pi}_i$ defined as above but replacing $\mathbf{H}$, $\mathbf{V}$, $\mathbf{\Lambda}$, $\mathbf{W}$, $\mathbf{R}$, $\mathbf{Q}$ by $\mathbf{H}_i$, $\mathbf{V}_i$, $\mathbf{\Lambda}_i$, $\mathbf{W}_i$, $\mathbf{R}_i$, $\mathbf{Q}_i$. The resulting equation for $\mathbf{\Pi}_i$ can alternative be written as

$$\mathbf{\Pi}_i^{-1} = (\mathbf{\Lambda}_i\mathbf{\Pi}_{i-1}\mathbf{\Lambda}_i^T + \mathbf{W}_i\mathbf{Q}_i\mathbf{W}_i^T)^{-1} + \mathbf{V}_i^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{V}_i. \tag{3.7}$$

It is instructive to note the similarity of this filter and the one in section 3.1.1. Indeed, putting

$$\mathbf{L}_i = \mathbf{V}_i\mathbf{\Lambda}_i\cdots\mathbf{\Lambda}_1, \qquad (\mathbf{L}_0 = \mathbf{V}_0)$$

and

$$\mathbf{U}_i = \mathbf{\Lambda}_1^{-1}\cdots\mathbf{\Lambda}_i^{-1}\mathbf{\Pi}_i(\mathbf{\Lambda}_i^T)^{-1}\cdots(\mathbf{\Lambda}_1^T)^{-1}, \qquad \mathbf{U}_0 = \mathbf{\Pi}_0.$$

Then clearly $\mathbf{P}_i^a = \mathbf{L}_i\mathbf{U}_i\mathbf{L}_i^T$ and (3.6), (3.7) can be rewritten in terms of $\mathbf{L}_i$, $\mathbf{U}_i$ as

$$\begin{aligned} \mathbf{L}_i &= \mathbf{M}_{i-1,i}\mathbf{L}_{i-1}, \\ \mathbf{U}_i^{-1} &= [\mathbf{U}_{i-1} + (\mathbf{L}_i^T\mathbf{L}_i)^{-1}\mathbf{L}_i^T\mathbf{Q}_i\mathbf{L}_i(\mathbf{L}_i^T\mathbf{L}_i)^{-1}]^{-1} + \mathbf{L}_i^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{L}_i. \end{aligned} \tag{3.8}$$

The last equation is the same as (3.2) except for an extra term involving $\mathbf{Q}_i$. The first equation is also the same as the first equation of (3.1). Thus this algorithm reduces to the one in section 3.1.1 in the case where $\mathbf{Q}_i = \mathbf{0}$. However, the algorithm in term of $\mathbf{L}_i$ and $\mathbf{U}_i$ may be numerically unstable, since $\mathbf{L}_i$ and $\mathbf{U}_i^{-1}$ tend to diverge through repeated multiplication by $\mathbf{M}_{i-1,i}$. The algorithm (3.6), (3.7) is numerically stable since $\mathbf{V}_i$ is "renormalized" at each step by the division by $\mathbf{\Lambda}_i$. If the first algorithm is used, $\mathbf{L}_i$ needs to be renormalized periodically to avoid numerical instability.

3.2.2. Stability consideration

A remarkable property of the Riccati equation (2.5) referred to above is that, when $\mathbf{M}_{i-1,i}$, $\mathbf{H}_i$, $\mathbf{Q}_i$, $\mathbf{R}_i$ are independent of i and the observability and controllability conditions are met, its solution converges to a limit such that the matrix $\mathbf{M} - \mathbf{KHM}$ is stable, $\mathbf{K}$ denoting the corresponding limit of the $\mathbf{K}_i$ as given by (2.7). Thus, consider again the near linear autonomous case so that we may apply the above result to the $\mathbf{\Pi}_i$ sequence, which satisfies a similar Riccati equation. This yields that $\mathbf{\Pi}_i$ converges to $\mathbf{\Pi}$ such that the matrix

$$\mathbf{\Lambda} - \mathbf{\Pi}\mathbf{V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV\Lambda} \tag{3.9}$$

is stable. We must emphasize that such an argument does not applies in the case $\mathbf{Q}_i = \mathbf{0}$ considered in section 3.1.2, since the controllability condition is violated. Here the sequence $\mathbf{\Pi}_i$ comes from a KF restricted to a low dimensional space so the controllability condition holds quite generally. (In fact it suffices that $\mathbf{WQW}^T$ be non singular.) Therefore, the gain matrices $\mathbf{K}_i$ of our filter converge to $\mathbf{K} = \mathbf{V\Pi V}^T\mathbf{H}^T\mathbf{R}^{-1}$. Then a direct computation yields

$$(\mathbf{M} - \mathbf{KHM})\mathbf{V} = \mathbf{V}(\mathbf{\Lambda} - \mathbf{\Pi V}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HV\Lambda}).$$

which leads to the same formula as (3.4) but with $\mathbf{\Pi\Lambda}^{-1T}\mathbf{\Pi}^{-1}$ replaced by the matrix (3.9). Since the latter matrix is stable, using the same argument as that in the end of section 3.1.2, we obtain that our filter is indeed stable.

3.3. Version with forgetting factor

In the above derivations of the filter, we have ignored the linearization error in (2.1'). Denote this error by $\eta'(t_i)$, then (2.8) will contain an extra term $(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\eta'(t_i)$, showing that that this error has a similar effect to the dynamic noise $\eta(t_i)$, although it is unlikely to satisfy the independence and zero mean assumption. Thus, in the case where dynamic noise exists, one may simply increase the matrix $\mathbf{Q}_i$ to take into account the linearization error effect. Applying this idea, in the version with no dynamic noise in 3.1.1, one should introduce an artificial matrix $\mathbf{Q}_i$, which bring us back to the general version in 3.2.1. However, the choice of this $\mathbf{Q}_i$ is not obvious because nothing is known concerning the linearization error. A possible course of action is to chose a convinient $\mathbf{Q}_i$ for computation purposes. Thus we propose taking $(\mathbf{L}_i^T\mathbf{L}_i)^{-1}\mathbf{L}_i^T\mathbf{Q}_i\mathbf{L}_i(\mathbf{L}_i^T\mathbf{L}_i)^{-1} = \alpha\mathbf{U}_{i-1}$, where α is a positive constant. This choice is not standard, since $\mathbf{Q}_i$ is not explicit and depends on i through $\mathbf{U}_{i-1,i}$, but since the linearization error should increase with the filter error $\mathbf{x}^a(t_{i-1}) - \mathbf{x}^t(_{i-1})$, this choice seems logical. The main reason for it is that it simplifies the algorithm: (3.8) becomes

$$\mathbf{U}_i^{-1} = \rho\mathbf{U}_{i-1}^{-1} + \mathbf{L}_i^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{L}_i. \tag{3.10}$$

where $\rho = 1/(1+\alpha)$. This results in an essentially the same filter as the no dynamic noise version in 3.1.1, except that the matrix $\mathbf{U}_{i-1}$ is amplified by a factor of $1/\rho$ before entering the updating equation. Note that this filter can also be used in the case where dynamic noise is indeed present, but no precise knowledge about $\mathbf{Q}_i$ is available, so that a convenient specification such as the one above might be used. But, although convenient, this is clearly a very approximate approach to the complex but essential problem of an adequate specification of the model error, which should anyway be investigated further.

The factor ρ may be interpreted as a forgetting factor. In fact, as is shown in Pham et al. [1995], the filter in section 3.1.1 can be viewed as a recursive implementation of the minimization of the least squares criterion with constraint and penalty on the initial state. If, instead of the usual sum of the squares of the errors, one takes the weighted sum of the

squares with weight decreasing with k as ρ^k when one goes back k steps, then the above filter is obtained. This kind of weighting is precisely what a forgetting factor produces in adaptive algorithms. The primary purpose is to down-weight the earlier observations with respect to the recent ones, thus enabling the algorithm to follow changes in the underlying process. As a beneficial side effect, it also attenuates the unaccounted errors in the past, thus reducing their propagation and possible accumulation. This is a type of error compensation technique [Jazwinski, 1970]. It actually enhances the stability of the filter. Although the arguments in section 3.2.2 are no longer valid here ($\mathbf{Q}_i$ being not constant), those in section 3.1.1 may be adapted to show this stability.

3.4. Initialization of the filter: the EOFs technique

To initialize the Kalman algorithm, one needs some initial state vector $\mathbf{x}^a(t_0)$ and its initial error $\mathbf{P}_0^a$. In practice, the initial state estimate is based on some a priori knowledge and one has at best a vague idea of the covariance matrix of the error. A choice thus needs to be made for $\mathbf{x}^a(t_0)$ and $\mathbf{P}_0^a$. To this end, we propose to use the EOFs technique applied to observed, or possibly simulated, state sequences from the system. Simulation might prove necessary if no sufficiently long sequence of observed state vectors is available. In the no dynamic noise case, it is also quite easy to generate long sequence of state vectors from to the model equation (2.1). The initial state may be set arbitrarily if one has taken care to wait until the model has been settled into a stable regime (from the statistical point of view) thus discarding the spin-up phase irrelevant for our purposes.

One can take as $\mathbf{x}^a(t_0)$ the average of the simulated (or observed) state vectors and as $\mathbf{P}_0^a$ a low rank approximation of the sample covariance matrix $\mathbf{P}_0$ of these vectors *. To obtain such an approximation, we propose to use the EOFs approach, which provides, in a certain sense explained below, the best approximation. Let V_1, V_2, ..., be the eigenvectors of $\mathbf{P}_0$, of unit L^2-norm, ordered according to their eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$. These vectors are called empirical orthogonal functions (EOFs). In the EOFs method, $\mathbf{P}_0$ approximates to $\mathbf{L}_0\mathbf{U}_0\mathbf{L}_0^T$ where

$$\mathbf{L}_0 = [V_0 \ \cdots V_r], \qquad \mathbf{U}_0 = \mathrm{diag}(\lambda_1, \ldots, \lambda_r).$$

This is justified by the following result: *"if ξ_0 is a random vector of mean zero and covariance matrix $\mathbf{P}_0$, then $\mathbf{L}_0\mathbf{L}_0^T\xi_0$ realizes the minimum of $\mathrm{E}||\xi_0 - \Pi_r\xi_0||^2$ among all orthogonal projectors Π_r onto a linear subspace of dimension r"*. Further, direct computation shows that the random vector $\mathbf{L}_0\mathbf{L}_0^T\xi_0$ has covariance matrix $\mathbf{L}_0\mathbf{U}_0\mathbf{L}_o^T$ and the error vector $\xi_0 - \mathbf{L}_0\mathbf{L}_0^T\xi_0$ has expected squared norm $\sum_{j>r}\lambda_j$. Since the expected squared norm of ξ_0 is $\mathrm{Tr}(\mathbf{P}_0)$, the trace of $\mathbf{P}_0$, the relative error in squared L^2-norm is $\sum_{j>r}\lambda_j/\mathrm{Tr}(\mathbf{P}_0)$. This ratio can be used to assess the accuracy of the approximation and thus to choose the appropriate value for r.

* The rank of $\mathbf{P}_0$ is in fact the number of generated vectors minus 1 and can (and should) thus be rather high (although it is usually is still far less than the system dimension)

To apply the SEEK filter, one also needs to specify the matrices $\mathbf{R}_i$ and $\mathbf{Q}_i$, which are generally unknown. It is for this reason that the no noise version with forgetting factor in section 3.3. has a certain appeal, since it does not require knowledge of $\mathbf{Q}_i$. If $\mathbf{R}_i$ is taken as σ^2 times the identity matrix, this σ^2 need not be known either. Indeed, from (2.7), (3.1) and (3.10), one obtains

$$\mathbf{K}_i = \mathbf{L}_i(\mathbf{U}_i/\sigma^2)\mathbf{L}_i^T\mathbf{H}^T, \qquad \sigma^2\mathbf{U}_i^{-1} = \rho\sigma^2\mathbf{U}_{i-1}^{-1} + \mathbf{L}_i\mathbf{H}_1^T\mathbf{H}_1\mathbf{L}_i.$$

This means that only the $\mathbf{U}_i/\sigma^2$ enter the computation. Usually $\mathbf{U}_0$ is very large with respect to σ^2. In any case, since $\mathbf{x}^a(t_0)$ is chosen somewhat arbitrarily, it is safe to err on the conservative side, by taking $\mathbf{U}_0/\sigma^2$ extremely large which amounts to putting $\sigma^2\mathbf{U}_0^{-1} = 0$.

4. An example of assimilation

As an example illustrating the application of our filter, we have considered a simple application of a quasigeostrophic (QG) model of ocean circulation, commonly used to describe the dynamical evolution of geophysical fluids such as the atmosphere or the ocean [*Pedlosky*, 1987]. In the present case, it is considered in its layered formulation [*Holland*, 1978]. The ocean is thus subdivided into N layers of constant density. Vertical momentum is transferred vertically by the pressure coupling between layers. Horizontal circulation in each layer is represented by N horizontal streamfunction fields $\psi_1, \ldots, \psi_N$. This model has been extensively used in our group for process studies or more realistic application problems. Although these models are less complete than the fully diabatic primitive equation ones, they are known to be nicely adapted to representing fairly faithfully the basic features of some oceanic processes such as mid-latitude ocean eddy interactions.

4.1. Simulation setup

In the mid-latitudes the ocean circulation is often characterized by very turbulent activity arizing from strong interactions between the so-called mesoscale eddies. The prototype region of this mesoscale activity is the Gulf Stream system where the barotropic and baroclinic instability of the intense inertial currents produces strong mesoscale eddies. The energy present in these eddies is at least of the order of the mean current energy and often much greater. Other western boundary ocean current system in the world also exhibit such behavior. Turbulent eddy activity is also encountered in the quieter areas of the ocean universe such as the interior regions of ocean basins, although to a lesser degree.

In the present case, a very schematic box-model of this turbulent eddy activity was considered. The domain is a simple square ocean 4000 km wide. A schematic wind stress pattern forced a circulation pattern made up of a number of mesoscale eddies engaged in strong nonlinear interactions.

The numerical domain was divided into a 41×41 grid. The stratification was assumed to be 3 layered. The quasigeostrophic equations were solved with an integration step of 22500 sec., with 6 such steps (the first being Euler forward the others leap-frog) between each observation, yielding a time interval between observations of 37.5 hours.

Observations are supposed to be of the altimetric type, i.e. observations of the instantaneous dynamical topography. This sea-surface height was directly proportional to the surface streamfunction which is a prognostic variable of the model. By definition, only the first layer is observed, but the whole state vector, all the 3 layers, was of course simulated. One challenge raised by the assimilation process, in addition to that of coping with the non-linearity, is to succeed in reconstructing flow evolution at great depth whilst disposing only of surface information.

The approach chosen is that of the so called "twin experiments". A reference experiment is performed and its finding noted so that the reference fields can be compared to the fields produced during the assimilation experiments. The assimilation experiment is performed using the "pseudo-observations" which are extracted (at the surface only) from the reference experiment. The assimilation will be successful if the flow converges with time towards the reference situation especially in flow at depth.

This test case has clearly many drawbacks due to the relatively simple model formulation together with the use of synthetic data and the simplistic domain configuration. It should be therefore regarded as a simple test case which has to be proved satisfactory before more realistic problems are even contemplated.

4.2. Specification of initial state and error covariance matrix

Following the strategy explained in section 3.4, the choice of the initial flow field and the corresponding error covariance matrix is made through the model simulation itself. Note that this is done once only, and the results can be (and have been) filed for use in different assimilation experiments. In the present study, data for the assimilation experiments is again simulated (i.e. synthetic data) but in a way unrelated to that in the above simulation.

To generate the flow fields we always consider a statistically steady regime in which the flow has been integrated for a sufficiently long time in order to avoid the transitional spin-up phase. In the present experiment, a long sequence of 8000 state vectors was generated of which the first 4000 vectors were discarded to avoid such transient effects. The remaining sequence of 4000 state vectors (6250 days) was reduced into a sequence of 1000 by retaining only 1 vector out of 4. This operation, known as decimation, was to reduce the calculation since successive states are quite similar. From this decimated sequence, we estimate the covariance matrix of the state vector and perform an EOFs analysis. Figure 1 plots the

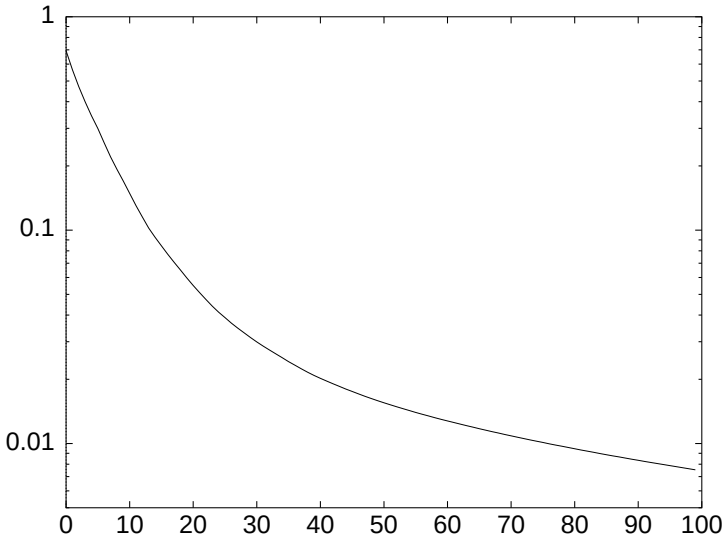

Figure 1: Relative error in square norm versus the number of retained EOFs

relative error in square L^2 norm versus the number of retained EOFs. It can be seen that the covariance matrix of $\mathbf{x}^t(t_0)$ is well approximated by a matrix of much lower rank. Indeed, the error decreases rapidly with the number of retained EOFs. For 51 retained EOFs, the relative error in square norm is already down to 1.55%. However, this error decreases much more slower in the range of higher numbers of retained EOFs (for 80 EOFs it is still 0.96%).

4.3. Results of assimilation experiments

We again generated a sequence of reference flow fields starting from an initial field totally unconnected to that used in the simulation of the above section. Practically, we take as initial filed, a computed field at a random time far removed from the time segment which was used in this simulation. Sequences of state vectors of length 200, corresponding to 312.5 days, were thus obtained. The first layer was taken as the (pseudo)-observation and was inputted into our filter algorithm (in some experiments, an observational noise was also added). The other layers were used as reference fields to evaluate the performance of the algorithm.

Figure 2 plots the assimilation results using the SEEK filter of rank 51 with no forgetting factor. It can be seen that the errors (relative in L^2 norm, layer by layer) decrease very rapidly for the first few steps (the initial errors are 95%, 99% and 101% which correspond to points far outside the figure). However, after several tens of steps, the errors start to increase again. This seems to result from to the fact that the filter under-evaluates the error and does not therefore make sufficient correction. Also, using a SEEK filter of rank only 51 induce an intrinsic error limitting the filter's performance.

In a second assimilation experiment we used the same setup as before except that the filter now had a forgetting factor of 0.5. The results are plotted in Figure 3. It can be seen

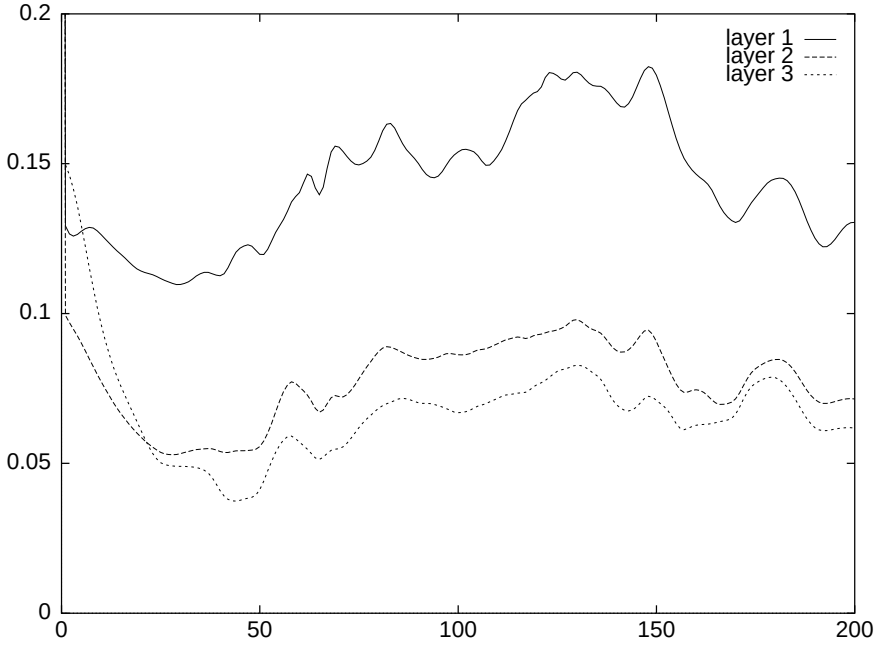

Figure 2: Relative assimilation errors for rank 51 with no forgetting factor

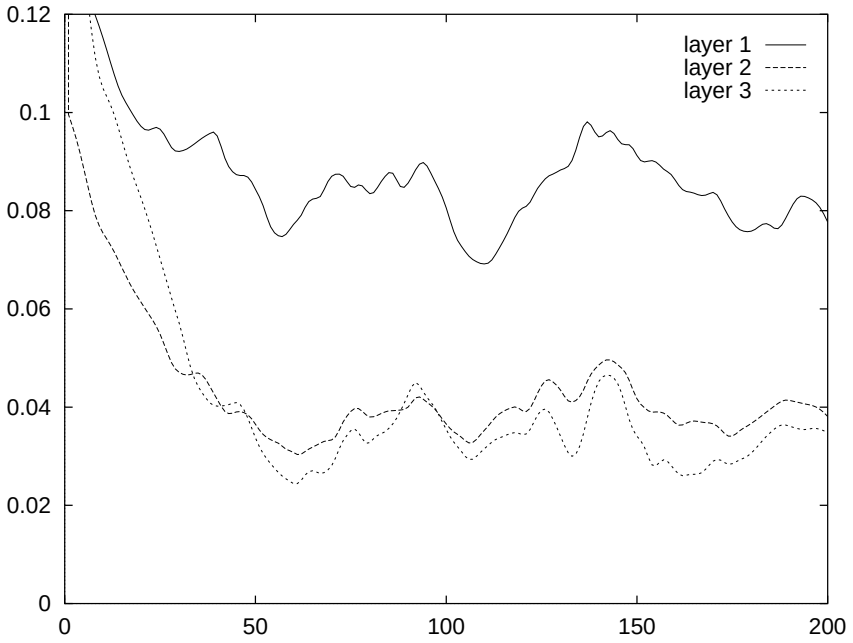

Figure 3: Relative assimilation errors for rank 51 with forgetting factor .5

that the introduction of the forgetting factor significantly enhances performance.

The chosen rank of the initial error covariance matrix has a direct effect on the performance of the algorithm. This is illustrated in figure 4, where the setup is the same as in figure 3 but the rank is increased to 80. One can see that the algorithm performs much better. Improvement in performance can also be achieved by a better initial state estimate. In figure 5, we have deliberately started the algorithm at a point $\mathbf{x}^a(t_0)$ much closer to the

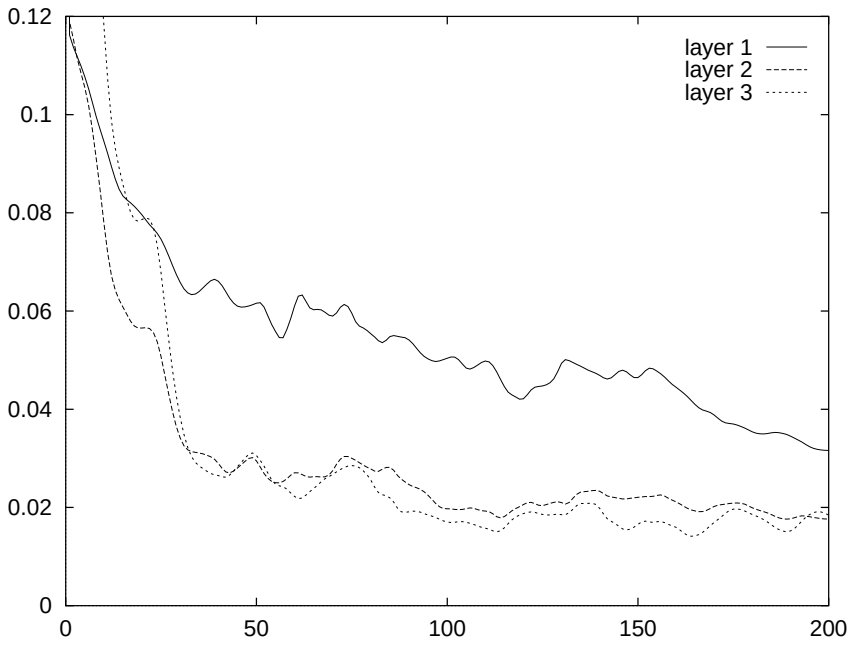

Figure 4: Relative assimilation errors for rank 80 with orgetting factor .5

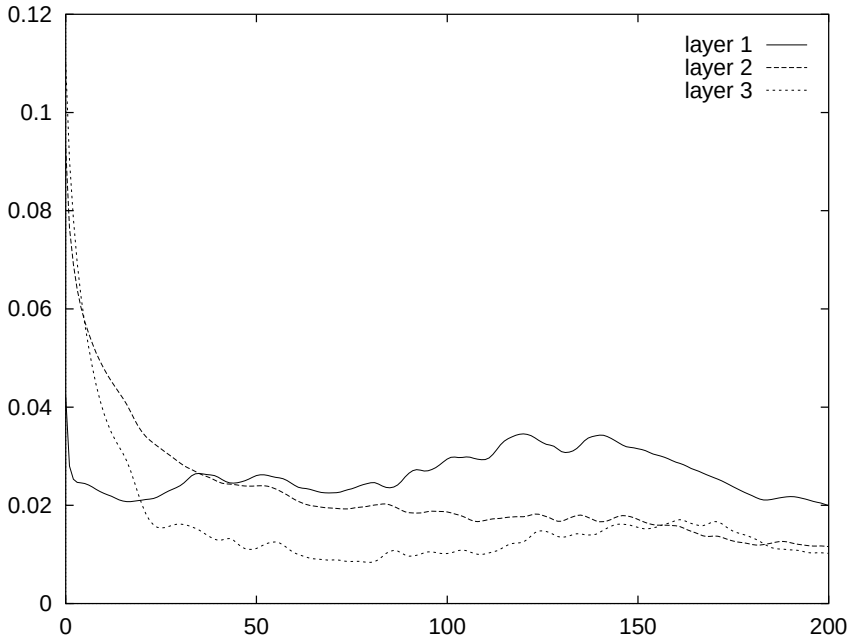

Figure 5: Relative assimilation errors for rank 80 with forgetting factor .5, special initial estimate

reference (the relative errors are 4.2% 9.2% and 10.9%). It can be seen from the figure that the error is significantly lower. However this gain of performance is attenuated over time and is not quite as appreciable near the end of the assimilation period.

Note that in the above simulation experiment, observations are simply taken as being components of the state vector corresponding to the first layer, *no noise being added* (whereas in the Kalman filter model an observational noise is assumed). But it would seem

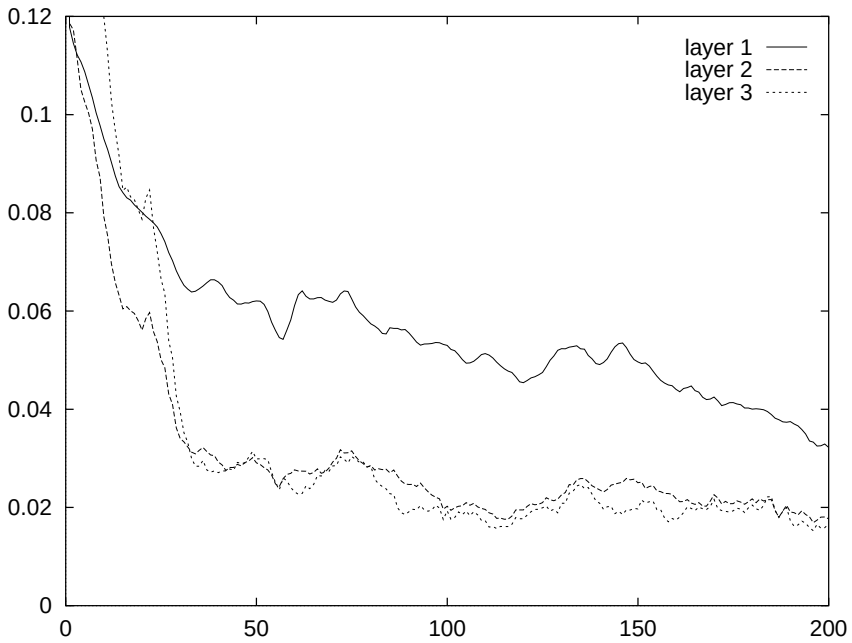

Figure 6: Relative assimilation errors for rank 80 with forgetting factor .5, obs. noise 30000

that observational noise has no appreciable effect on the performance of the filter. This is illustrated in figure 6, which corresponds to the same situation as that of figure 4, except that the observation is corrupted by a (spatially) white noise of standard deviation 30 000 (this corresponds to roughly 10% of the average standard deviation of the state variables). It can be seen that the curves in this figure are very similar to those in figure 4. They are in fact slightly lower, most, but not all, of the times.

5. Conclusion

A modified extended Kalman filter for the assimilation of oceanographic data into numerical models has been presented and a pilot implementation in a test case using altimeter (pseudo) data from a strongly nonlinear quasigeostrophic model is shown, which preliminary proves its feasibility. The basic feature for this "Singular Evolutive Extended Kalman" (SEEK) filter is the use of a low rank approximation to the error covariance matrix. This results in the introduction of a reduced size basis of statistical functions on which one relies in order to describing the evolution of the error covariance matrix. The basis evolves in time following the model evolution. This results from the nonlinearity of the model: on the one hand the gradient of the system transition operator depends on the system state so that the error propagation depends on the actual state of the ocean, on the other hand the reduced statistical basis was designed specifically to concentrate on the correction in those directions for which the error is amplified. Although EOFs analysis has been found to be a convenient and fairly effective way of initializing the reduced basis, this technique, due to its statistical nature, could only provide a good basis "on the average". Evolution with time is therefore

needed in order to adapt the basis to the particular state of the ocean under analysis at that time. We believe this evolutive property of our basis is crucial to keeping its dimension low. Fixed basis indeed has been considered in the litearature, but since it remains the same for possibly quite different ocean states, it is likely that a much higher number of basis functions would be required. Finally, the introduction of a memory effect ("forgetting factor") was also found to be quite beneficial and to greatly improve the performance of the algorithm.

The chosen test situation was that of an ocean box-model within which several mesoscale eddies were interacting in an intrinsically active turbulent process. The wind forcing was constant in time and space. The dissipation process was mainly through bottom friction but also by lateral friction. At this preliminary stage, the approximation was made that the model simulates perfectly the real ocean. Simply the observations, actually generated from a sequence of simulations of the same model, were assumed to be noisy. A simple white noise was chosen at a first approximation. Note that, because of the quasigeostrophic formulation, the observations of the ocean dynamical topography are exactly proportional to the corresponding streamfunction fields of the model. The functional dependence between the observations and the state vector is therefore trivial.

In test conditions, our SEEK filter was found to be fairly effective in monitoring the flow state and evolution disposing of surface-only pseudo-altimeter data. Further work will consider a more realistic situation, both from the model point of view (real coastal geometry, bathymetry, real wind forcing) and the observational one (using real data from the Topex/Poseidon satellite). However, this preliminary application was a necessary step before realistic applications and provides us with encouraging results as regard to that purpose.

Acknowledgments. This work was performed within the framework of the IDOPT project supported by INRIA and CNRS. The authors would also like to thank J. Blum, the project leader, with whom they have had many fruitful discussions.

References

Cane M. A. and R. J. Patton, 1984: A numerical model for low-frequency equatorial dynamics. *J. Phys. Oceanogr.*, 14,1853–1863.

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. Submitted to *J. Geophys. Res.*.

Dee D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Q. J. R. Meteorological. Soc.*, **117**, 365–384.

Evensen G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, 97, 17,905–17,924.

Evensen G., 1993: Open boundary conditions for the extended Kalman filter with a quasi-geostrophic ocean model. *J. Geophys. Res.*, 98 (C9), 16,529–16,546.

Evensen G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 (C5), 10,143–10,162.

Fu L.L., I. Fukumori and R. N. Miller, 1993: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific ocean. II: A linear, wind-driven model. *J. Phys. Oceanogr.*, 23, 2162–2181.

Fukumori I., J. Benveniste, C. Wunsch and D.B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady state smoother. *J. Phys. Oceanogr.*, 23, 1831–1855.

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 (C12), 6777–6793.

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), 25, 027–25,039.

Gauthier P., Courtier P. and P. Moll, 1993: Assimilation of simulated wind lidar data with a Kalman filter. *Mon. Wea. Rev.*, **121**, 1803–1820.

Gaspar P. and C. Wunsch, 1989: Estimates from altimeter data of barotropic Rossby waves in the northwestern Atlantic ocean. *J. Phys. Oceanogr.*, 19, 1821–1844.

Ghil, M., 1986: Sequential estimation and satellite data assimilation in meteorology and oceanography, in Y.K. Sasaki, T. Gat-Chen, L. Whie, M.M. Zaman, C. Ziegler, L.P. Chang and D. J. Rusk (editors), Variational Methods in Geosciences, Elsevier, Amsterdam, pp. 91–100.

Ghil, M. and P. Manalotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 23, 141–265.

Ghil M., Cohn S. E. and Dalcher A., 1982: Sequential estimation; data assimilation and initilization, in D. Williamson (Editor), The Interaction Between Objective Analysis and Initialization, *Publ. Meteorol. 127*, (Proc. 14th Standstead Seminar), McGill University, Montreal, pp.83–97.

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.

Hager, W. and Horowitz, L. L., 1976: Convergence and stability property of the discrete Riccati operator equation and the associated optimal control and filtering pronlems. *Siam J. Control and Optimization*, **14**, 2, 295–312

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

Holland, W. R., 1978: The role of mesosccale eddies in the general circulation of the ocean. Numerical experiments using a wind-driven quasi-geostrophic model. *J. Phys. Oceanogr.* **8**, 363–392.

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.*

Jazwinski, A. A., 1970: *Stochastic and Filtering Theory.* Mathematics in Sciences and Engineering series series **64**. New-York Academic Press.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D, J. Basic Eng.*, **82D**, 35–45.

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.

Miller R. N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.* , **51**, 1037–1056.

Pham D. T., J. Verron et M. C. Roubaud, 1995: Assimilation of Oceanic altimeter data using a quasi-geostrophic model and the extended Kalman filter. Rapport technique du projet IDOPT INRIA-CNRS, RT141, Octobre 1995.

Pedlosky J., 1987: *Geophysical fluid dynamics.* Springer-Verlag. New York Heidelberg.