We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

5,000
Open access books available

125,000
International authors and editors

140M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX INDEXED — CLARIVATE ANALYTICS

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Adaptive Filter as Efficient Tool for Data Assimilation under Uncertainties

Hong Son Hoang and Remy Baraille

Abstract

In this contribution, the problem of data assimilation as state estimation for dynamical systems under uncertainties is addressed. This emphasize is put on high-dimensional systems context. Major difficulties in the design of data assimilation algorithms is a concern for computational resources (computational power and memory) and uncertainties (system parameters, statistics of model, and observational errors). The idea of the adaptive filter will be given in detail to see how it is possible to overcome uncertainties as well as to explain the main principle and tools for implementation of the adaptive filter for complex dynamical systems. Simple numerical examples are given to illustrate the principal differences of the AF with the Kalman filter and other methods. The simulation results are presented to compare the performance of the adaptive filter with the Kalman filter.

Keywords: adaptive filter, innovation process, minimum prediction error, simultaneous perturbation stochastic approximation, filter stability

1. Introduction

In this chapter, the adaptive filter (AF) is considered as a computational device that yields estimates of the system state by minimizing recursively (in time) the error between the predicted output of the device and its observed signal in real time. As the main objective of the AF is to produce estimates of the state in high-dimensional systems (HdSs), we shall focus the attention on the mathematical form of the AF in a state-space form as that used in the Kalman filter (KF) [1]. In this chapter, the HdS is referred to as a system whose state dimension is of order $O(10^7) - O(10^8)$.

The assimilation problem in this chapter is formulated as a standard filtering problem. For simplicity, let the dynamical system be described by the equation

$$x(t+1) = \Phi x(t) + w(t), x(0) = x_0, t = 0, 1, \ldots \tag{1}$$

where $x(t)$ is the system state at the t time instant. At each time instant t, we are given the observation for the system output

$$z(t+1) = Hx(t+1) + v(t+1), t = 0, 1, \ldots \tag{2}$$

In (1) and (2), $w(t)$ is the model error (ME), $v(t)$ is the observation error (ObE), and Φ represents the system dynamics. In general, the system (1) and (2) may be nonlinear with $\Phi x = f(x)$, $Hx = h(x)$. The filtering problem for a partial observed dynamical system (1) and (2) is to obtain the best possible estimate for the state $x(t)$ at each instant t, given the set of observations $Z(1:t) = [z(1), ..., z(t)]$.

There exist different techniques to solve estimation problems. The simplest approach is related to linear estimator [2], since it requires only first two moments. Linear estimation is frequently used in practice when there is a limitation in computational complexity. Among others, the widely used methods are maximum likelihood, least squares, method of moments, the Bayesian estimation, minimum mean square error (MMSE), etc. For more details, see [3].

There are limitations of optimal filters. In practice, the difficulties are numerous: the statistics of signals which may not be available or cannot be accurately estimated; there may not be available time for statistical estimation (real-time); the signals and systems may be non-stationary; memory required and computational load may be prohibitive. All these difficulties become insurmountable, especially for HdSs.

In order to deal with real-time applications, the AFs appear to be a valuable tool in solving estimation problems when there is no time for statistical estimation and when we are dealing with non-stationary signals and/or systems environment. They can operate satisfactorily in unknown and possibly time-varying environments without user intervention. They improve their performance during operation by learning statistics from current signal observations. Finally, they can track variations in the signal operating environment [4].

It is well-known that the MMSE estimator in the class of Borel measurable (with respect to (wrt) $Z(1:t)$) functions is given by the conditional mean

$$\hat{x}(t) = E[x(t)/Z(1:t)] \tag{3}$$

Under standard conditions, related to the noise sequences $w(t), v(t)$ (Gaussian i.i.d.—identically independent (temporal) distributed), the estimate (3) $\hat{x}(t)$ for $x(t)$ can be obtained from the KF in the recursive form

$$\hat{x}(t+1) = \hat{x}(t+1/t) + K\zeta(t+1), \tag{4}$$

$$\hat{x}(t+1/t) = \Phi\hat{x}(t), \zeta(t+1) = z(t+1) - \hat{x}(t+1/t) \tag{5}$$

$$K(t) = M(t)H^T [HM(t)H^T + R]^{-1} \tag{6}$$

$$M(t+1) = \Phi P(t)\Phi^T + Q \tag{7}$$

$$P(t) = [I - K(t)H]M(t)[I - K(t)H]^T + K(t)RK^T(t) \tag{8}$$

In (4)–(8), Q, R are the covariance matrix for $w(t)$ and $v(t)$, respectively. One sees that $K = K(M)$—the gain matrix, is a function of $M := M(t+1)$—the error covariance matrix (ECM) for the state prediction error (PE) $e(t+1/t)$ which is defined as $e(t+1/t) := x(t+1) - \hat{x}(t+1)$, and $\zeta(t+1)$ is known as innovation vector. Note from (7) and (8) that the ECM $M(t+1)$ can be found by solving the matrix nonlinear Algebraic Riccati equation (ARE). Generally speaking, a solution of the ARE is not unique. Conditions must be introduced for ensuring an existence of a unique non-negative definite solution [5]. It is remarkable that the ECMs P, M in (7) and (8) do not depend on observations; therefore, they can be computed in advance, offline, given the system matrices and the noise covariances. The same remark is valid for the gain matrix K in (6). In contrast, the gain in the AF is observation-dependent [6] (see Section 2).

Under the most favorable conditions (perfect knowledge of all system parameters and noise statistics), for a dynamical system with dimension of order 10^7–10^8, it is impossible to solve the ARE (due to computational burden), not to say about storing M, P. To overcome these difficulties, the AF is proposed. Mention that the KF is also an MMSE filter in the complete Hilbert space of random variables,

$E|\xi|^2 < \infty$, with scalar product $(\xi_1, \xi_2)_H = E < \xi_1 \xi_2 >$ and the norm $||\xi|| = \left[E\left(|\xi|^2 \right) \right]^{\{1/2\}}$.

For the nonlinear models, there are KF variants, among those are the extended KF (EKF) [7], the unscented KF (UKF, [8]), and the Ensemble Kalman filter (EnKF, [9]). In the EnKF, the ECM is a sampled ECM whose samples are generated using samples of the state variable, and consequently the ECM in the KF becomes a sampled ECM. For an example of application of the EnKF for data assimilation in geophysical data assimilation with high dimensional model, see [10]. Another class of ensemble filtering technique is a class of particle filters (PF, [11]). The basic idea of the PF (also the EnKF) is to use a discrete set of weighted n particles to represent the distribution of $x(t)$, where the distribution is updated at each time by changing the particle weights according to their likelihoods.

Despite a possible implementation of the KF variants, they might still be seriously biased because the accuracy of the KF update requires linearity of the observation function and Gaussianity of the distribution of system state $x(t)$. In reality, the KF (4)–(8) may be biased and unstable, even divergent [12]. Today, the PF algorithms are ineffective for HdS data assimilation.

In this chapter, we shall show how the AF can be efficient in dealing with uncertainties existing in the filtering problem (1) and (2). In Section 2, a brief outline of the AF is given. The main features of the AF, which are different to those of the KF, are presented. This concerns the optimal criteria, stabilizing gain structure, optimization algorithms. Section 3 shows in detail how the AF is capable of dealing with uncertainties in the specification of ME covariance. The hypothesis on a subspace of ME is presented in Section 4 from which one sees how one can make order reduction for representing the bias and ME covariance. Simple numerical examples on one- and two-dimensional systems are given in Section 5 to illustrate in details the differences between the AF and the traditional KF. Numerical experiments on low and high dimensional systems are given in Section 6 to demonstrate how the AF algorithm works. The performance comparison between the AF and KF, for both situations of perfect knowledge of ME statistics and that with ME uncertainties, is also presented. Conclusions and perspectives of the AF are summarized in Section 7.

2. Adaptive filter

The AF is originated from [13]. It is constructed for estimating the state of a dynamical system based on partially observed quantities related in some way to the system state. As reported before, for linear systems contaminated by Gaussian noise, the MMSE estimate can be obtained by the KF. Since publication of [1] in 1960, an uncountable number of works are done for solving engineering problems by KF, in all engineering fields, as well as many modifications have been proposed. The reasons for the need in modification of the KF are numerous, but mostly related to nonlinear dynamics, parameter uncertainties in specification of system parameters, bias of ME, unknown statistics of ME, model reduction. With the rapid progress of computer technology (computer power, memory, …), various simplified versions of KF are suggested for solving filtering (or data assimilation) for HdSs, in particular, in meteorology and oceanography.

Direct application of the KF to HdSs is impossible due to the limit in computer power, memory, and computational time. In particular, the KF requires to solve the matrix AREs (7) and (8) for computing ECMs $M(t), P(t)$. Storing such matrices is impossible, not to say on computational time.

Different simplified approaches are proposed for overcoming difficulties in the application of the KF. The example of successful tool for solving data assimilation problems in HdSs is the EnKF [9]. In the EnKF, an ensemble of error samples, of small size, is generated on the basis of model states to approximate the ECMs. In practice of data assimilation for HdSs, it is possible to generate only ensembles of moderate sizes (of order $O(100)$) by model integrations over the assimilation window (time interval between two successive arrivals of observations) since one such integration takes several hours! The other approach like PF is based on sampling from conditional distributions. Theoretically, this approach is more appropriate for nonlinear problems because no linearization is required as in the EKF (Extended KF based on linearization technique). However, even for filtering problems with state dimensions of order $O(10)$, relatively large ensembles of size $O(10000)$ would be required in order to produce reasonably good performance.

The AF in [13] is based on the different idea. Here, no linearization is required for nonlinear filtering problems. For the problem (1) and (2), the filter is of the form (4) and (5) but the gain $K = K(\theta)$ is assumed to be of a given stabilizing structure [6]. It means that K is parametrized by some vector of unknown parameters $\theta \in \Theta$ so that the filter (4) and (5) with the gain $K(\theta), \forall \theta \in \Theta$ is stable. It is well-known that under mild conditions, the solution of the ARE will tend (quickly) to stationary solution M_∞ and so the gain (6), to the stationary gain K_∞. Moreover, the innovation $\zeta(t+1) = z(t+1) - \hat{z}(t+1/t)$, representing the error for the output prediction $\hat{z}(t+1/t) := H\hat{x}(t+1/t) = H\Phi\hat{x}(t)$, is unbiased and of minimum variance. This fact leads to the idea to seek the optimal vector θ by minimizing

$$J(\theta) := E[\Psi(\zeta)] \to \arg\,\min \theta \qquad (9)$$

here $E[.]$ denotes the *average* in a *probabilistic sense*. For stationary systems (1) and (2), if we assume the validity of the ergodic hypothesis, the *average* value in a *probabilistic sense*, expressed in (7), is almost everywhere equivalent to the time average (for large time of running the dynamical system). The optimal θ^* can be found by solving the equation

$$\nabla_\theta J(\theta) = \nabla_\theta E[\Psi(\zeta)] = E[\nabla_\theta \Psi(\zeta)] = 0 \qquad (10)$$

A stochastic approximation (SA) algorithm for solving (10) can be written out

$$\theta(t+1) = \theta(t) - \gamma(t)\nabla_{\theta(t)}\Psi[\zeta(t+1)] \qquad (11)$$

Conditions related to the sequence of positive scalar $\gamma(t)$ for ensuring a convergence $\{\theta(t)\}$ in the procedure (11) are

$$\gamma(t) > 0, \sum_{t=0}^{\infty}\gamma(t) = \infty, \sum_{t=0}^{\infty}\gamma^2(t) < \infty \qquad (12)$$

One of the most advantages of the SA algorithm (11) is that, instead of computing the gradient of the cost function (9) (which requires knowledge of probability distribution), the algorithm (11) is based on the knowledge of only the gradient of sample cost function Ψ (wrt to θ) which can be easily evaluated numerically.

Comment 2.1. Generally speaking, the convergence rate of the algorithm (11) and (12) is $O(1/t)$. It is possible to improve the convergence rate of the SA by averaging of the iterates,

$$\theta_m(t) = \frac{1}{t} \sum_{t'=1}^{t} \theta(t') \qquad (13)$$

For more details, see [14].

Comment 2.2. For high HdSs, even with θ being of moderate dimension, instead of the algorithm (12) or (13), the SPSA (Simultaneous Perturbation Stochastic Perturbation) algorithms in [15, 16] are of preference. That is due to the fact that integration of HdS over the assimilation window is very expensive. These algorithms generate stochastic perturbation $\delta\theta = (\delta\theta_1, \dots, \delta\theta_n)^T$ with components as Bernoulli i.i.d. realizations. Each ith component of the gradient-like (pseudo-gradient) vector is computed as the divided difference $\delta\Psi/\delta\theta_i$ where $\Psi := \Psi[\theta + \delta\theta] - \Psi[\theta]$. This allows to evaluate the gradient-like vector by only two or three time integration of the numerical model.

For details on the SPSA algorithm and its convergence rate, see [15, 16].

3. Covariance uncertainties and AF

3.1 Covariance uncertainties

3.1.1 Adjoint approach

As seen from (11) and (12), implementation of SA algorithms is much simpler for searching optimal gain parameters compared to the other optimization methods. The SA algorithms require only numerical computing derivatives $\nabla_\theta \Psi[\zeta(t+1)]$ of the sample cost function $\Psi[\zeta(t+1)]$ wrt θ evaluated at $\theta(t)$ and $\gamma(t)$ is a scalar which can be chosen a priori, for example, as $\gamma(t) = \frac{1}{t}$. That is possible due to introducing the ergodic hypothesis on of the system (1) and (2) from which there exists an asymptotic optimal gain

First, consider the situation when the vector of parameters consists of are all elements of K, $\theta = K$. Compute the innovation vector,

$$\zeta(t+1) = z(t+1) - \hat{z}(t+1/t) = z(t+1) - H\Phi\hat{x}(t)$$

$$\hat{x}(t) = \hat{x}(t-1/t) + K(\theta)\zeta(t)$$

$$\delta_\theta \zeta(t+1) = -H\Phi\delta_\theta\hat{x}(t) = -H\Phi\delta_\theta K(\theta)\zeta(t) = -H\Phi\delta K \zeta(t).$$

$$\delta_K \Psi[\zeta(t+1)] = \delta_K <\zeta(t+1), \zeta(t+1)> \; = 2<\zeta(t+1), \delta_K\zeta(t+1)>$$

$$= -2<\zeta(t+1), H\Phi\delta K\zeta(t)> \; = -2<\Phi^T H^T\zeta(t+1), \delta K\zeta(t)>.$$

Let us compute derivatives of the sample cost function Ψ wrt the elements K_{ij} of the gain K. To do so, one needs to integrate the adjoint operator Φ^T s.t. the forcing $f := -H^T\zeta(t+1)$ which yields $\psi := -\Phi^T H^T\zeta(t+1)$ and hence

$$\delta\Psi/\delta K_{ij} = -\psi_i\zeta(t,j), \qquad (14)$$

here ψ_i is the ith component of ψ, $\zeta(t,j)$ the jth component of $\zeta(t)$. The AF now takes the form

$$\hat{x}(t+1) = \hat{x}(t+1/t) + K\zeta(t+1), \qquad (15)$$

$$\hat{x}(t+1/t) = \Phi\hat{x}(t), \zeta(t+1) = z(t+1) - H\hat{x}(t+1/t), \qquad (16)$$

$$K(t+1) = K(t) - \gamma(t)\nabla_K\Psi[\zeta(t+1)], \qquad (17)$$

where $\nabla_K\Psi(\zeta(t+1))$ is the gradient vector whose components are computed by (14). In the AF (15)–(17), no matrix ARE (see (7) and (8) in the KF) is involved. The AF (15)–(17) is quite realizable for HdSs, since at each assimilation instant we need to integrate only the direct model to produce the forecast (16) and (eventually) an adjoint model over the assimilation window for computing $\nabla_K\Psi[\zeta(t+1)]$ [13].

3.1.2 Simultaneous perturbation stochastic approximation (SPSA) approach

Remark that in the form (14) the adjoint operator Φ^T would be available to implement the AF. It is well-known that construction of numerical code for Φ^T is a very difficult and heavy task, especially for meteorological and oceanic numerical models which are HdSs and nonlinear (linearization is required).

A comparison study of the AF with other assimilation methods is done in [17]. Compared to the AF, the widely used variational method (VM) minimizes the distance between the observations available (for example, the observations of the whole set $Z[1:T]$) and the outputs of the dynamical system. This optimization problem is carried out in the phase space, hence is very difficult and expensive. Theoretically, a simplification is possible subject to (s.t.) the condition of linearity of the dynamical system: in this case, one can reformulate the VM minimization problem as searching the best estimate for the initial system state $x(0)$. For HdSs, to ensure a merely high quality estimate for $x(0)$, it is necessary: (i) to take the observation window as large as possible; (ii) to parameterize the initial state by some parameters (using a slow manifold, for example). Iterative minimization procedures require usually $O(10)$ iterates involving integrating the direct and adjoint models over the window $[1, T]$. For an unstable dynamics, integration of direct and adjoint equations over a long period naturally amplify the initial errors during assimilation process. For a more detailed comparison between the AF and VM, see [17].

Thus, if the ergodic conditions hold, there exists an optimal stationary gain and the AF in limit will approach to the optimal one *in the given class of stable filters*. It is important to emphasize that up to this point, no covariance matrices Q, R are specified. It means that the AF in the form (12) is robust to uncertainties in the specification of the covariances of the ME and ObE.

3.2 Stability of the AF

One of the main features of the AF is related to its stability.

For simplicity of presentation, in the previous section, the AF algorithm is written out under the assumption (13). In practice, application of the AF in the form (13) is not recommended since instability may occur. It is easy to see that the transition matrix of the filter is given by

$$L = (I - KH)\Phi \qquad (18)$$

It is evident that if we do not take care on the structure of K, varying stochastically all elements of K can lead to instability of L and the filter will be exploded. Moreover,

for HdSs, the number of elements of K is very large. It is therefore primordial to choose a parametrized stabilizing structure for K (depending on θ) to ensure a stability of L and reducing a number of tuning parameters. This question is addressed in [6]. One of possible structures for K is of the form

$$K(t) = P_r \Theta K_e, K_e := M_e H_e^T \left[H_e M_e H_e^T + \alpha I \right]^{-1}, H_e := HP_r \qquad (19)$$

where $P_r \in R^{nxr}$ is a matrix with dimensions $n \times r$, r is the dimension of the reduced space (equal to the number of unstable eigenvectors (EiVecs)of Φ), the matrix M_e is a strictly positive symmetric definitive playing the role of the ECM in the reduced space R_e, Θ is a diagonal matrix with diagonal elements θ_i whose values belong to $(\epsilon, 2 - \epsilon)$, i.e. $\theta_i \in (\epsilon, 2 - \epsilon)$, with $\epsilon \in (0, 1)$ whose value depends on the modulus of the first stable eigenvalue (EiV) [6]. We will refer to the filter s.t. (19) with $\Theta = Id$ (Id is the identity matrix of appropriate dimension) as a *nonadaptive filter* (NAF). In the AF, the parameters θ_i are adjusted each time when a new observation arrives, to minimize the cost function (9). Thus θ_i is a time-varying function. As to the matrix P_r, its choice is important to ensure a filter stability. One simple and efficient procedure (called Prediction Error Sampling Procedure—PeSP) to generate P_r is to use the power orthogonal iteration method [18] which allows to compute real leading Schur vectors (SchVecs) of Φ_i. The advantage of using the SchVecs compared to the EiVecs, is that they are real and their computation is stable. It is seen that the optimal AF is found in a class of stable filters which is stable even for an unstable numerical model. As to the VM, the optimal trajectory is found on the basis of only the numerical model with the initial state as a control vector. It means that for unstable dynamics, the errors in the forcing or numerical errors arising during computations will be amplified and lead to large estimation error growth. More seriously, the VM requires a large set of observations and large number of iterations (i.e., many forward and backward integrations of the direct and adjoint models) which naturally leads to increase of estimation error too.

3.3 On improving the initial gain

Consider the gain structure (19). Suppose that M_e has been chosen in agreement with the required stability conditions. Before tuning the parameters θ_i to minimize the cost function (9), remark that stability of the filter is still ensured for the following gain:

$$K(t) = P_r \Lambda \Theta K_e, \qquad (20)$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)$, $\lambda_k \in (0, 1)$ since then for $\Gamma = \Lambda \Theta = \text{diag}(\gamma_1, \ldots, \gamma_r)$ it implies $\gamma(t) = \lambda(t)\theta(t) \in (\epsilon, 2 - \epsilon)$, where $\epsilon \in (0, 1)$. Writing the equation for the filtered error (FE) $e_f(t) := \hat{x}(t) - x(t)$ one sees that the matrix L in (18) also represents the transition of the FE $e_f(t)$. It means that it is possible to choose a more optimal initial gain by solving, for example, the minimization problem

$$J_o(\Lambda) = \|L(\Lambda)\|_2^2 \rightarrow \min{}_\Lambda \qquad (21)$$

The problem (21) is solved without using the observations, hence it is offline. Once the optimal Λ^* has been found, the standard AF is implemented s.t. the filter gain

$$K(t) = P_r \Theta K_e^*, K_e^* := \Lambda^* K_e \qquad (22)$$

It is seen that using the structure (22) this optimization procedure does not require the information on the ME statistics.

4. Joint estimation of state and model error in AF

The previous section shows how the AF is designed to deal with the difficulty in specification of covariances of the ME and ObE. This is done without exploiting a possibility to determine, more or less correctly, a subspace for the ME. If such a subspace can be determined without major difficulties, it would be beneficial for better estimating the AF gain and improving the filter performance. In [19], the hypothesis of the structure of the ME has been introduced and a number of experiments have been successfully conducted.

There is a long history of joint estimation of state and ME for filtering algorithms, in particular, with the bias and covariance estimation. One of the most original approaches, dealing with the treatment of bias in recursive filtering (known as bias-separated estimation—BSE), is carried out by Friedland in [20]. He has shown that the MMSE *state estimator* for a linear dynamical system augmented with bias states can be decomposed into three parts: (1) bias-free state estimator; (2) bias estimator; and (3) blender. This BSE approach has the advantage that it requires fewer numerical operations than the traditional augmented-state implementation and avoids numerical ill-conditioning compared to the case of bias-separated estimation by filtering technique.

It is common to treat the bias as part of the system state and then estimate the bias as well as the system state. There are two types of ME—deterministic (DME) and stochastic (SME). Generally speaking, a suitable equation can be introduced for the ME. In the presence of bias, under the assumption on constant b, instead of (1) one has

$$x(t+1) = \Phi x(t) + b(t) + w(t), b(t+1) = b(t), t = 0, 1, 2 \dots \tag{23}$$

To introduce a subspace for the variables $w(t), b(t)$ the SME and DME in (23), let

$$w = G_w \rho, b = G_b d$$

$$G_w \in R^{n \times n_w}, G_b \in R^{n \times n_d}, n \geq n_w, n \geq n_b \tag{24}$$

Generally speaking, G_w, G_b are unknown, and finding reasonable hypothesizes for them is desirable but not self-evident. In [19], one hypothesis for G_w, G_b has been introduced (it will be referred to as Hypothesis on model error—HME).

The information on G_w, G_b, given in (25), allows to better estimate the DME b and SME w for improving the filter performance, especially for $n_b < n, n_w < n$ in a HdS setting. The difficulty, encountered in practice of operational forecasting systems, is that (practically) nothing is given a priori on the space of the ME values. To overcome this difficulty, one simple hypothesis has been introduced in [19]. This hypothesis is postulated by taking into consideration the fact that for a large number of data assimilation problems in HdSs, the model time step δt (chosen for ensuring a stability of numerical scheme and for guaranteeing a high precision of the discrete solution) is much smaller than Δt—the assimilation window (time interval between two successive observation arrivals).

Suppose that $\Delta t = n_a \delta t$ where n_a is a positive integer number.

Hypothesis (on the subspace of ME—HME) [19]. Under the condition that n_a is relatively large, the ME belongs to the subspace spanned by all unstable and neutral EiVecs (or SchVecs) of the system dynamics Φ.

5. Simple numerical examples

5.1 One-dimensional system

To see the difference between the AF and the KF in doing with ME uncertainties, introduce the one-dimensional system

$$x(t+1) = \Phi x(t) + w(t), z(t+1) = hx(t+1) + v(t+1), t = 0, 1, \ldots \quad (25)$$

In (25), Φ is the unique eigenvalue (also the singular value) of the system dynamics.

i. For simplicity, let $\Phi = 1, h = 1$. This corresponds to the situation when the system is *neutrally stable*. The filter fundamental matrix (18) now is $L(K) = (1 - K)$ which is stable if $K \in (0, 2)$. For the KF gain (4)–(8), as $K_{kf}(t) = \frac{M_{kf}(t)}{M_{kf}(t)+R}$ we have $K_{kf}(t) \in (0, 1)$, $M_{kf}(t)$ is the solution of (7). That is true for any $M_{kf}(t) \geq 0, R > 0$. It means then the KF is stable. Mention that if $Q > 0$ always $M_{kf}(t) > 0$. In general, $K_{kf}(t) = M_{kf}(t) [M_{kf}(t) + R]^+$ where $[A]^+$ is the pseudo-inverse of A [21].

For the AF, we have in this case $P_r = 1$. Consider the gain $K_{af}(\theta) := P_r \theta K_e$, where K_e is the gain of the form $K_e = \frac{M_e}{M_e+R}$, $M_e > 0, R > 0$, M_e is constant. We have then for the NAF ($\theta = 1$) $0 < K_e < 1$ and $K_{naf} = K_e$.

For the AF, the transition matrix (18) reads $L_{af}(\theta) = (1 - \theta K_e)$. For $\theta \in (0, 2)$, $|L_{kf}(\theta)| \in (0, 1)$, $K_{af}(\theta) \in (0, 2)$ and the AF is stable. It is evident that there is a larger margin for varying the gain in the AF than that in the KF since $K_{kf}(t) \in (0, 1)$. One sees that the stationary KF is a member of the class of stable AFs (19). The performance of AF is optimized by solving the problem (9) using the procedure (11) and (12) or SPSA algorithms (*Comment 2.2*).

ii. Let $\Phi < 1$, i.e., the system (1) is *stable*. The results in (i) are valid for the AF structure. In this situation, the filter is stable even for $K_{af} = 0$.

iii. Let $\lceil \Phi \rceil > 1$—the system (1) is *unstable*. Consider two situations (a) $\Phi > 1$; (b) $\Phi < -1$. As before $P_r = 1$.

For $\Phi > 1$ we have

$$\frac{\Phi - 1}{K_e \Phi} < \theta < \frac{\Phi + 1}{K_e \Phi} \quad (26)$$

In particular, when $\Phi \to 1$, approximately $\theta \in \left(0, \frac{2}{K_e}\right)$. When $Q \gg R$ (that is usually in practice), approximately $\theta \in (0, 2)$ as in the situation (i). For large $\Phi \gg 1$, $\frac{\Phi-1}{K_e \Phi} \to \frac{1}{K_e}$ (left-hand limit), $\frac{\Phi+1}{K_e \Phi} \to \frac{1}{K_e}$ (right-hand limit) and there remains no margin for varying θ (or $Q \gg R$) and $K_{af} \to 1$. It is important to emphasize that as K_e is chosen by designer, we can define the interval for varying θ if the amplitude of Φ is more or less known. In practice, one can vary $\theta \in [\epsilon, 2 + \epsilon]$ with small $\epsilon > 0$ for Φ close to 1, and with ϵ close to 1 for large Φ.

For $\Phi < -1$ we have

$$\frac{\Phi + 1}{K_o \Phi} < \theta < \frac{\Phi - 1}{K_o \Phi} \tag{27}$$

It is seen from (27) that when $\Phi \to -1$, approximately $\theta \in \left(0, \frac{2}{K_e}\right)$. As for the situation $\Phi \ll -1$, $\frac{\Phi + 1}{K_o \Phi} \to \frac{1}{K_o}$ (left-hand limit), $\frac{\Phi - 1}{K_o \Phi} \to \frac{1}{K_o}$ (right-hand limit) when $Q \gg R$ approximately $\theta \to \frac{1}{K_o}$ hence $K_{af} \to 1$.

It is important to stress that the KF gain is computed on the basis of Q and R (under the condition that the statistics of the initial state will be forgotten as t becomes large); whereas, the gain of the AF is updated on the basis of samples of the innovation vector. It means that the KF is optimal in the MMSE sense (under the condition of exact knowledge of the required statistics) whereas the AF is optimized during the assimilation process using PE realizations of the system output (innovation vector). The KF gain can be computed in an offline fashion, whereas the AF gain is a function of observation and computed in online.

5.2 Two-dimensional system: specification of covariances

5.2.1 Stable filter

To see the role of the correction subspace $R[P_r]$ in ensuring a stability of the AF, let us consider the system (1) and (2) s.t.

$$\Phi = diag\,(\Phi_{11}, \Phi_{22}), H = diag\,(1, 1) \tag{28}$$

Consider the AF with the gain (19) s.t. $P_r = Id$, i.e., two columns of P_r are in fact the EiVecs associated with two EiVs Φ_{11} and Φ_{22}. Let us denote the AF gain $K_{af} = P_r \Theta M H^T [HMH^T + R]^{-1} = \Theta K_e$ with $K_e := M_e H^T [HM_e H^T + R]^{-1}$ which is structurally identical to that of the KF. For the nonadaptive filter $K_{naf} = K_e$ and with the choice $M_e = diag\,(M_{11}, M_{22})$, taking into account (28) one gets

$$K_{naf} = diag\,(K_{11}, K_{22}), K_{ii} = \frac{M_{ii}}{M_{ii} + R_i}, i = 1, 2 \tag{29}$$

$$L_{naf} = diag(l_{11}, l_{22}), l_{ii} = \left(1 - \frac{M_{ii}}{M_{ii} + R_i}\right)\Phi_{/ii}, i = 1, 2 \tag{30}$$

The filter transition matrix (30) is obtained on the basis of $L_{naf} = (I - K_{naf}H)\Phi$ and the assumption (29). It is easily to see that L_{naf} has two EiVs, $\lambda_i = l_{ii}, i = 1, 2$ where l_{ij} est. the (ij) element of L_{naf}.

Stability of the filter depends on the condition $|l_{ii}| < 1, i = 1, 2$. For $M_{ii} > 0, R_i > 0$, if Φ_{ii} is *stable* or *neutrally stable*, i.e. $|\Phi_{ii}| \leq 1, i = 1, 2$, we have $|l_{ii}| < 1$. For *unstable* Φ_{ii} ($|\Phi_{ii}| > 1, i = 1, 2$), the filter is unstable if $\Phi_{ii} > \frac{M_{ii} + R_i}{R_i}$ (situation $\Phi_{ii} > 1$) or $\Phi_{ii} < -\frac{M_{ii} + R_i}{R_i}$ (situation $\Phi_{ii} < -1$). These conditions should be taken into account when the EiVs of Φ are large.

For the AF gain (19) ($P_r = I$),

$$l_{ii} = \left(1 - \frac{\theta_i M_{ii}}{M_{ii} + R_i}\right)\Phi_{ii}, \tag{31}$$

From (31) conditions for $|l_{ii}| < 1$ can be obtained as done in Section 5.1 with the one-dimensional system since $l_{ii}, i = 1, 2$ are independent one from another. The length of the interval I_i for varying θ_i depends on the value of Φ_{ii} (see (26)).

This example shows that for $P_r = Id$, it is always possible to construct a stable AF whatever are the EiVs of Φ (stable or unstable). There are some constraints for M_{ii} (they are positive) and for R_i (small positive). Optimality of the AF is obtained by searching recursively (in time) the optimal θ_i during assimilation process. Thus, in the AF, a correct specification of ME and ObE statistics (second order) is not important as happens in the KF.

5.2.2 Unstable filter

Consider the situation when P_r is constructed from only one vector. Let $P_r = (1, 0)^T$—the EiVec associated with Φ_{11} (the results remain the same if we choose $P_r = (0, 1)^T$—the EiVec associated with Φ_{22}).

$$\Phi = diag\,(\Phi_{11}, \Phi_{22}), |\Phi_{11}| > 1, |\Phi_{22}| > 1, H = diag(1, 1) \tag{32}$$

We show now that the filter with the gain (19) is unstable. We have (for $\Theta = Id$),

$$H_e = HP_r = P_r, K = P_rK_e, K_e = M_eH_e^T\left[H_eM_eH_e^T + R\right]^{-1} = M_eP_r\left[P_rM_eP_r^T + \alpha I\right]^{-1} \tag{33}$$

if we put $R = \alpha I$. For $L_{naf} = [I - KH]\Phi$, taking into account (32), (33) it implies

$$L_{naf} = \begin{bmatrix} \dfrac{\alpha\Phi_{11}}{m_e + \alpha} & 0 \\ 0 & \Phi_{22} \end{bmatrix} \tag{34}$$

As $\frac{\alpha}{m_e + \alpha}$ can be made as small as desired by choosing small $\alpha > 0$, the first EiV $l_{11} = \frac{\alpha\Phi_{11}}{m_e + \alpha}$ can be made stable. However, the second EiV in (34) $l_{22} = \Phi_{22} > 1$ is unstable. It implies that the filter with the gain (19) s.t. $P_r = (1, 0)^T$ is unstable. This happens even for $\Theta \neq Id$. It means that when the projection subspace $R[P_r]$ does not contain all unstable and neutral EiVecs of the system dynamics, it is impossible to guarantee a stability of the filter.

5.3 Two-dimensional system: estimation of ME

Consider the filtering problem (1) and (2), the dynamical system (1) describes a sequence of system states at time instants $t = 0, 1, \ldots$ when the observations are available. It means that Φ represents the transition of system state over the (observation) time window Δt separating arrivals of two successive observations. In practice, the interval Δt is much larger than the *model time step δt* which is the step size in approximating the temporal derivative. The choice of δt is important for guaranteeing a stability of discretized scheme and having high is important for guaranteeing a stability of discretized scheme and having high precision of the discretized solution (wrt the continuous solution). We have then $\Delta T = n_a\delta t$, where n_a is a relatively large positive integer. For example, in the HYCOM model at SHOM (French marine) for the Bay of Biscay configuration, the interval Δt between two observation arrivals is 7 days which is equivalent to integrating 1200 model time

steps δt. It means $n_a = 1200$. Symbolically we have then the equations for model time step integration

$$x'(\tau + 1) = \Phi' x'(\tau) + \psi'(\tau), \psi'(\tau) := b'(\tau) + w'(\tau) \tag{35}$$

In (35), Φ' represents the integration of numerical model over one model time step δt. Hence

$$\Phi = [\Phi']^{n_a} \tag{36}$$

The contribution of $\psi'(\tau)$, over the assimilation window $[t - 1, t]$ (for simplicity and without loss of generality, one supposes $t - 1 := 0, t := n\_a$) is

$$\psi(t) = b(t) + w(t), b(t) := \sum_{\tau=0}^{n_a} \nu_1(\tau)_1, \nu_1(\tau) := [\Phi']^{n_a - 1 - \tau} b'(\tau),$$

$$w(t) := \sum_{\tau=0}^{n_a} \nu_2(\tau), \ \nu_2(\tau) := \Phi^{n_a - 1 - \tau} w'(\tau), [\Phi']^{-1} := 0 \tag{37}$$

The HME in Section 4 says that the SME $w(t)$ and DME $b(t)$, as functions of n_a, belong to the subspace spanned by leading EiVs (or SchVecs) of Φ for a relatively large n_a. The initial filtering problem now has the form (1) and (2) s.t. (36) and (37) where $t = \tau/n_a$.

To illustrate this HME, continue the two-dimensional system in Section 5.2.2 and suppose that $|\Phi'_{11}| > 1, |\Phi'_{22}| < 1$. Applying HME in this case is equivalent to saying that the values of MEs $b(t), w(t)$, approximately, belong to the subspace $R[u_1]$ spanned by the first EiVec $u_1 = col(1, 0)$, associated with the EiV Φ'_{11}. Here $y = col(y_1, \dots, y_n)$ denotes the vector-column with components $y_1, \dots, y_n$. It follows that the covariance matrix of $w(t)$ is assumed to be of the form $Q = \sigma_w^2 u_1 u_1^T$ and $b(t)$—of the structure $b(t) = cu_1$, c is a scalar to be estimated. For the algorithm of joint estimation of state and bias (in term of c), see [19].

6. Simulation results

6.1 One-dimensional system

In this section, the filtering problem (25) in Section 5.1 is considered s.t.

$$\Phi = 0.99, H = 1.02, Q = 0.09, R = 0.01.$$

The true system states and observations are simulated using the initial state $x(0) = 1$ and $w(t), v(t)$ are zero mean Gaussian mutually uncorrelated and temporal uncorrelated sequences.

To see the performance of the AF, unknown system states are estimated on the basis of the AF algorithm. To obtain a reference, the standard KF is also implemented for solving this filtering problem. In the filtering algorithms, the estimate of the initial state is $\hat{x}(0) = 2$. The gain K_{naf} in the NAF is taken as that of the KF at $t = 0$, i.e., $K_{naf} = K_{kf}(0)$.

Figure 1 shows the temporal evolution of the parameters $\theta_m(t)$ during assimilation process.

The gains in the KF and AF during the assimilation process are displayed in **Figure 2**. Mention that the KF gain is computed s.t. true statistics Q, R. In the AF, $\theta_m(t)$ has been used for computation of the AF gain, i.e., $K_{af} = \theta_m(t)K$. From **Figure 2**, one

sees that initialized by the same value, the two gains become different during assimi-
lation process. The KF gain has reached a stationary regime very quickly.

The mean temporal RMS (root mean square) of the innovation is shown in
Figure 3. It is interesting to remark that no significant difference is observed
between two curves and a slightly better performance is produced by the KF.

In **Figure 4**, we show RMS of the state FE produced by the KF and AF under the
condition that the variance Q is known exactly. One sees that the KF, as expected,
produces the best results.
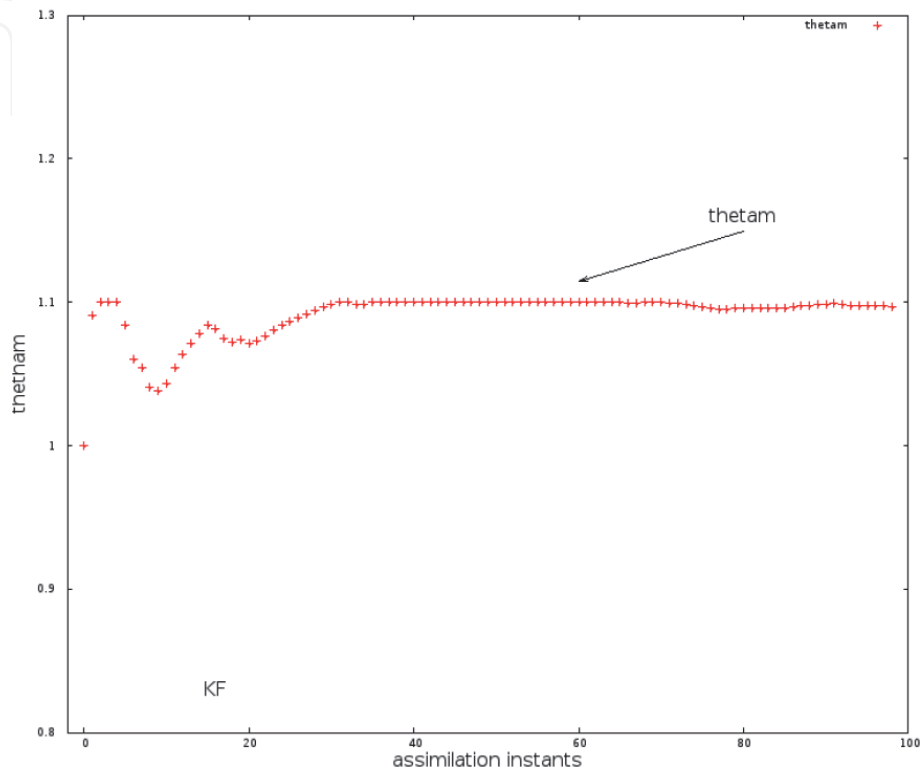
Figure 1.
Temporal evolution of the parameter $\theta_m(t)$ in the AF gain.
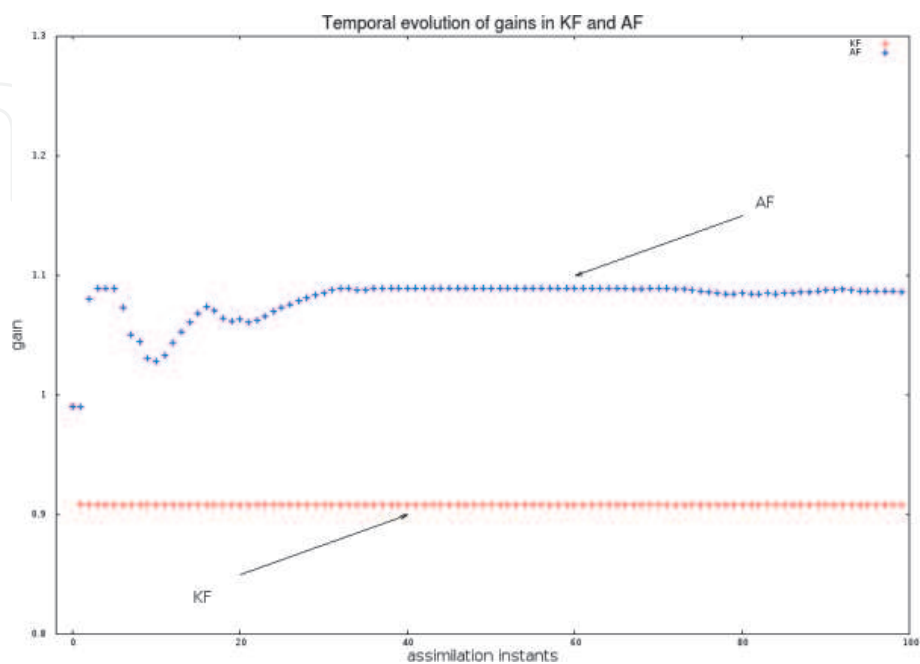
Figure 2.
Temporal evolution of gains in KF and AF during data assimilation.
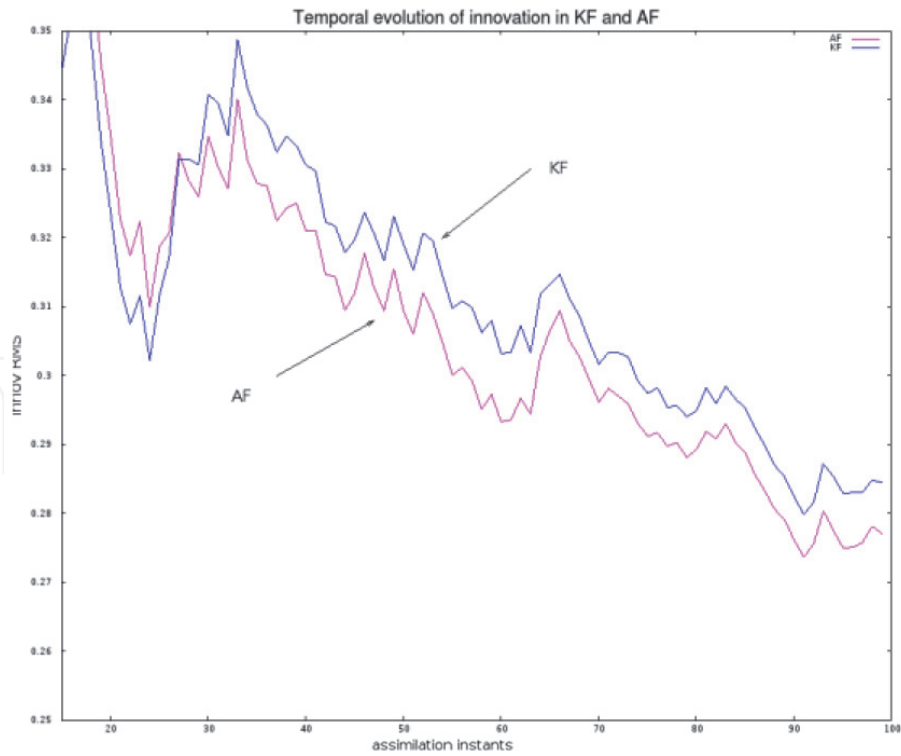
Figure 3.
RMS of innovation produced by the KF and AF.
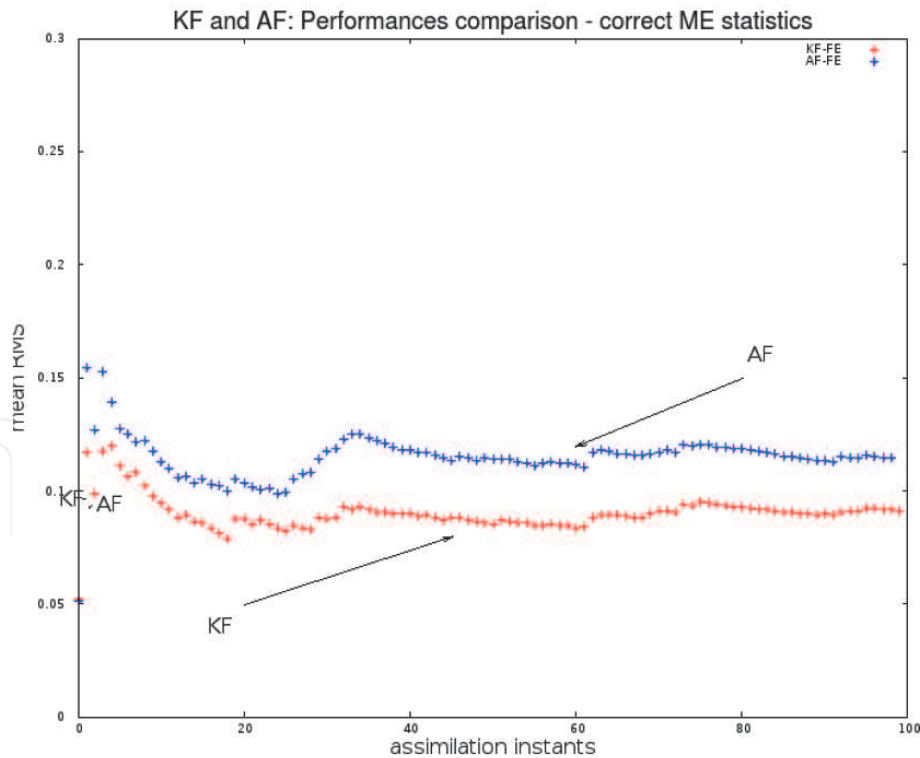
Figure 4.
RMS of the state FE produced by the KF and AF under the condition that the variance of ME is known exactly. It is seen that when the ME is correctly specified, the KF behaves better than the AF.

Figure 5 shows the RMS of FE as a function of the variance Q. Here, the value of Q varies from 0.1 to 1.9. Note that the true value of Q is 0.1. The red curve represents the RMS of FE produced by the KF at the end of the assimilation period (as a function of Q). The green curve has the same meaning, but for the FE
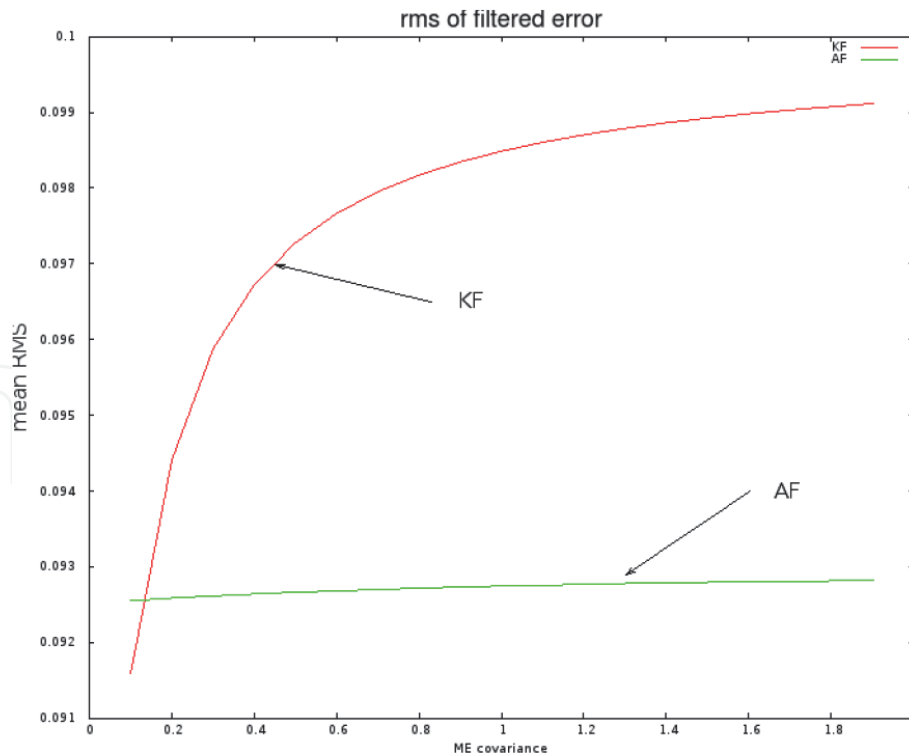
Figure 5.
RMS of FE as a function of Q. The true value of Q is equal to 0. It is noted that the KF behaves better than the AF s.t. true Q but is more and more degraded as the ME becomes greater and greater. At the same time, the FE of the AF remains very robust.

produced by the AF. It is interesting to note that when Q is correctly specified, the KF behaves better than the AF, but misspecification of Q leads to growing of the error in the KF. The AF is robust wrt the error in the specification of Q. This fact says in favor of the AF as an efficient tool for overcoming uncertainties in the ME.

6.2 Two dimensional system

6.2.1 Illustration of hypothesis HME

According to the notations in Section 5.3, consider the two-dimensional system (1) with $\Phi' = [\Phi'_{ij}], \Phi'_{11} = 1.02, \Phi'_{12} = 0.1, \Phi'_{21} = 0, \Phi'_{22} = 0.9, H = (1, 1)$ with the true DME $b'(\tau) = \text{col}(0.1, 0.1)'$. Thus the first EiV is unstable, the second—stable [19].

Numerically one finds that the first SchVec is equal to $u_1 = (-1, -7.0 \, E - 7)^T$.

Figure 6 [19] shows the simulation results obtained on the basis of (37). One sees that, for $n_a > 10$, the second component of $w(t)$ is close to 0 whereas the first component becomes bigger and bigger (in absolute value) as n_a increases. Here, $w'(\tau)$ is a sequence of independent two-dimensional Gaussian random vectors of zero mean and variance 1. This means that the values of $w(t)$ become more and more close to the subspace $R[u_1]$ spanned by u_1, hence the HME is practically valid for $n_a > 10$ in this example. Mention that, as a rule, in ocean numerical models, n_a is of order $o(100)$ ($n_a = 800$ or the MICOM model in the experiment in Section 6.3). See also [22].

6.2.2 Assimilation results

First simulation of the sequence of true system states $x'(\tau), \tau = 1, \dots, 390$ has been carried out (see (35)). The observations are picked at $\tau = 15, 30, \dots, 390$.
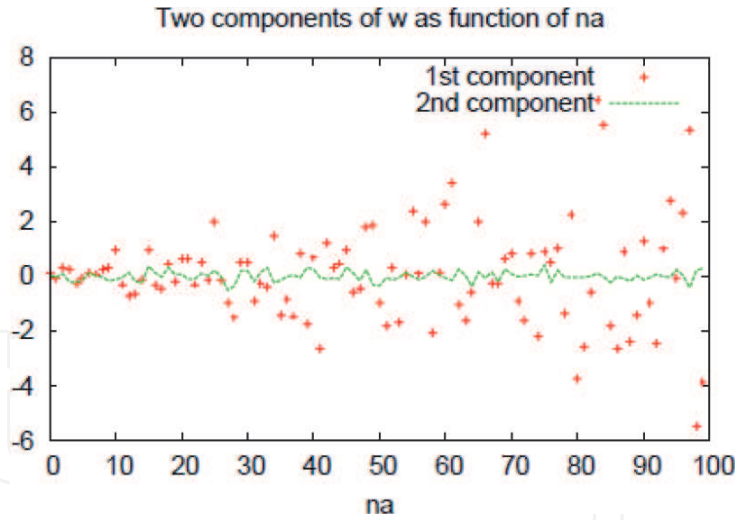
Figure 6.
Two components of $w(t)$ as functions of n_a.

In terms of $x(t)$, the filtering problem then is of the form (1) and (2) s.t. $t = \frac{\tau}{15}$, $\Phi = \Phi'^{15}$ (if no bias exists). When there is a bias, instead of $w(t)$ stands $\psi(t) := b(t) + w(t)$.

In the experiment, the true system states $x'(t)$ are generated by (35) s.t. $b'(\tau) = col(0.1, 0.1)$, $w'(t)$ is zero mean with the covariance $Q' = Id$. The observation error is of zero mean and covariance $R = 0.16$. For the state $x(t)$ in the KF and NAF, the forecast is obtained at each assimilation instant t as $\hat{x}(t + 1/t) = \Phi\hat{x}(t) + \hat{b}(t)$. The simulation yields $b(t) = (0.2296, 2.0589E - 02)$ which results from applying (37) s.t. $b'(\tau) = col(0.1, 0.1)$.

Figure 7 depicts the time evolution of the KF and AF gains. One sees here as in the experiment with 1D system (**Figure 2**) that the KF gain is stabilized very quickly compared to that of the AF gain.

Figure 8 (from [19]) shows the sample time average RMS of the state FE produced by the three filters NAF, KF, and AF. One sees that the AF outperforms the NAF and KF.
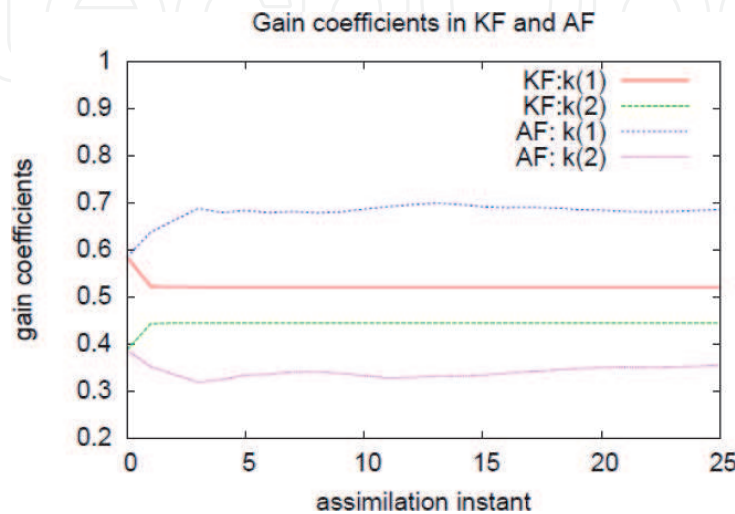

Figure 7.
Gain coefficients in KF and AF: The gains in KF and AF are identical at the beginning of the assimilation process.
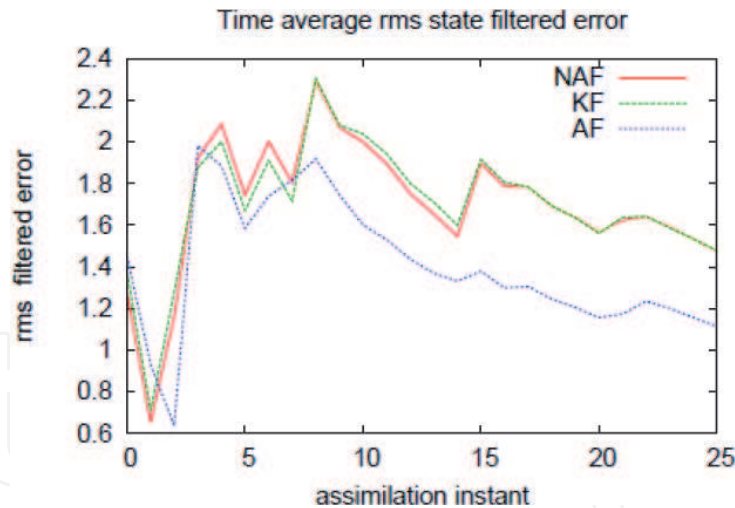
Figure 8.
Sample time average RMS of the state filtered error produced by the NAF, KF, and AF.

6.3 Data assimilation in the high-dimensional ocean model

To illustrate the effectiveness of the AF in dealing with uncertainties in HdSs, this section presents the results on data assimilation in the oceanic numerical model MICOM (Miami Isopycnic Coordinate Ocean Model) [19]. This MICOM describes the oceanic circulation in the North Atlantics. The model has four vertical layers with the state consisting of three variables $x = (h, u, v)$ where h is layer thickness, (u, v) are two velocity components. The horizontal grid is 140×180. Totally at each time instant t we have the state $x(t)$ of dimension 302400 ($140 \times 180 \times 4$ *(layers)* $\times 3$ *(variables)*) . For more details on the configuration of this experiment, see [22].

The experiment is carried out on estimating the oceanic circulation using sea surface height (SSH) measurements. The SSH observation is available each 7 days *(ds)* (hence the observation window $\Delta T = 7ds$). Mention that simulating the circulation over $7\,ds$ requires 800 model time steps (δt) integration.

6.3.1 AF with optimal initial gain

First, in order to examine whether the method of optimal gain initialization, described in Section 3.2, is really useful for improving the filter performance, the optimization problem (21) has been solved. Symbolically, in the gain (20), $P_r = \left[Id, (QG)^T \right]^T$ where Id is the identity operator on the space of layer thickness h, QG is the quasi-geostrophy operator computing the correction for velocity using the SSH innovation ζ. The gain K_e computes the correction for h using ζ. The ECM $M = M_v \otimes M_h$—Kronecker product of M_h—ECM of horizontal variable, M_v—ECM with vertical variable (see below for details). The two parameter matrices Θ and Λ are related to parameterization of M_v. The problem (21) is solved s.t. $\Theta = Id$—identity operator.

The optimal parameters $\lambda_i, i = 1, \dots, 4$ are found by solving the minimization problem (21) using SPSA algorithm. **Figure 9** shows the averaged values (see *Comment 2.1*) of $\lambda_i, i = 1, \dots, 4$ resulting during the optimization process. All $\lambda_i, i = 1, \dots, 4$ are initialized as $\lambda_i = 1, i = 1, \dots, 4$.

The two NAFs are performed, one (denoted as NAFI) is with the gain (20) s.t. $\Theta = Id, \Lambda = Id$, and the other (denoted by NAFOI) s.t. $\Theta = Id$ and $\lambda_i, i = 1, \dots, 4$ obtained by solving (21) (their values are those displayed at the end of the
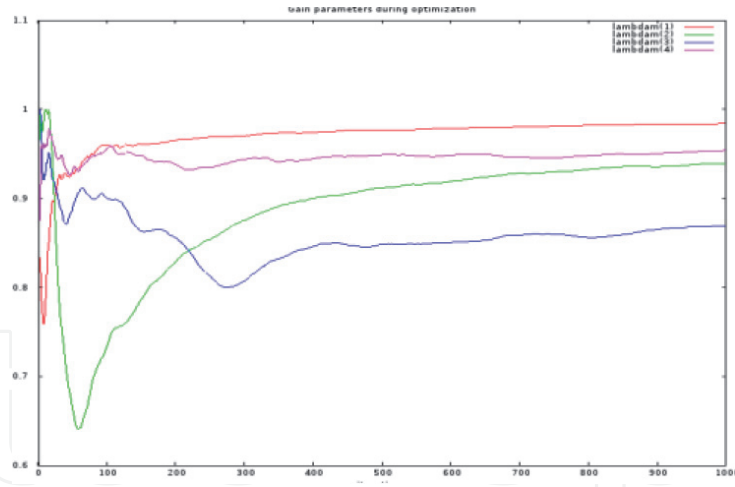
Figure 9.
Control parameters $\lambda_i, i = 1, \ldots, 4$ during optimization process.

optimization process in **Figure 9**). The performances of these two NAFs are shown in **Figure 10**. One sees here that the NAFOI has improved considerably the quality of estimates of the velocity u-component compared with the NAFI. This result justifies that offline optimization (21) is an interesting strategy for finding the optimal initial gain in the NAF.

6.3.2 Estimating the ECM of ME

In practice, for real operational systems, information on the space of ME is not available or very poorly known. Usually, there is a big difference between the model and the real physical process and if the ME statistics are taken more or less properly, in some way, in the filtering algorithm, one can improve the filter performance and reduce the estimation error.

This idea is tested here by applying the HME in Section 4. We carry out the procedure for estimating the ECM of the ME by first constructing the subspace for the ME. For more details on the structure of the ECM M in the AF, see [23]. According to [23], the ECM M is assumed to be of the structure $M = M_v \otimes M_h$—the Kronecker product of M_h with M_v where M_h is the ECM of the horizontal variable, M_v—ECM with vertical variable. **Figure 11** displays RMS of FE for the u velocity
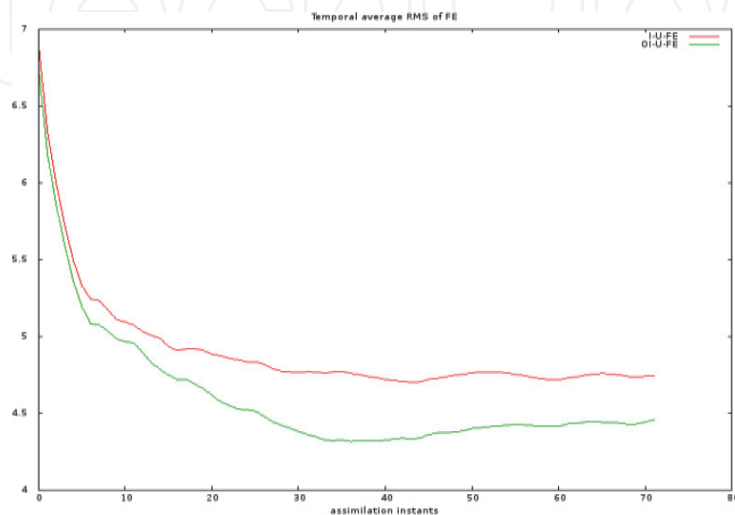

Figure 10.
Temporal average RMS of FE for total velocity u-component produced by the NAF s.t. the initial gain (red curve) and the NAF s.t. optimal initial gain resulting from solving (22) (green curve).
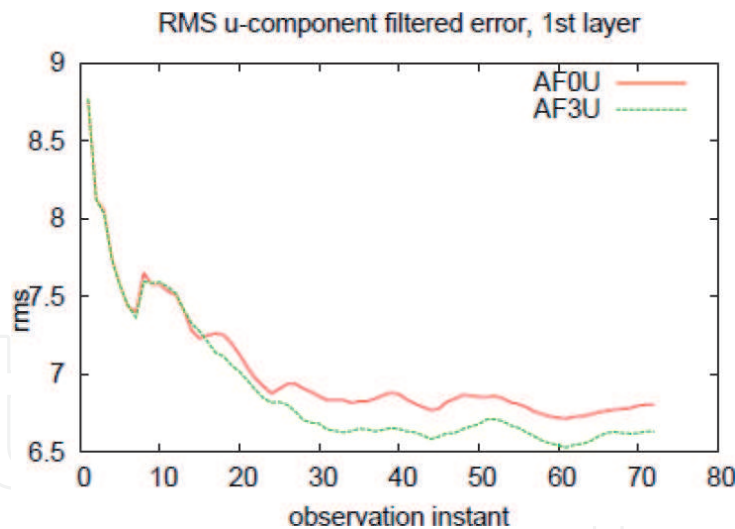
Figure 11.
Performance of the AF: (i) AF0U—no ME ECM has been taken into account; (ii) AF3U—with ME ECM computed in accordance with the HME.

component at the surface resulting from two AFs. The curve *AF0U* corresponds to the AF whose nonadaptive version has the gain computed on the basis of the ECM *M* using an ensemble of PE samples (generated by the PeSP in [18]). The curve *AF3U* shows the performance of the AF with the modified ECM (by adding the vertical ME covariance Q_v to the vertical ECM M_v). More precisely, Q_v is assumed to belong to the subspace spanned by three leading EiVs of M_v. This choice is justified by the fact: the eigenvalue decomposition of M_v has the first three EiVs with the explained variances 67, 17, 15%, respectively. As the fourth EiVec has only the explained variance 0.7E-07%, it is dropped from the subspace constructed for the vertical ME. The better performance of the *AF3U*, in comparison with that of the *AF0U*, is apparently seen in **Figure 11**.

The above experiment shows in details how, on the basis of HME, the subspace for the ME can be constructed, and how one estimates the ECM for the model error. The superior performance of the *AF3U* over that of the *AF0U* validates the usefulness of the HME which can serve as an important tool for estimating the ME and improving the performance of the AF for solving the data assimilation problems with HdSs.

7. Conclusions

One of the key assumptions to ensure the optimal performance of the KF is that a priori knowledge of the system model is given without any uncertainty. This assumption, however, is never valid in practice for dynamical systems under consideration. The uncertainties exist everywhere in modeling a real process like structural uncertainty, model parameterization, model resolution, model bias or ME statistics. For HdSs, order reduction introduced either in the original numerical model or in the filtering algorithms, inevitably leads to uncertainty in the ME, especially in geophysical numerical models.

Our focus in this chapter is to show how the AF solves efficiently filtering problems for systems operating in an uncertain environment.

As seen from this chapter, the AF has proven to be efficient to deal with uncertainties in the specification of the ME statistics, system bias or model reduction. The reasons of the success of the AF are that (i) it belongs to the class of

parametrized stable filters; (ii) it is defined as the best member minimizing mean PE for the system outputs; (iii) The tuning parameters are chosen as elements of stabilizing gain and they are of no physical sense.

It is obvious from this chapter that the performance of the AF is comparable with that of the KF when perfect knowledge of all ME statistics is given, and it outperforms the KF in presence of uncertainties. This happens since the AF acquires knowledge during assimilation process, regardless of uncertainties existing in the filtering problems. From the computational point of view, implementation of the AF consumes much less memory and computational time than the KF or other assimilation methods.

Simple numerical examples and simulation results, presented in Sections 5 and 6, clearly demonstrate the advantages gained through application of the AF in dealing with uncertainties. These positive results encourage a wide application of the AF in different fields of technology and applied sciences like automatic control, finance, aerospace, space exploration, meteorology, and oceanography. A more in-depth and significant research on the capacity of the AF to deal with uncertainties is surely a challenge for the near future.

Author details

Hong Son Hoang* and Remy Baraille
REC/HOM/SHOM, Toulouse, France

*Address all correspondence to: hhoang@shom.fr

IntechOpen

References

[1] Kalman REA. New approach to linear filtering and prediction problems. Journal of Basic Engineering. 1960;**82**: 35-45. DOI: 10.1115/1.3662552

[2] Kailath T, Sayed AH, Hassibi B. Linear Estimation. NJ, Upper Saddle River: Prentice-Hall; 2000

[3] Liptser RS, Shiryaev AN. Statistics of Random Processes—I. General Theory. Berlin and Heidelberg: Springer-Verlag; 2001

[4] Sayed AH. Fundamentals of Adaptive Filtering. NJ: Wiley; 2003

[5] Kucera V. The discrete Riccati equation of optimal control. Kybernetika. 1972;**8**(5):430-447

[6] Hoang HS, Talagrand O, Baraille R. On the design of a stable adaptive filter for high dimensional systems. Automatica. 2001;**37**:341-359

[7] Simon D. Optimal State Estimation. Hoboken, NJ: John Wiley and Sons; 2006. ISBN: 978-0-471-70858-2

[8] Gustafsson F, Hendeby G. Some relations between extended and unscented Kalman filters. IEEE Transactions on Signal Processing. 2012; **60**(2):545-555

[9] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics. 2003;**53**(4):343-367. DOI: 10.1007/s10236-003-0036-9

[10] Chen Y, Snyder C. Assimilating vortex position with an ensemble Kalman filter. Monthly Weather Review. 2007;**135**(5):1828-1845. DOI: 10.1175/MWR3351.1

[11] Del Moral P. Non linear filtering: Interacting particle solution. Markov Processes and Related Fields. 1996;**2**(4): 555-580

[12] Fitzgerald R. Divergence of the Kalman filter. IEEE Transactions on Automatic Control. 1971;**16**(6):736-747

[13] Hoang HS, De Mey P, Talagrand O, Baraille R. A new reduced-order adaptive filter for state estimation in high dimensional systems. Automatica. 1997;**33**(8):1475-1498

[14] Polyak BT. New method of stochastic approximation type. Automation and Remote Control. 1990; **51**(7):937-946

[15] Spall JC. Introduction to Stochastic Search and Optimization. New Jersey: Wiley; 2003. ISBN 978-0-471-33052-3

[16] Hoang HS, Baraille R. Stochastic simultaneous perturbation as powerful method for state and parameter estimation in high dimensional systems. In: Baswell AR, editor. Advances in Mathematics Research. Nova Science Publishers; 2015;**20**:117-148. ISBN: 978-1-63482-741-6c

[17] Hoang HS, Baraille R. A comparison study on performance of an adaptive filter with other estimation methods for state estimation in high-dimensional system. In: Hokimoto T, editor. Advances in Statistical Methodologies and their Application to Real Problems. Rijeka, Croatia: IntechOpen; 2017. pp. 29-52. DOI: 10.5772/67005

[18] Hoang HS, Baraille R. Prediction error sampling procedure based on dominant Schur decomposition. Application to state estimation in high dimensional oceanic model. Journal of Applied Mathematics and Computing. 2011;**218**(7):3689-3709

[19] Hoang HS, Baraille R. On estimation of model error by an adaptive filter. WSEAS Transactions on Systems and Control. 2019;**14**:158-168

[20] B. Friedland B. Treatment of bias in recursive filtering. IEEE Transactions on Automatic Control. 1969;**14**:359-367

[21] Albert AE. Regression and the Moore-Penrose Pseudo-Inverse. London: AP; 1972

[22] Hoang HS, Baraille R, Talagrand O. On an adaptive filter for altimetric data assimilation and its application to a primitive equation model, MICOM. Tellus 57A; 2005:153-170. DOI: 10.1111/ j.1600-0870.2005.00094.x

[23] Hoang HS, Baraille R. On the efficient low cost procedure for estimation of high-dimensional prediction error covariance matrices. Automatica. 2017;**83**:317-330. DOI: 10.1016/j.automatica.2017.06.018