ADAPTIVE KALMAN FILTERING : AN ESTIMATION OF OBSERVATIONAL AND SYSTEM NOISE STATISTICS

I Blanchet
SHOM/CMO/BRESM
Toulouse, France

Summary : The use of the Kalman filter as an assimilation scheme in oceanography is increasing. The Kalman filter gives the best linear unbiased estimator of the model state only if the first and second order statistics of the observational and system noise are correctly specified. If not, an adaptive filter can be used to estimate both the state vector and the noise statistics. Here, we present two different adaptive Kalman filter schemes. The adaptive algorithms have been used in a reduced space linear model of the tropical Pacific to estimate the system noise covariance matrix. The adaptive filter has also been implemented for serially correlated system or observational noise.

1. INTRODUCTION

Since oceanic initial conditions are poorly known and numerical ocean models are imperfect, data assimilation is an essential component for an ocean prediction system. Among several assimilation methodologies, the Kalman filter (Kalman 1960, hereafter KF) is an efficient approach to assimilate different types of observations and propagate the corrections to other model variables. Moreover, the KF provides estimates of the forecast and analysis errors. However, because of the computational burden, the use of such a sophisticated assimilation method for oceanographic observations, has been limited to very simple and coarse resolution models (e.g. Miller and Cane, 1989 ; Ghil and Malanotte-Rizzoli, 1991 ; Evensen, 1992).

Consequently, a number of simplifications have been proposed for applications of the filtering techniques to the atmosphere and the ocean (Todling and Cohn, 1994 ; Cohn and Todling, 1996). One of the solution we can mention is the « *reduced order* » filter which computes the forecast error covariances in a space of reduced dimension, using coarser resolution models (Fukumori and Malanotte-Rizzoli, 1995) or some other reduction operator (Cohn and Todling, 1996 ; Cane et al., 1996). Those schemes are briefly discussed in section 2.3.

The optimality of the KF depends on the assumption that we know perfectly the statistics of the observational and system error. Actually, misspecification of the statistics can lead to the divergence of the KF (Jazwinsky, 1970 ; Fitzgerald, 1971). Obviously, in the oceanic context, the statistics of the errors, particularly the system errors, are most often very hard to correctly specify. Moreover, the specification of the observational and system errors in the reduced space have to account for the truncation errors.

However, with the adaptive KF the observations can be used to improve the representation of the noise statistics as well as the estimation of the ocean state (Gelb 1974). In the present paper, we focus on two adaptive schemes that are implemented for a reduced order KF. We briefly discuss their application in a twin experiment performed in the tropical Pacific context.

2. THE EXTENDED KALMAN FILTER

2.1 General notations for the assimilation problem

The notations used in the present document are those recommended by Ide et al. (1997). The evolution of the ocean model from time t_{i-1} to time t_i is governed by the equation :

$$\mathbf{x}^f(t_i) = M_{i-1}\left[\mathbf{x}^a(t_{i-1})\right] \tag{1}$$

where the ocean forecast $\mathbf{x}^f$ and analysis $\mathbf{x}^a$ are column vectors of dimension n, where n is usually given by the number of prognostic variables of the numerical ocean model times its number of grid points or by the number of spectral modes. M_i is the dynamics operator, which is generally non linear.

The true ocean state $\mathbf{x}^t$ differs from (1) by random or systematic errors and its evolution is given by:

$$\mathbf{x}^t(t_i) = M_{i-1}\left[\mathbf{x}^t(t_{i-1})\right] + \eta_{i-1} \tag{2}$$

where η accounts for the modeling and forcing errors, hereafter referred to as system errors. Generally, the system noise is assumed to be white in time, to have zero mean with spatial covariance matrix $\mathbf{Q}$, which is positive definite :

$$E[\eta_i] = 0 \;\; ; \;\; E[\eta_i \; \eta_j^T] = \mathbf{Q}_i \, \delta_{ij} \tag{3}$$

The observations at time t_i are related to the model variables by the equation:

$$\mathbf{y}_i^o = H_i\left[\mathbf{x}^t(t_i)\right] + \varepsilon_i \tag{4}$$

where $\mathbf{y}_i^o$ is a column vector of dimension p_i equal to the number of observations available at t_i, and H is the observation operator. As M, this operator can also be non linear and it can depend on time when the observation network is not fixed. ε_i is the observational noise, which is assumed to be white, with zero mean and spatial covariance matrix **R**..

$$E[\varepsilon_i] = 0 \;;\; E[\varepsilon_i\, \varepsilon_j^T] = \mathbf{R}_i\, \delta_{ij} \tag{5}$$

The observational and system errors are assumed to be uncorrelated.

2.2 Principle and basic algorithm

The KF is a sequential-estimation approach to data assimilation. It gives an estimate $\mathbf{x}^a(k)$ of the true ocean state as a linear combination of the observations and the model forecast :

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i \mathbf{d}_i \tag{6}$$

where $\mathbf{d}_i$ is the innovation vector given by :

$$\mathbf{d}_i = \mathbf{y}_i^o - H_i\left[\mathbf{x}^f(t_i)\right] \tag{7}$$

K is the gain matrix which depends on the relative accuracy of the predictions and the observations :

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1} \tag{8}$$

In the Extended Kalman Filter (EKF, Gelb, 1974; Ghil and Malanotte-Rizzoli, 1991), the linearization $\mathbf{M} \equiv M'$ et $\mathbf{H} \equiv H'$ are introduced and the forecast errors covariance matrix is calculated as

$$\mathbf{P}^f(t_i) = \mathbf{M}_{i-1}\mathbf{P}^a(t_{i-1})\mathbf{M}_{i-1}^T + \mathbf{Q}_{i-1} \tag{9}$$

The forecast error covariance matrix appears as the sum of the estimated errors propagated by the model dynamics and the system errors (9). This is the most costly step of the KF since it requires $2\,n$ integrations of the tangent linear model **M** to obtain the term $\mathbf{M}\,\mathbf{P}^a\,\mathbf{M}^T$.

The analysis errors covariance matrix is given by

$$\mathbf{P}^a(t_i) = [\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]\mathbf{P}^f(t_i)[\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]^T + \mathbf{K}_i\mathbf{R}_i\mathbf{K}_i^T \tag{10}$$

which for the Kalman gain (8) yields the simplified equation

$$\mathbf{P}^a(t_i) = [\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]\mathbf{P}^f(t_i) \tag{11}$$

From a numerical point of view, (10) should be preferred to (11) to make sure that $\mathbf{P}^a$ is symmetric positive definite, even though the calculation is more costly. When simplifications are introduced in the calculation of **K** then only (10) holds.

2.3 Order reduction

2.3.1 *Principle of the reduced order Kalman Filter*

We assume that the simplification operator is denoted by **S** (r,n) where r is the dimension of the reduced space, with $r << n$. For the reduced order EKF (ROEKF), the analysis is given by :

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{S}^{-I}\mathbf{Kr}_i\mathbf{d}_i \tag{12}$$

Kr (r,p) is the reduced gain matrix, which writes :

$$\mathbf{Kr}_i = \mathbf{Pr}^f(t_i)\mathbf{Hr}_i^T\left[\mathbf{Hr}_i\mathbf{Pr}^f(t_i)\mathbf{Hr}_i^T + \mathbf{Rr}_i\right]^{-1} \tag{13}$$

the forecast errors covariance matrix in the reduced space being given by $\mathbf{Pr}^f$:

$$\mathbf{Pr}^f(t_i) = \mathbf{Mr}_{i-1}\mathbf{Pr}^a(t_{i-1})\mathbf{Mr}_{i-1}^T + \mathbf{Qr}_{i-1} \tag{14}$$

The analysis errors covariance matrix is given by

$$\mathbf{Pr}^a = [\mathbf{I} - \mathbf{KrHr}]\mathbf{Pr}^f \tag{15}$$

with $\mathbf{Hr} = \mathbf{H}\ \mathbf{S}^{-I}$ and $\mathbf{Mr} = \mathbf{S}\ \mathbf{M}\ \mathbf{S}^{-I}$. The observational and system error matrix, **Rr** and **Qr** respectively, are defined in the reduced space taking into account the truncation errors.

2.3.2 *Defining the simplification operator*

A first approach to reduce the dimension of the problem consist in computing the forecast error-covariances with a coarser-resolution model than the model used to forecast the state itself (LeMoyne and Alvarez 1991 ; Hoang et al., 1995). Using such a simplification operator and the steady state limit of the forecast error covariance matrix Fukumori and Malanotte-Rizzoli (1995) have developed the first application of an approximate KF for a non linear primitive equation ocean model.

To avoid the drawbacks of coarse resolution, Cane et al. (1996) have adopted a different approach : using a multivariate empirical orthogonal function (EOF) analysis, they reduced the state space for the forecast covariance update to a small set of basis functions, which nonetheless represented all of the significant structures that were predicted by the model. The procedure was shown to lead to a substantial saving without any loss of accuracy compared to a grid point KF.

Cohn and Todling (1996) have proposed two suboptimal *Kalman-like* filters for reducing the computational cost of the KF algorithm. The first scheme, the Partial Singular value decomposition Filter (PSF), is based on the most dominant singular modes of the tangent linear propagator. This scheme assumes that most of the propagated analysis error covariances is due to a small collection of the model's most rapidly growing singular modes. The second scheme, the Partial Eigen decomposition Filter (PEF), is based on the most dominant eigenmodes of the propagated analysis error covariance matrix. The propagated analysis error covariance is then replaced by the leading part of its decomposition. An experiment with unstable dynamics was performed with these two schemes. otherwise it diverges.

The compared theoretical properties of these reduced order filters will not be detailed here. But, it can be shown that all the unstable modes (and neutral modes) of **M** should be included in the sub-space defined by **S** (Cohn and Todling, 1996 ; Hoang et al., 1996) to insure the stability of the ROEKF. Thus, detectability (observability of unstable modes) and stabilizability (controllability of unstable modes) issues should constrain the choice of the reduction space for the implementation of the ROEKF (e.g. Kucera, 1972).

3. ADAPTIVE KALMAN FILTERING

3.1 Principle and basic algorithm

In principle, an adaptive filter can estimate both the system and the observational errors. However, adaptive algorithms that try to update both the observational noise and the system noise are not robust, since it is not easy to distinguish between errors in **Q** and **R** (Groutage et al. 1987 ; Maybeck 1982 ; Daley 1992-b). Since the observational errors are generally much better known than the system errors, we will focus on the estimation of **Q**.

3.1.1 *Definition of the maximum likelihood function*

The adaptive schemes that are described below make use of the innovation sequence covariance to estimate the noise statistics (Maybeck, 1982 ; Dee, 1995 ; Blanchet et al., 1997). The adaptive Kalman filters are implemented in the reduced space since their computational cost in the full space is prohibitive.

Let us define $\underline{\alpha}$ as the M-dimensional vector of the unknown noise parameters describing $\mathbf{Q_r}$. $\underline{\alpha}$ is estimated by maximizing the conditional probability density function $p(D_N(t_k) \mid D(t_{k-N}), \underline{\alpha})$ where $D_N(k)$ denotes the innovation sequence during the last N steps $(\mathbf{d}(t_{k-N+1}), \ldots, \mathbf{d}(t_k))$ and $D(t_i)$ denotes the innovation sequence during the first i steps $(\mathbf{d}(t_1), \ldots, \mathbf{d}(t_i))$. We assume that the system noise statistics are approximately constant on the assimilation window defined by the N steps. This window has been defined in order to avoid the divergence of the estimation of $\mathbf{Q_r}$ that could be due to a »bad » observation. This might happen when only one innovation sample is used as it was the case in Dee (1995). Dee pointed out that single-sample covariance estimation is only reasonable if the number of observations is more than two orders of magnitude larger than the number of parameters to be estimated. He also showed that the simultaneous estimation of several parameters led to a large variance of the single-sample estimates. Since the number of observations is rather small in the oceanographic context, we would recommend to use a sufficiently large window.

Using repeated applications of Bayes' rule, we can write

$$p\left(D_N(t_k)\middle|\, D(t_{k-N}),\underline{\alpha}\right)=\prod_{i=k-N+1}^{k} p\left(d_i\middle| D(t_{i-1}),\underline{\alpha}\right) \tag{16}$$

Assuming that $\mathbf{d}$ is a gaussian variable with zero mean and covariance matrix $C_0(\underline{\alpha})$, we have :

$$p\left(d_i\middle| D(t_{i-1}),\underline{\alpha}\right)=(2\pi)^{-p/2}\left(\det\left(C_0(t_i,\underline{\alpha})\right)\right)^{-1/2}\exp\left[-\frac{1}{2}d_i^T C_0^{-1}(t_i,\underline{\alpha})d_i\right] \tag{17}$$

The maximum likelihood estimator obtained by maximizing (16) can be obtained as well by minimizing the following functional $J(\underline{\alpha})$

$$J(\underline{\alpha})=\sum_{i=k-N+1}^{k}\left[\ln\left(\det\left(C_0(t_i,\underline{\alpha})\right)\right)+d_i^T C_0^{-1}(t_i,\underline{\alpha})d_i\right] \tag{18}$$

The innovation covariance matrix C_0 can be related to the system noise covariance matrix $\mathbf{Q_r}$:

$$C_0(t_i,\underline{\alpha})=H_{r_i}P_{r_i}^f H_{r_i}^T+R_{r_i}=H_{r_i}M_{r_{i-1}}P_{r_{i-1}}^a M_{r_{i-1}}^T H_{r_i}^T+H_{r_i}Q_{r_{i-1}}H_{r_i}^T+R_{r_i} \tag{19}$$

With such an adaptive scheme, a parameter estimate is produced only every N samples, yielding slower initial convergence to a good estimate, specially if the initial guess is far from the true value. Moreover, if we admit that the parameters can vary slowly, then it introduces an inherent lag in estimating a parameter change. The effects can be alleviated by changing the size of the interval from N to a divisor of N. This effectively reduces the lag in responding to measurement information that the parameter values are different than previously estimated.

3.1.2 *Maybeck estimator*

Since it is assumed that $\mathbf{Q_r}(\underline{\alpha})$ is approximately constant over a window of N assimilation steps and that $\mathbf{R_r}$ is known, Maybeck (1982) showed that after introducing some simplifications an explicit form for $\mathbf{Q_r}(\underline{\alpha})$ can be derived from (18) :

$$\hat{Q}r(t_k)=\frac{1}{N}\sum_{i=k-N+1}^{k}\left\{Kr(t_i)d_i d_i^T Kr^T(t_i)-\left[Mr_{i-1}Pr^a(t_{i-1})Mr_{i-1}^T-Pr^a(t_i)\right]\right\} \tag{20}$$

At each time step $k \geq N$, the adaptive filter requires the storage of the last N innovation samples, and the estimation of (20). The initial value $\hat{Q}r(0)$ must be specified. To guarantee the semi-definite positiveness of $\hat{Q}r$: we calculate its eigenvalues at each time step and reset the negative eigenvalues to zero. This formulation has a very low cost and is easy to implement.

Maybeck (1982) pointed out the different factors which can influence the choice of the window length. Even though large N yields less susceptibility to « bad » measurements and reduces the high frequency oscillations of the estimator, we found that the convergence of the estimator was really slowed when N was too large and we decided for a very small interval.

It has been shown (Blanchet et al., 1997), that it is necessary to drastically reduce the number of parameters to be estimated, so that they are properly constrained by the assimilation. The number of terms of the covariance matrix Qr, that are allowed to vary, is to be determined according to the number of independent observations that are available, i.e. according to the rank of Hr (Blanchet, 1997). Actually, every term of the covariance matrix Qr can be estimated if and only if $p \geq r$ and Hr is of full rank, i.e. of rank r (Mehra, 1970 ; see also Hoang et al. 1994).

3.1.3 *Maximum likelihood estimator*

The second approach solves directly the minimization problem (18) by using whichever minimization algorithm is the most efficient. We might like to use a conjugated gradient algorithm in which case the minimization scheme is based on the gradient of the functional J which is given by :

$$\frac{\partial J}{\partial \alpha_j}=\sum_{i=k-N+1}^{k}\left\{\mathrm{Trace}\left[\left(C_0^{-1}(t_i)-C_0^{-1}(t_i)d_i d_i^T C_0^{-1}(t_i)\right)\frac{\partial C_0(t_i)}{\partial \alpha_j}\right]-2\frac{\partial x^f(t_i)}{\partial \alpha_j}^T H_i^T C_0^{-1}(t_i)d_i\right\} \tag{21}$$

with
$$\frac{\partial C_0}{\partial \alpha_j}=Hr\frac{\partial Pr^f}{\partial \alpha_j}Hr^T+\frac{\partial Rr}{\partial \alpha_j} \tag{22}$$

$\frac{\partial Pr^f}{\partial \alpha_j}(t_i)$ is evaluated by deriving the equations (13), (14) et (15) and depending on $Qr(\underline{\alpha})$ and eventually $Rr(\underline{\alpha})$:

$$\frac{\partial Kr}{\partial \alpha_j}=\frac{\partial Pr^f}{\partial \alpha_j}Hr^T C_0^{-1}-Pr^f Hr^T C_0^{-1}\frac{\partial C_0}{\partial \alpha_j}C_0^{-1} \tag{23}$$

$$\frac{\partial Pr^f}{\partial \alpha_j}(t_i)=Mr_{i-1}\frac{\partial Pr^a}{\partial \alpha_j}(t_{i-1})Mr_{i-1}^T+\frac{\partial Qr}{\partial \alpha_j} \tag{24}$$

$$\frac{\partial Pr^a}{\partial \alpha_j} = [I - KrHr]\frac{\partial Pr^f}{\partial \alpha_j} - \frac{\partial Kr}{\partial \alpha_j} HrPr^f \tag{25}$$

By deriving (1), and using the tangent linear operator, one writes :

$$\frac{\partial x^f(t_i)}{\partial \alpha_j} = M_{i-1}\frac{\partial x^a(t_{i-1})}{\partial \alpha_j} \tag{26}$$

$\frac{\partial x^a(t_i)}{\partial \alpha_j}$ can be related to $\frac{\partial x^a(t_{i-1})}{\partial \alpha_j}$ by deriving (12) and making use of (7) and (26)

$$\frac{\partial x^a(t_i)}{\partial \alpha_j} = \left[I - S^{-I}Kr_iH_i\right]\frac{\partial x^f(t_i)}{\partial \alpha_j} + S^{-I}\frac{\partial Kr_i}{\partial \alpha_j}d_i = \left[I - S^{-I}Kr_iH_i\right]M_{i-1}\frac{\partial x^a(t_{i-1})}{\partial \alpha_j} + S^{-I}\frac{\partial Kr_i}{\partial \alpha_j}d_i \tag{27}$$

We now note A the filter transition matrix :

$$A_{i-1} = \left[I - S^{-I}Kr_i H_i\right]M_{i-1} \tag{28}$$

Using (26) and (27), the last term in (21) can be written :

$$\sum_{i=k-N+1}^{k} -2\left(\frac{\partial x^f(t_i)}{\partial \alpha_j}\right)^T H_i^T C_0^{-1}(t_i)d_i = \sum_{i=k-N+1}^{k}\left\{\sum_{\tau=t_{i-T}}^{t_i} d^T(\tau-1)\left(\frac{\partial Kr(\tau-1)}{\partial \alpha_j}\right)^T S\varphi_i^*(\tau)\right\} \tag{29}$$

For each i, φ is the solution of the following adjoint equations :

$$\varphi_i^*(t_i) = -2M_{i-1}^T H_i^T \Sigma_i^{\prime} d_i \tag{30}$$

$$\varphi_i^*(\tau-1) = A^*(\tau-2)\varphi_i^*(\tau) \quad , \text{ for } t_{i-T+1} \le \tau \le t_i \tag{31}$$

with

$$A_{i-1}^* = M_{i-1}^*\left[I - H_i^T Kr_i^T S\right] \tag{32}$$

When the adjoint of the numerical model M^* is not available, the gradient (21) is approximated by using the matrix Ar^* and integrating the adjoint equations in the reduced space :

$$\varphi r_i^*(t_i) = -2Mr_{i-1}^T Hr_i^T C_0^{-1}(t_i)d_i \tag{33}$$

$$\varphi r_i^*(\tau-1) = Ar^*(\tau-2)\varphi r_i^*(\tau) \quad , \text{ for } t_{i-T+1} \le \tau \le t_i \tag{34}$$

with

$$Ar_{i-1}^* = Mr_{i-1}^T\left[I - Hr_i^T Kr_i^T\right] \tag{35}$$

Thus, (29) is approximated by replacing $\left(S\varphi_i^*(\tau)\right)$ with $\varphi r_i^*(\tau)$. In such a case, we will have to build the matrix Mr in the reduced space to be able to explicit its adjoint Mr^T. Building Mr is an expensive task.

T is the number of backward integrations that are required to evaluate the gradient. The norm of the transition matrix $\mathbf{A_r}$ being strictly smaller than 1 in order to ensure the filter stability, the gradient can be obtained in good approximation with a relatively small value of T.

The forecast errors will remain bounded, if the system is detectable in the full space and if all the unstable modes of the numerical model are included in the reduced space. But, the stability of the adaptive Kalman Filter also relies on the convergence of the estimation of $\mathbf{Q_r}$, which means that we have to choose carefully the parameters that we want to estimate. Actually, all the elements of $\mathbf{Q_r}$ can be uniquely determined if and only if $p \geq r$ and $\mathbf{H_r}$ is of rank r (Mehra, 1970).

Using this algorithm, we could possibly estimate $\mathbf{R_r}$ as well, even though we would not be sure that the simultaneous estimation of $\mathbf{R_r}$ and $\mathbf{Q_r}$ is well behaved. Using this adaptive algorithm makes it easy to take into account and estimate an auto-correlation of the observational or system noise (Blanchet, 1997). This is simply done by augmenting the vector of unknown parameters $\underline{\alpha}$ and modifying the expression of $\mathbf{C_0}$ according to the Kalman Filter equations that can be derived for auto-correlated noises (Daley, 1992-a).

3.2 Practical aspects of implementation

The most expensive step of the ROEKF is the construction of the model transition matrix $\mathbf{M_r}$ in the reduced space, thus this calculation should not be repeated too often (Fukumori and Malanotte-Rizzoli, 1995 ; Cane et al., 1996). However, because of the limited validity of the tangent linear approximation for the EKF, the transition matrix has to be updated periodically. The memory requirement for the simplification matrix S can also be quite huge, unless the projection onto the reduced space is defined as an operator instead of a matrix.

Adaptive KF can simultaneously estimate the state and the system error statistics, but these algorithms require an even greater amount of computation and storage than the ROEKF. The maximum likelihood estimator has not yet been used for huge systems and undoubtedly some work has to be undertaken to lower its computational requirement, in order to make its cost acceptable. For example, the computational cost of the gradient calculation used to solve the minimization problem in section 3.1.3 is directly proportional to the number of unknown parameters. So the cost of the maximum likelihood estimator could be reduced by using a minimization algorithm that does not need an explicit formulation for the gradient.

Actually, a way of reducing drastically that cost would be to implement an optimal interpolation scheme (OI : Gandin, 1963) with the maximum likelihood estimator (18) for updating directly the forecast error covariance matrix Pr^f(instead of Qr). Thus, we can avoid most of the expensive calculation steps in the adaptive assimilation scheme. The counterpart is that one has to define properly the structure of the forecast errors, which strongly depends on the dynamics and the system errors (14), since very few parameters can be estimated by the adaptive scheme. The parameterization of Pr^f is one of the bigger task to be undertaken but it is actually shared by all the assimilation schemes.

3.3 An application of the adaptive Kalman Filter for an ocean model

The application of adaptive filters to meteorology and oceanography has been very limited. The adaptive KF discussed in section 3.1.2 and 3.1.3 have been applied in the context of tropical oceanography (Blanchet et al., 1997). In this study, we assess the ability of the two adaptive assimilation schemes at estimating an unbiased, stationary system noise. The adaptive algorithms are implemented in a reduced space linear model for the tropical Pacific, as described in Cane et al. (1996). Using a twin experiment approach, the algorithms are compared by assimilating sea level data at fixed locations mimicking the tropical Pacific tide gauges network.

It is shown that the description of the system error covariance matrix requires too many parameters for the adaptive problem to be well-posed. However, the adaptive procedures are efficient if the number of noise parameters is dramatically reduced and their performance is shown to be closed to optimal, i.e. based on the true system noise covariance. The two adaptive procedures that of Maybeck and the maximum likelihood estimator give comparable results in terms of the analysis and forecast (figure 1, Table 1). They both are very efficient at whitening the innovation sequence, meaning that all information has been extracted from the observations. However, the structure of the matrices estimated in each case is quite different, meaning that the matrix giving the optimal gain might not be unique when the number of observations is smaller than the dimension of the state vector (Mehra, 1970). The ML estimator proved to be more costly than the Maybeck one.

These results show that it is worthwhile to concentrate on estimating system errors statistics to improve the forecast. For non linear dynamics, Malanotte-Rizzoli et al. (1996) also found out in that the most critical point was the specification of the system noise since they obtained much better results, even with the steady state filter, by specifying full covariances instead of spatially uncorrelated noise as was the case in their previous study (Fukumori and Malanotte-Rizzoli, 1995).

In Blanchet (1997), the estimation of biased or auto-correlated statistics has also been addressed. The estimation of auto-correlation parameters in the system or observational noise can only be achieved with the maximum likelihood estimator.

	observations				true state (EOFs space)			
	correlation		rms (cm)		correlation		rms x 10^{-4}	
	forecast	analysis	forecast	analysis	forecast	analysis	forecast	analysis
unfiltered	0.68		**6.21**		0.66		**7.48**	
Maybeck	0.77	0.95	**4.96**	2.11	0.77	0.80	5.81	**5.18**
TKF	0.79	0.94	**4.72**	2.37	0.81	0.86	5.05	**4.05**

Table 1 :performance of the Maybeck algorithm when estimating 16 parameters. Rms differences and correlations are given for the observations at the 34 stations (left) and for the true state of the ocean (right) for both the forecasts and the analysis. For reference, values are given for the unfiltered run (no assimilation) and the KF run with true noise statistics.

4. CONCLUSION

The key issue of the assimilation is the specification of the observation error and forecast error covariance matrices. There is no doubt that the EKF is of great interest to help us to find out how the analysis errors are propagated. On the other hand, we showed that using the EKF for high dimensional models implies some simplifications like order reduction and time-invariant linearization of model dynamics.

The determination of the forecast error covariance matrix depends not only on the error propagation but also on the specification of the system errors. Building the system error covariance matrix requires a good knowledge of the physics that has been neglected in the model. In the reduced space, the system errors also depends on the dynamical coupling terms between the retained and the discarded modes. Depending on the kind of simplification that is used, this might be very hard to infer. The observation errors consist of the actual measurement error and the representativity error of the observations errors in the reduced space. The latter will also depends on the way we define the reduced space and may introduce auto-correlated errors.

Actually, the necessary condition on a good knowledge of the noise statistics is never satisfied in practice. The use of wrong a priori statistics will lead to erroneous estimates. Thus, it is compulsory to get improved noise statistics compared to the very simple choices that are usually made. We doubt that « hand tuning » the representation of these errors is an appropriate tool for a huge problem and the implementation of an adaptive scheme might prove very useful. One can imagine a prediction system, where the analysis will be solved on a limited time window by OI or a variational approach, while the adaptive Kalman filter will evaluate the noise statistics. As already mentioned the maximum likelihood estimator will also allow the estimation of the statistics of auto-correlated noises.

An other promising approach which has not been discussed here, is the so called nonlinear adaptive filter which estimate directly the gain matrix instead of the noise statistics in order to improve the state forecasts (Hoang et al., 1996).

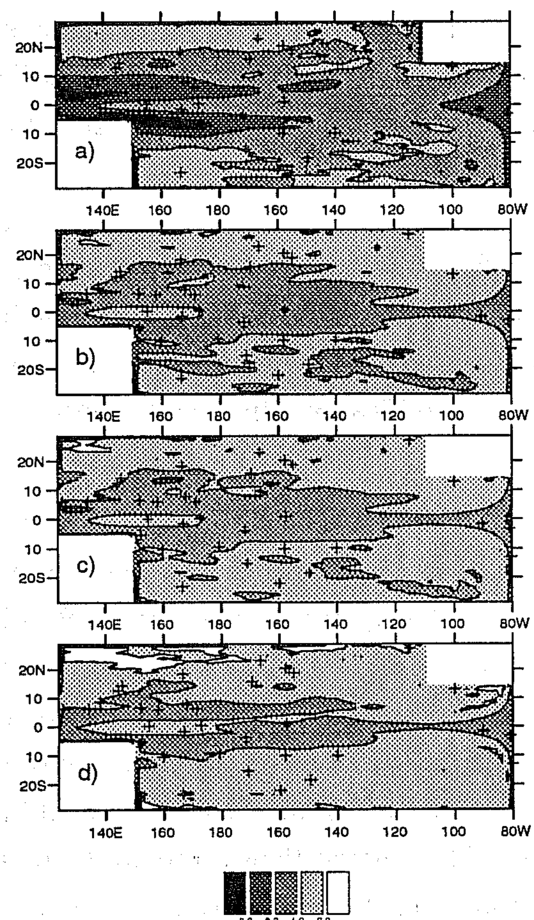

Figure 1: Rms difference between true and predicted sea level using (a) the a priori system noise covariance matrix (b) that estimated by Maybeck adaptive procedure (c) that estimated by the ML function and (d) the true one. Contour interval : 2 cm.

5. REFERENCES

Blanchet, I., C. Frankignoul, and M. A. Cane, 1997 : A comparison of adaptive Kalman filters for a tropical Pacific ocean model. *Mon. Wea. Rev.*, **125**, 40-58.

Blanchet, I., 1997: Assimilation à l'aide d'un filtre de Kalman adaptatif dans un modèle d'océan. Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris, France.

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 space Kalman Filter. *J. Geophys. Res.* , **101**, 22599-22618.

Cohn, E., and R. Todling, 1996 : Approximate data assimilation schemes for stable and unstable dynamics. *J. Met. Soc. Jpn*, **74**, 63-75.

Daley, R., 1992-a : The Effect of Serially Correlated Observation and Model Error on Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **120**, 164-177.

Daley, R., 1992-b : The lagged innovation covariance : A performance diagnostic for

Dee, D. P., 1995 : On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128-1145.

Evensen, G., 1992 : Using the extended Kalman Filter with the multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905-17924.

Fitzgerald, R. J., 1971 : Divergence of the Kalman Filter. *IEEE Trans. Automat. Contr.*, **AC-16**, 736-747.

Fukumori, I., and P. Malanotte-Rizzoli, 1995 : An approximate Kalman filter for ocean data assimilation : An example with an idealized Gulf Stream mode. *J. Geophys. Res.*, **100**, 6777-6793.

Gandin, L. S., 1963 : "Objective Analysis of Meteorological Fields." Gidrometeorol. Izd. Leningrad (in Russian), (English Translation by Israel Program of Scientific Translations, Jerusalem, 1965).Gelb, A., 1974 : "Applied Optimal Estimation." MIT Press, Cambridge, Massachusetts, pp 374.

Ghil, M., and P. Malanotte-Rizzoli, 1991 : Data assimilation in meteorology and oceanography. *Advances in Geophysics* **33**, 141-266.

Groutage, F. D., R. G. Jacquot, and R. L. Kirlin, 1987 : Techniques for adaptive state estimation through the utilization of robust smoothing. In *Control and Dynamic system* **23**, 273-308.

Hoang, S., P. De Mey, and O. Talagrand, 1994 : A simple algorithm of stochastic approximation type for system parameter and state estimation. In *Proc. 33rd IEEE Conf. on Decision and Control*, Lake Buena Vista, FL, pp 747-752.

Hoang, S., P. De Mey, O. Talagrand, and R. Baraille, 1995 : Assimilation of altimeter data in a multilayer quasi-geostrophic ocean model by simple nonlinear adaptive filter. In *Proc. International Conf. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, March, Vol. **2**, 521-527.

Hoang, S., R. Baraille, O. Talagrand, P. De Mey, and X. Carton, 1996 : On the design of a stable adaptive filter. In *Proc. 35th IEEE Conf. on Decision and Control*, Kobe, Japan.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997 : Unified Notation for Data Assimilation : Operational, Sequential and Variational. *J. Met. Soc. Japan*, Special Issue on "Data Assimilation in Meteorology and Oceanography : Theory and Practice.", **75**, 181-189.

Jazwinsky, A. H., 1970 : "Stochastic Processes and Filtering Theory." Academic Press, New York, 376 pp.

Kalman, R. E., 1960 : A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82D**, 35-45. *J. Phys. Oceanogr.* **23**, 2541-2566.

Kucera, V., 1972 : The discrete Riccati Equation of Optimal Control. *Kybernetica-8*, **5**, 430-447.

LeMoyne, H., and J. C. Alvarez, 1991 : Analysis of dynamic data assimilation for atmospheric phenomena. Effect of the model order. *Atmosfera* **4**, 145-164.

Malanotte-Rizzoli, P., I. Fukumori, and R. E. Young, 1996 : A Methodology for the Construction of a Hierarchy of Kalman Filters for Nonlinear Primitive Equation Models. In *Modern Approaches to Data Assimilation in Ocean Modeling*. P. Malanotte-Rizzoli editor., Elsevier, Amsterdam, 455 pp.

Maybeck, P. S., 1982 : "Stochastic models, estimation, and control." Vol. 2. Academic Press, N.Y., 423 pp.

Mehra, R. K., 1970 : On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Automat. Contr.* **AC-15**, 175-184.

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

Todling, R., and E. Cohn, 1994 : Suboptimal Schemes for Atmospheric Data, Assimilation Based on the Kalman Filter. *Mon. Wea. Rev* . **122**, 2530-2557.