Kalman filtering in hydrological modelling

DAIHM Technical Report 2003-1

Jean-Philippe Drécourt
jpd@dhi.dk

DHI Water & Environment
Agern All 11
2970 Hørsholm
Denmark

May 20, 2003

Contents

Notations

The notations in this report are based on [24], that attempts to provide a unified notation for the data assimilation field. The following is a summary. The meaning of the variables is explained as they are presented in the report.

Vectors (bold lower case Roman)

$\mathbf{d}$	Innovation vector
$\mathbf{x}$	State vector
$\delta\mathbf{x}$	Incremental state vector
$\mathbf{y}^{\mathrm{o}}$	Observations
$\mathbf{y}$	Estimated observation values calculated from the state vector $\mathbf{x}$
η	Model error
ε	Observational error

Operators (upper-case Italic)

H	Observation operator
J	Cost function
M	Model operator

Matrix(bold upper-case Roman)

$\mathbf{H}$	Linearized observation operator
$\mathbf{I}$	Identity matrix
$\mathbf{K}$	Gain matrix
$\mathbf{M}$	Model operator
$\mathbf{P}$	Forecast error covariance
$\mathbf{Q}$	Model error covariance
$\mathbf{R}$	Observational error covariance

Superscripts

$(\cdot)^{-1}$	Inverse
$(\cdot)^{\mathrm{T}}$	Transpose
$(\cdot)^{\mathrm{a}}$	Analysis
$(\cdot)^{\mathrm{b}}$	Background
$(\cdot)^{\mathrm{f}}$	Forecast
$(\cdot)^{\mathrm{o}}$	Observed
$(\cdot)^{\mathrm{t}}$	True

Subscripts

$(\cdot)_i$	Discrete time index i or space index on a finite-difference grid
$(\cdot)_{(i)}$	Refers to the ith column in a covariance matrix

Chapter 1

Introduction

Data assimilation stems from the need to improve the output of our models. With the evolution of sciences, the discovery of Quantum mechanics and the uncertainty principle of Heisenberg, it became obvious that it would be impossible to create models that would reproduce to the perfection the behaviour of nature.

Even though the computers become everyday more powerful, they cannot cope with the complexity of the world, and especially our inability to capture all the details of the system we want to model. As a user of the models, we have to deal with those imperfections and try to correct them as efficiently as possible. In the simplest cases, a review of the conceptualization of the system to be modeled, combined with a thorough calibration of the model lead to results that are acceptable.

In the most complex cases, there is a need for real life data to correct the model behaviour. This is what data assimilation is about: combining model predictions and real world data to make a better estimate of the state of the system we want to model.

The Kalman filter [30] is the most well known sequential data assimilation scheme. It has been developed in the sixties by R. E. Kalman to try to solve the Wiener problem in a generally easier way. The filter has the advantage to be sequential. It needs only the system variables of the previous time step and the forcing terms and observations of the current time step.

This report reviews the main theoretical aspects of the Kalman filter: its derivation for linear systems, the extension to non-linear systems, the development of suboptimal schemes that have lower computational demands and the adaptation of the filter to systems that are biased. The last chapter is dedicated to the application of Kalman filtering to hydrological processes.

Chapter 2

A brief introduction to estimation theory

2.1 Probabilistic Estimation

The goal of this section is to review some of the approaches to estimate a n-vector $\mathbf{x}$ representing a given system, given p-vector $\mathbf{y}^{\mathrm{o}}$ of observations of the system. The estimate is referred to as $\hat{\mathbf{x}}$ and the estimation error as $\eta = \hat{\mathbf{x}} - \mathbf{x}$. Because of the imperfection of the observations, they are represented as random variables. The state vector is also considered as one realization of a random variable.

Different approaches can be taken to derive the estimate [44]. They are described in the following sections.

2.1.1 Bayesian approach

The Bayesian approach to estimation theory is based on the minimization of a function $\mathcal{J}$, that represents the risk involved in the estimation of the state $\mathbf{x}$. It is defined as the mean of a cost function J.

$$\mathcal{J}(\hat{\mathbf{x}}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} J(\eta) p_{\mathbf{x}\mathbf{y}^{\mathrm{o}}}(\mathbf{x}, \mathbf{y}^{\mathrm{o}}) d\mathbf{x} d\mathbf{y}^{\mathrm{o}} \tag{2.1}$$

$p_{\mathbf{x}\mathbf{y}^{\mathrm{o}}}$ is the joint probability density of the random variables $\mathbf{x}$ and $\mathbf{y}^{\mathrm{o}}$. The different way of solving this problem are based on the choice of the cost function.

In most cases, one wants the estimate to be unbiased, i.e. its expected value is equal to the expected value of the state.

$$\mathrm{E}\{\hat{\mathbf{x}}\} = \mathrm{E}\{\mathbf{x}\} \tag{2.2}$$

2.1.2 Minimum variance estimate

The minimum variance estimate minimizes the risk function based on the cost function:

$$J(\eta) = \eta^{\mathrm{T}} \mathbf{\Sigma} \eta \tag{2.3}$$

where the $n \times n$ matrix $\mathbf{\Sigma}$ is symmetric, non-negative.

The minimum of the risk function $\mathcal{J}$ is found for the value of $\hat{\mathbf{x}}^{\mathrm{MV}}$ (see [44] for the derivation of the solution):

$$\hat{\mathbf{x}}^{\mathrm{MV}} = \mathrm{E}\left\{\mathbf{x} \,|\, \mathbf{y}^{\mathrm{o}}\right\} \tag{2.4}$$

One can prove that this estimate is unbiased.

2.1.3 Maximum *a posteriori* probability estimation

The maximum *a posteriori* probability estimate is based on uniform cost function, described in any meaningful norm as:

$$J(\eta) = \begin{cases} 0, & \|\eta\| < \zeta \\ 1/2\zeta, & \|\eta\| \geq \zeta \end{cases} \qquad \zeta > 0 \tag{2.5}$$

The minimization of the risk function is equivalent (c.f. [44]) to maximizing the *a posteriori* probability distribution of $\mathbf{x}$, i.e. solving for $\mathbf{x}$:

$$\frac{\partial p_{\mathbf{x}\,|\,\mathbf{y}^{\mathrm{o}}}}{\partial \mathbf{x}}(\mathbf{x} \,|\, \mathbf{y}^{\mathrm{o}}) = 0 \tag{2.6}$$

After taking the natural logarithm and using Bayes' theorem, it is equivalent to solve:

$$\frac{\partial \left[\ln(p_{\mathbf{y}^{\mathrm{o}}\,|\,\mathbf{x}}(\mathbf{y}^{\mathrm{o}} \,|\, \mathbf{x})) + \ln(p_{\mathbf{x}}(\mathbf{x}))\right]}{\partial \mathbf{x}} = 0 \tag{2.7}$$

In general, the estimate is not necessarily unbiased.

2.1.4 Maximum likelihood estimation

If there is no prior information about the random variable, then its variance is infinite and therefore the following equation is valid for any $\mathbf{x}$:

$$\frac{\partial \ln p_{\mathbf{x}}(\mathbf{x})}{\partial \mathbf{x}} = 0 \tag{2.8}$$

It simplifies the equation 2.7 to:

$$\frac{\partial p_{\mathbf{y}^{\mathrm{o}}\,|\,\mathbf{x}}(\mathbf{y}^{\mathrm{o}}\,|\,\mathbf{x})}{\partial \mathbf{x}} = 0 \tag{2.9}$$

Both the maximum likelihood and the maximum a-posteriori estimates are mode estimates. It is therefore important to know which distribution is maximized. Moreover it implies that the estimate is not necessarily unbiased.

2.2 Least squares estimation

The least squares estimation method has the advantage not to rely on any statistical assumption. It just tries to get the best least squares fit among the data.

In the linear case, we can define a relation between the measurements and the state of the system for one experiment:

$$\mathbf{y}_i^{\mathrm{o}} = \mathbf{H}_i \mathbf{x} + \varepsilon_i \tag{2.10}$$

By collecting the information on k experiments, it is possible to write the set of equations in a compact form. Assuming that $\mathbf{y}_i^{\mathrm{o}}$ is a q_i vector, then we can define the $\tilde{q}$ vector $\tilde{\mathbf{y}}^{\mathrm{o}}$, with $\tilde{q} = \sum_{i=1}^{k} q_i$:

$$\tilde{\mathbf{y}}^{\mathrm{o}} = \tilde{\mathbf{H}}\mathbf{x} + \tilde{\varepsilon} \tag{2.11}$$

where

$$\tilde{\mathbf{y}}^{\mathrm{o}} \equiv \left[\mathbf{y}_1^{\mathrm{oT}} \cdots \mathbf{y}_k^{\mathrm{oT}}\right]^{\mathrm{T}} \tag{2.12}$$

$\tilde{\varepsilon}$ is also a $\tilde{q}$-vector

$$\tilde{\varepsilon} \equiv \left[\varepsilon_1^{\mathrm{T}} \cdots \varepsilon_k^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.13}$$

and $\tilde{\mathbf{H}}$ is $\tilde{q} \times n$ matrix:

$$\tilde{\mathbf{H}} \equiv \left[\mathbf{H}_1^{\mathrm{T}} \cdots \mathbf{H}_k^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.14}$$

The goal of least squares estimation is to find an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ that minimizes the cost function:

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\tilde{\mathbf{y}}^{\mathrm{o}} - \tilde{\mathbf{H}}\mathbf{x}\right)^{\mathrm{T}} \tilde{\mathbf{O}}^{-1} \left(\tilde{\mathbf{y}}^{\mathrm{o}} - \tilde{\mathbf{H}}\mathbf{x}\right) \tag{2.15}$$

The weighting matrix $\tilde{\mathbf{O}}^{-1}$ is symmetric positive definite and represents the confidence in each experiment.

It leads (c.f. [44]) to the following estimate:

$$\hat{\mathbf{x}} = \mathbf{P}\tilde{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{O}}^{-1}\mathbf{y}^{o} \tag{2.16}$$

with

$$P = \left(\tilde{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{O}}^{-1}\tilde{\mathbf{H}}\right)^{-1} \tag{2.17}$$

It is assumed that the inverse exists.

The least squares estimate is equivalent to the minimum variance estimate when one chooses $\mathbf{O} = \mathbf{R}$, where $\mathbf{R}$ is the observational error covariance, the covariance of the random variable ε.

Chapter 3

The linear Kalman filter

The Kalman filter is a minimum variance estimation in the special case where the system is a linear stochastic dynamical system. The system will be described by the coupled equations:

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{M}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \eta_k \\ \mathbf{y}_k^o &= \mathbf{H}_k \mathbf{x}_k + \varepsilon_k \end{cases} \tag{3.1}$$

Let n be the dimension of the state vector $\mathbf{x}$ and q the dimension of the observation vector $\mathbf{y}^o$. It is assumed that the number of observations is constant through time.

$\mathbf{u}_k$ is the forcing term. In the rest of the chapter, $\mathbf{B}_k$ is set to zero to simplify the notation, as the influence of the forcing term is purely deterministic and does not affect the estimation process.

This chapter is based on the classical assumptions of the linear Kalman filter. η_k and ε_k are independant sequences of zero-mean white noise, with $cov\{\eta_k\} = \mathbf{Q}_k$ and $cov\{\varepsilon_k\} = \mathbf{R}_k$. The noise sequences are also independent of the initial conditions $\mathbf{x}_0$.

It is possible to derive the Kalman filter using two methods mentioned in the previous chapter, the minimum variance estimation and the maximum a-posteriori estimation.

3.1 Minimum variance estimate

3.1.1 Orthogonal projection and minimum variance

In this section, we will show that the minimum variance estimate of $\mathbf{x}_{k+1}$ given the set of observations $\{\mathbf{y}_0^o, ..., \mathbf{y}_{k+1}^o\}$ is the orthogonal projection of $\mathbf{x}_{k+1}$ onto the space defined by a linear combination of the observations. It

is demonstrated by Kalman in [30] for $q = n$. Chui and Chen in [6] try to extend it to any value of q but do not project properly the observation space onto the state space. The way they define the linear span of the observations (p. 34) leads to the definition of the whole state space, whatever the number of non-zero observations.

We introduce here the Moore-Penrose pseudo-inverse of the observation matrix $\mathbf{H}^{\oplus}$ that projects the observations onto the state space, i.e. $\mathbf{H}^{\oplus}\mathbf{y}^{\mathrm{o}}$ is a n-vector. The results of section 2.1.2 mean also that the minimum variance estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is a linear combination of the observation, it is an element of the ensemble Y defined by:

$$Y = \{\mathbf{y} \ : \ \mathbf{y} = \sum_{i=0}^{r} \beta_i \mathbf{H}^{\oplus}\mathbf{y}_i^{\mathrm{o}}, \ \beta_i \in \mathbb{R}\} \tag{3.2}$$

We define the covariance of two vectors of random variables as:

$$cov(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left\{(\mathbf{x} - \mathrm{E}\left\{\mathbf{x}\right\})(\mathbf{y} - \mathrm{E}\left\{\mathbf{y}\right\})^{\mathrm{T}}\right\} \tag{3.3}$$

And the variance as $var(\mathbf{x}) = cov(\mathbf{x}, \mathbf{x})$.

Being a minimum variance estimate, $\hat{\mathbf{x}}$ minimizes its variance, which is the same as minimizing its trace. The following theorem shows that the minimum variance estimate of $\mathbf{x}$ is its orthogonal projection onto Y

Theorem 3.1 *$\hat{\mathbf{x}} \in Y$ minimizes $tr(var(\mathbf{x} - \hat{\mathbf{x}}))$ if and only if $cov(\mathbf{x} - \hat{\mathbf{x}}, \mathbf{y}_i^{\mathrm{o}}) = \mathbf{O}$, where $\mathbf{O}$ is a $n \times q$ matrix of zeros. Moreover $\hat{\mathbf{x}}$ is unique*

If we suppose that there exist an observation $\mathbf{y}_{j0}^{\mathrm{o}}$ so that $tr(var(\mathbf{x} - \hat{\mathbf{x}}, \mathbf{y}_{j0}^{\mathrm{o}})) = \mathbf{C}$ with $\mathbf{C}$ non-zero, then we can try to find a vector that has a lower variance trace. This vector is probably in the same direction as $\mathbf{y}_{j0}^{\mathrm{o}}$. So we can calculate the trace of $var(\mathbf{x} - \hat{\mathbf{x}} - \alpha\mathbf{H}^{\oplus}\mathbf{y}_{j0}^{\mathrm{o}}), \ \alpha \in \mathbb{R}$. The vector $\hat{\mathbf{x}} - \alpha\mathbf{H}^{\oplus}\mathbf{y}_{j0}^{\mathrm{o}}$ is in Y as a linear combination of elements of Y. Replacing the known values, we get:

$$\begin{aligned} & tr(var(\mathbf{x} - \hat{\mathbf{x}} - \alpha\mathbf{H}^{\oplus}\mathbf{y}_{j0}^{\mathrm{o}})) \\ = \ & tr(var(\mathbf{x} - \hat{\mathbf{x}}) + \alpha^2\mathbf{H}^{\oplus}\mathbf{H}^{\oplus\mathrm{T}}var(\mathbf{y}_{j0}^{\mathrm{o}}) - \alpha\mathbf{C}\mathbf{H}^{\oplus\mathrm{T}} - \alpha\mathbf{H}^{\oplus}\mathbf{C}^{\mathrm{T}}) \end{aligned} \tag{3.4}$$

To prove the theorem, we need to find at least one α so that the trace of $\alpha^2\mathbf{H}^{\oplus}\mathbf{H}^{\oplus\mathrm{T}}var(\mathbf{y}_{j0}^{\mathrm{o}}) - \alpha C\mathbf{H}^{\oplus\mathrm{T}} - \alpha\mathbf{H}^{\oplus}\mathbf{C}^{\mathrm{T}}$ is negative.

Knowing that:

- The trace is a linear operator

- $\mathbf{H}^{\oplus}\mathbf{H}^{\oplus \mathrm{T}}$ is a symetric, non-negative matrix, so its trace is strictly positive
- $tr(\mathbf{A}^{\mathrm{T}}) = tr(\mathbf{A})$

We can define any α:

$$\alpha < \frac{2tr(\mathbf{C}\mathbf{H}^{\oplus \mathrm{T}})}{tr(\mathbf{H}^{\oplus}\mathbf{H}^{\oplus \mathrm{T}} var(\mathbf{y}^{\mathrm{o}}_{j_0}))} \tag{3.5}$$

that verifies $tr(var(\mathbf{x} - \hat{\mathbf{x}} - \alpha\mathbf{H}^{\oplus}\mathbf{y}^{\mathrm{o}}_{j_0})) < tr(var\mathbf{x} - \hat{\mathbf{x}})$. It contradicts the definition of $\hat{\mathbf{x}}$.

The uniqueness of $\hat{\mathbf{x}}$ is derived in [6]. □

It is then possible to define an orthonormal base on Y, $\mathbf{e}_1, ..., \mathbf{e}_s$, with $s \leq r$, and to express $\hat{\mathbf{x}}$ by[1]:

$$\hat{\mathbf{x}} = \sum_{i=1}^{s} cov(\mathbf{x}, \mathbf{e}_i)\mathbf{e}_i \tag{3.6}$$

It is important to notice that there is here no assumption about the Gaussianity of the variables.

3.1.2 The Kalman filter

The derivation of the Kalman filter follows easily the results demonstrated in the previous section, and detailed in [6]. We will just summarize the algorithm here.

At time step 0, it is necessary to set up the covariance matrix $\mathbf{P}^{\mathrm{a}}_0$ of the initial conditions $\mathbf{x}^{\mathrm{a}}_0$. The model propagation is given by:

$$\mathbf{x}^{\mathrm{f}}_k = \mathbf{M}_{k-1}\mathbf{x}^{\mathrm{a}}_{k-1} \tag{3.7}$$

$$\mathbf{P}^{\mathrm{f}}_k = \mathbf{M}_{k-1}\mathbf{P}^{\mathrm{a}}_{k-1}\mathbf{M}^{\mathrm{T}}_{k-1} + \mathbf{Q}_{k-1} \tag{3.8}$$

The innovation vector is defined by:

$$\mathbf{d}_k = \mathbf{y}^{\mathrm{o}}_k - \mathbf{H}_k\mathbf{x}^{\mathrm{f}}_k \tag{3.9}$$

The analysis step is:

$$\mathbf{x}^{\mathrm{a}}_k = \mathbf{x}^{\mathrm{f}}_k + \mathbf{K}_k \cdot \left(\mathbf{y}^{\mathrm{o}}_k - \mathbf{H}_k\mathbf{x}^{\mathrm{f}}_k\right) \tag{3.10}$$

[1] In [6], the orthonormal base has as many vectors as the number of observations, which is the case only if the observations are completely uncorrelated.

and

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\, \mathbf{P}_k^{\mathrm{f}} \tag{3.11}$$

The Kalman gain is defined by:

$$\mathbf{K}_k = \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} \left[\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} \right]^{-1} \tag{3.12}$$

The analysis step is in fact a linear combination of the observations and the model estimate. If the model estimate is more certain than the observation, i.e. $\mathbf{P}_k^{\mathrm{f}} \ll \mathbf{R}_k$, then the gain is close to zero and $\mathbf{x}_k^{\mathrm{a}} \to \mathbf{x}_k^{\mathrm{f}}$. In case $\mathbf{P}_k^{\mathrm{f}} \gg \mathbf{R}_k$, then the gain is close to one, and the analysis is close to the observations.

3.2 Maximum a-posteriori estimate

In this section, we attempt to retrieve the Kalman filter from another approach: determining the MAP of $\mathbf{x}_k^{\mathrm{f}}$, given the set of measurements $(\mathbf{y}_0^{\mathrm{o}}, ..., \mathbf{y}_k^{\mathrm{o}})$. We have to solve the following equation:

$$\left. \frac{\partial f(\mathbf{x}_k^{\mathrm{f}} \,|\, \mathbf{y}_0^{\mathrm{o}}, ..., \mathbf{y}_k^{\mathrm{o}})}{\partial \mathbf{x}_k^{\mathrm{f}}} \right|_{\mathbf{x}_k^{\mathrm{f}} = \mathbf{x}_k^{\mathrm{a}}} = 0 \tag{3.13}$$

The measurements from previous time steps are not dependent on $\mathbf{x}_k^{\mathrm{f}}$, so it is equivalent to solve:

$$\left. \frac{\partial f(\mathbf{x}_k^{\mathrm{f}} \,|\, \mathbf{y}_k^{\mathrm{o}})}{\partial \mathbf{x}_k^{\mathrm{f}}} \right|_{\mathbf{x}_k^{\mathrm{f}} = \mathbf{x}_k^{\mathrm{a}}} = 0 \tag{3.14}$$

Following the results of equation 2.7, we have to solve:

$$\left. \frac{\partial \ln(f(\mathbf{y}_k^{\mathrm{o}} \,|\, \mathbf{x}_k^{\mathrm{f}}))}{\partial \mathbf{x}_k^{\mathrm{f}}} \right|_{\mathbf{x}_k^{\mathrm{f}} = \mathbf{x}_k^{\mathrm{a}}} + \left. \frac{\partial \ln(f(\mathbf{x}_k^{\mathrm{f}}))}{\partial \mathbf{x}_k^{\mathrm{f}}} \right|_{\mathbf{x}_k^{\mathrm{f}} = \mathbf{x}_k^{\mathrm{a}}} = 0 \tag{3.15}$$

Without any assumption about the distributions, the problem cannot be solved in general. In this case, it is assumed that the two distributions are Gaussian, and written:

$$f(\mathbf{y}_k^{\mathrm{o}} \,|\, \mathbf{x}_k^{\mathrm{f}}) = \alpha \exp \left[\left(\mathbf{y}_k^{\mathrm{o}} - \mathbf{H}_k \mathbf{x}_k^{\mathrm{f}} \right)^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathbf{y}_k^{\mathrm{o}} - \mathbf{H}_k \mathbf{x}_k^{\mathrm{f}} \right) \right] \tag{3.16}$$

$$f(\mathbf{x}_k^{\mathrm{f}}) = \beta \exp\left[\left(\mathbf{x}_k^{\mathrm{f}} - \mathrm{E}\left\{\mathbf{x}_k^{\mathrm{f}}\right\}\right)^{\mathrm{T}} \mathbf{P}_k^{\mathrm{f}\,-1} \left(\mathbf{x}_k^{\mathrm{f}} - \mathrm{E}\left\{\mathbf{x}_k^{\mathrm{f}}\right\}\right)\right] \tag{3.17}$$

α and β are normalization constants. $\mathbf{R}_k$ is the observation noise covariance at time step k, $\mathbf{P}_k^{\mathrm{f}}$ is the forecast error covariance before correction by the filter.

After derivation, one can find the same result as in previous section.

3.3 Comparison between the two methods

Even though the two methods lead to the same result, the assumptions to reach the result are different: in the case of the MAP estimate, it is necessary that the variables follow a Gaussian probability distribution. For the minimum variance estimate, this is not necessary.

As pointed out in [30], the two approaches are fundamentally similar. It is the minimization of a cost function of the error between the *model* estimate and the *observed* value. In the case of the minimum variance, the cost function is related to the square of the difference. As Kalman explains in his article [30], it is not then necessary to deal with Gaussian distributions. In the case of the MAP, the cost function is not specified. It demands then the use of Gaussian functions. In a nutshell, if the distribution is Gaussian, one gets both the minimum variance and the MAP estimate. If the distribution is not Gaussian, one gets only a minimum variance estimate, regardless of the a-priori distributions.

As an additional remark, it is interesting to mention that the linear Kalman filter is equivalent to the co-krigging of the model forecast error given the observations [42]. This geostatistical formulation gives a better insight of the behavior of the filter in situations where the state vectors are random fields as found in hydrology: the corrections of the model forecast are calculated at the points where observations are available. Then they are propagated over the whole field according to the correlation structure.

3.4 Kalman filter and colored noise

It is possible to derive the Kalman filter for model and observation noises that are correlated in time. See e.g. chapter 5 in [6]. The main results are summarized below.

Given the noise models as:

$$\begin{cases} \eta_k = \mathbf{\Lambda}_k \eta_{k-1} + \mu_k \\ \varepsilon_k = \mathbf{\Gamma}_k \varepsilon_{k-1} + \varphi_k \end{cases} \tag{3.18}$$

with the initial conditions $\eta_{-1} = 0$, $\varepsilon_{-1} = 0$, $\mathbf{\Lambda}_k$ and $\mathbf{\Gamma}_k$ constant matrices defining how the structure of the noise is propagated through time. The noise sequences μ_k and φ_k are uncorrelated Gaussian white noise with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively.

To make the system white, the noise vector is concatenated with the state vector. The model operator becomes:

$$\begin{bmatrix} \mathbf{M}_k & \mathbf{I} \\ 0 & \mathbf{\Lambda}_k \end{bmatrix}$$

Because of the presence of colored noise in both the observation and the model propagation, it is necessary to derive a recursive relationship between the analyzed values of the state vector (see the derivation in [6]):

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_k \\ \eta_k \end{bmatrix} &= \begin{bmatrix} \mathbf{M}_k & \mathbf{I} \\ 0 & \mathbf{\Lambda}_k \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1} \\ \eta_{k-1} \end{bmatrix} \\ &\quad + \mathbf{K}_k \left(\mathbf{y}_k^{\mathrm{o}} - \mathbf{\Gamma}_k \mathbf{y}_{k-1}^{\mathrm{o}} - \mathbf{H}'_k \begin{bmatrix} \mathbf{x}_{k-1} \\ \eta_{k-1} \end{bmatrix} \right) \end{aligned} \tag{3.19}$$

The observation operator is defined as:

$$\mathbf{H}'_k = [\mathbf{H}_k \mathbf{M}_k - \mathbf{\Gamma}_k \mathbf{H}_{k-1} \quad \mathbf{H}_k] \tag{3.20}$$

The Kalman gain is:

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{M}_k & \mathbf{I} \\ 0 & \mathbf{\Lambda}_k \end{bmatrix} \mathbf{P}'_{k-1} \mathbf{H}_k'^{\mathrm{T}} \left(\mathbf{H}'_k \mathbf{P}'_{k-1} \mathbf{H}_k'^{\mathrm{T}} + \mathbf{R}_k \right)^{-1} \tag{3.21}$$

And the covariance matrix of the analyzed concatenated state vector is calculated by:

$$\mathbf{P}'_k = \left(\begin{bmatrix} \mathbf{M}_k & \mathbf{I} \\ 0 & \mathbf{\Lambda}_k \end{bmatrix} - \mathbf{K}_k \mathbf{H}'_k \right) \mathbf{P}'_{k-1} \begin{bmatrix} \mathbf{M}_k^{\mathrm{T}} & 0 \\ \mathbf{I} & \mathbf{\Lambda}_k^{\mathrm{T}} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{Q}_k \end{bmatrix} \tag{3.22}$$

The initial conditions for the covariance matrix are:

$$\mathbf{P}'_0 = \begin{bmatrix} \left(var(\mathbf{x}_0)^{-1} + \mathbf{H}_0'^{\mathrm{T}} \mathbf{R}_0^{-1} \mathbf{H}'_0 \right)^{-1} & 0 \\ 0 & \mathbf{Q}_0 \end{bmatrix} \tag{3.23}$$

These equations give back the original Kalman filter equations when the operators $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are set to zero. If there is only colored noise in the model error, i.e. $\mathbf{\Gamma} = 0$, then it is possible to express the Kalman filter in its forecast/analysis form, using the same augmented state vector as defined here.

3.5 Bias aware Kalman filter (BAKF)

The drawback of the Kalman filter is that it assumes that the model estimates are unbiased. In the colored noise method, the bias can be defined as a noise with a high autocorrelation. As stated in [17], this approach is feasible as long as the dimension of the bias term is small compared to the state variables. It becomes computationally difficult to deal with a bias term of similar dimension as the state vector.

Two methods have been proposed to derive the Kalman filter in presence of bias. The first method, developed in [17] starts from the extended state vector to derive a method where the bias estimate is simply an add-on to the normal Kalman filter. It has been extended to more general cases in [35, 25, 26, 55, 27]. In [10], the same equations are derived by studying the difference between the observed data and the model forecast. The interest of the original implementation of the bias aware Kalman filter is the possibility to use it as an add-on to the original Kalman filter in case a bias problem is noticed. This section will summarize the results of these studies.

3.5.1 Where to put the bias?

Bias is defined in reference to a state of the system that is supposed to be unbiased. According to the knowledge available, it is possible to define the bias relative to unbiased measurements, or to unbiased model. Unless the dynamics of the bias are very well known, it is difficult to assess the bias on both the measurements and the model. In this derivation, it is assumed that the measurements or at least a subset of the measurements are unbiased and that the bias is entirely generated by the model.

3.5.2 Three different implementations

According to the degree of integration of the bias estimation in the whole filter algorithm, three implementation are possible.

Bias blind estimator This implementation is fully off-line: the bias is calculated independently of the original Kalman filter.

On-line bias estimate This implementation uses the unbiased state forecast to make the analysis but the input to the next time step of the model is the *biased* analysis.

On-line bias estimate with feedback The bias is calculated as before but the input to the model is the *unbiased* analysis.

3.5.3 Bias blind estimator

The forecast and analysis of the state variable are the same as with the usual Kalman filter:

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a \tag{3.24}$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f \right) \tag{3.25}$$

with the gain defined as

$$\mathbf{K}_k = \mathbf{S}_k^f \mathbf{H}_k^T \left[\mathbf{H}_k \mathbf{S}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1} \tag{3.26}$$

$$\mathbf{S}_k^f = \mathbf{M}_k \mathbf{S}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k \tag{3.27}$$

$$\mathbf{S}_k^a = \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right] \mathbf{S}_k^f \tag{3.28}$$

It is important to notice that the matrices $\mathbf{S}$ are the actual estimation error covariances only in the case when there is no bias. In the general case, they are smaller than the actual error covariance matrices (in a meaningful norm) as the estimates in the absence of bias are more accurate than the estimates with unknown bias. If $\mathbf{T}_k$ is the bias prediction error covariance matrix then according to [10]:

$$\mathbf{S}_k^f = \mathbf{P}_k^f - \mathbf{T}_k \tag{3.29}$$

If the bias propagation model is the constant model,

$$\mathbf{b}_k^f = \mathbf{b}_{k-1}^a \tag{3.30}$$

and the unbiased model estimate is:

$$\widetilde{\mathbf{x}}_k^f = \mathbf{x}_k^f - \mathbf{b}_k^f \tag{3.31}$$

then the analysis of the bias is calculated by:

$$\mathbf{b}_k^a = \mathbf{b}_k^f - \mathbf{L}_k \left[\mathbf{y}_k^o - \mathbf{H}_k \widetilde{\mathbf{x}}_k^f \right] \tag{3.32}$$

The matrix $\mathbf{L}_k$ is defined by:

$$\mathbf{L}_k = \mathbf{T}_k \mathbf{H}_k \left[\mathbf{H}_k \mathbf{T}_k \mathbf{H}_k^T + \mathbf{H}_k \mathbf{S}_k \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1} \tag{3.33}$$

The gain matrix $\mathbf{L}_k$ is the optimal Kalman gain for the system defined by the propagation equation:

$$\mathbf{b}_k^f = \mathbf{b}_{k-1}^f \tag{3.34}$$

and the observation equation:

$$\mathbf{d}_k = -\mathbf{H}_k \mathbf{b}_k^f + \mu_k \tag{3.35}$$

The innovation vector is also defined by:

$$\mathbf{d}_k = \mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f \tag{3.36}$$

and its noise component μ_k is unbiased and has a covariance of:

$$E\left(\mu_k \mu_k^T\right) = \mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \tag{3.37}$$

Pros and cons of this method This approach has the nice interest of being an add-on to the original Kalman filter. An original implementation of the filter does not need to be modified. The bias estimation is used when a bias problem is observed.

The problem of this method is that the original Kalman filter corrects values that are biased, which leads to biased estimates, i.e. not optimal estimates. The optimality of the results could be improved by providing a better estimate of the bias before the correction by the filter.

3.5.4 On-line bias estimate

In this implementation of the filter, there are two analysis values. $\mathbf{x}_k^a$ is the analysis from the biased value, that will be used in the next time step. $\widetilde{\mathbf{x}_k^a}$ is the unbiased estimate.

The algorithm reads:

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a \tag{3.38}$$

$$\mathbf{b}_k^f = \mathbf{b}_{k-1}^a \tag{3.39}$$

$$\widetilde{\mathbf{x}}_k^f = \mathbf{x}_k^f - \mathbf{b}_k^f \tag{3.40}$$

The bias is updated by:

$$\mathbf{b}_k^a = \mathbf{b}_k^f - \mathbf{L}_k \left[\mathbf{y}_k^o - \mathbf{H}_k \widetilde{\mathbf{x}}_k^f\right] \tag{3.41}$$

The two analysis values are then calculated:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left[\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f\right] \tag{3.42}$$

$$\widetilde{\mathbf{x}}_k^a = \mathbf{x}_k^a - \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right] \mathbf{b}_k^a \tag{3.43}$$

The propagation and update of the covariance matrix are identical to the off-line bias estimate.

3.5.5 On-line bias estimate with feedback

In this implementation, the unbiased value is used in the next time step. The bias is then an estimation of the bias introduced by the system at each time step.

The propagation equation reads:

$$\mathbf{x}_k^f = \mathbf{M}_k \widetilde{\mathbf{x}}_{k-1}^a \tag{3.44}$$

and the analysis equation is equivalent to the combination of the equations 3.42 and 3.43.

$$\widetilde{\mathbf{x}}_k^a = \left(\mathbf{x}_k^f - \mathbf{b}_k^a\right) + \mathbf{K}_k \left[\mathbf{y}_k^o - \mathbf{H}_k \left(\mathbf{x}_k^f - \mathbf{b}_k^a\right)\right] \tag{3.45}$$

3.5.6 The covariance estimation problem

Estimating the covariance matrix $\mathbf{P}_k^f$ has always been a problem. Generally speaking, the matrix is defined using some assumptions about the system. Then the covariance is propagated through the system using the model.

For the covariance of the bias, the model propagation is too simple to allow for a proper propagation of the correlation structure.

The paper [10] considers that the propagation of the bias covariance cannot be directly estimated from the simple assumption of a persistent model. It

	BAKF	Colored noise KF
Covariance matrix size	$2n^2$	$4n^2$
Kalman gain size	$2nq$	$2nq$
Model operator matrix size	$n^2 + 1$	$4n^2$

Table 3.1: Comparison of the storage requirements of the BAKF and the colored noise KF given a bias and state vector size of n and an observation vector size of q

is assumed that it should have the same correlation structure as the state and propose to model the bias covariance with:

$$\mathbf{T}_k = \gamma \mathbf{P}_k^f \tag{3.46}$$

where $0 \leq \gamma \leq 1$ is a tuning parameter that controls the stability of the bias estimate. It determines to which extent the information given by the observation should be used to correct the bias parameter, and what remains to correct the random component.

The main issue with this method is the necessity to evaluate this parameter. In [11], γ is chosen using a Fourier analysis of the residual after bias correction and assimilation over a given period of time. The parameter is optimized so that the spectrum is flat, i.e. when the residual uncertainty is close to white noise. This assumes that the data are ergodic (cf. section 6.1.1).

3.6 Colored noise filter against bias aware filter

3.6.1 Storage requirements and computational demands

The BAKF is designed to have smaller storage requirements than the colored noise filter. In table 3.6.1, the sizes of the different matrices are compared. To make the comparison possible, the sizes of the covariance matrices for the state and the bias in the BAKF are summed, as well as the two Kalman gains. The persistent model operator is considered as a scalar of dimension 1×1.

The computational costs of the two methods are similar. Both need $2n + 1$ model evaluations for the propagation of the state vector and the covariance matrix. For the bias propagation, the persistent model of the BAKF demands no extra computation and the autoregressive model of the colored

noise filter demands only one multiplication, which is normally negligible compared to the model evaluation.

3.6.2 Cross-diagonal correlation structure

The main difference between the two implementations of the filter are the way the bias interacts with the original state vector.

In the BAKF, the bias does not influence whatsoever the way the Kalman gain and the covariance matrix of the state vector is calculated. In the implementation with feedback, the state is corrected for bias before being processed for the next time step, but it does not affect the covariance matrices. It is impossible to calculate an expression of an augmented covariance matrix of the concatenated state and bias vectors because one cannot get an expression of the cross-diagonal terms of the covariance.

In comparison, the colored noise implementation does not impose any restriction of the cross-diagonal terms of the augmented covariance matrix, except the ones imposed by the autoregressive model of the noise part of the augmented vector.

Even though the two implementations seem similar, they have a different role:

- The BAKF intends to correct a Kalman filter implementation that would generate analysis errors that are constant in time compared to the dynamics of the system.
- The colored noise Kalman filter is in theory able to represent a system where the innovation is biased. It is then important to notice that the covariance matrix $\mathbf{Q}$ defines two completely different characteristics of the system: the rate of convergence of the colored noise towards a value that is supposed to be constant in time and the uncertainty on the state vector. This dual function can pose some practical problems.

Chapter 4

Towards non-linearity

The original Kalman filter theory is based on linear systems. For non-linear systems, the Kalman filter has been extended, using approximations of the non-linear system in order to use linear algebra. There are two main approaches:

- The extended Kalman filter (EKF), based on the local derivatives of the non-linear system. The description is based on [6], pp 108-111.
- The unscented Kalman filter (UKF), based on the unscented transformation that is supposed to preserve at least the two first moments through a linear system by judiciously choosing the points where the model is evaluated. The section is based on [29, 28, 53].

In this chapter, the system studied is then described as:

$$\left\{ \begin{array}{rcl} \mathbf{x}_k^{\mathrm{f}} & = & M(\mathbf{x}_{k-1}^{\mathrm{a}}, \mathbf{u}_k, \eta_k) \\ \mathbf{y}_k^{\mathrm{o}} & = & H(\mathbf{x}_k^{\mathrm{f}}, \varepsilon_k) \end{array} \right. \tag{4.1}$$

where $\mathbf{u}_k$ is the forcing term of the system. Note that the model and observation operators are not matrices any more but any function, linear or non-linear.

4.1 The extended Kalman filter

The extended Kalman filter is based on the linearization of the model and observation operators in the neighborhood of $(\mathbf{x}_k^{\mathrm{f}}, \mathbf{u}_k, 0)$. It is then possible

to rewrite equations 3.8 to 3.12, replacing the matrices $\mathbf{M}_k$ and $\mathbf{H}_k$ by:

$$\left\{ \begin{array}{rcl} \mathbf{M}_k & = & \left.\frac{\partial M}{\partial \mathbf{x}_k}\right|_{(\mathbf{x}_k^{\mathrm{f}},\mathbf{u}_k,0)} \\ \mathbf{H}_k & = & \left.\frac{\partial H}{\partial \mathbf{x}_k}\right|_{(\mathbf{x}_k^{\mathrm{f}},0)} \end{array} \right. \tag{4.2}$$

The model propagation does not need to be linearized.

This method handles well small non-linearities but is rather inefficient in case of very nonlinear systems as explained in [28]. Moreover the method is not suited for large dimension systems, as the calculation of the derivatives, using a finite difference method, demands $n+1$ model evaluations for each time step (n is the dimension of the state vector), and $q+1$ evaluations of the observation operator (q is dimension of the observation space).

The other possibility is to write a tangent linear model, but it is generally difficult for complex models or impossible for highly non-linear models.

4.2 The unscented Kalman filter

4.2.1 The unscented transform

The unscented transform attempts to solve the problem of the propagation of a probability distribution through a nonlinear transformation. With the EKF, the necessity of linearizing implies a loss of information about the model operator.

We are essentially interested in the propagation of the two first moments of the probability distribution. The unscented transform proposes a way to generate a population from a probability distribution so that the two first moments are propagated accurately. Instead of using a Monte Carlo approach, the points are chosen deterministically.

From [29], a n-dimensional random variable $\mathbf{x}_k$ with mean $\mathrm{E}\{\mathbf{x}_k\}$ and covariance $\mathbf{P}_k$ is approximated by $2n+1$ weighted points given by:

$$\begin{array}{rclrcl} \mathcal{X}_0 & = & \mathrm{E}\{\mathbf{x}\} & W_0 & = & \kappa/(n+\kappa) \\ \mathcal{X}_i & = & \mathrm{E}\{\mathbf{x}\} + \left(\sqrt{n+\kappa}\,\mathbf{P}^{1/2}\right)_i & W_i & = & 1/2(n+\kappa) \\ \mathcal{X}_{i+n} & = & \mathrm{E}\{\mathbf{x}\} - \left(\sqrt{n+\kappa}\,\mathbf{P}^{1/2}\right)_i & W_{i+n} & = & 1/2(n+\kappa) \end{array} \tag{4.3}$$

where $\kappa \in \mathbb{R}$ is a fitting parameter that varies according to the initial distribution. Its choice is empirical. For Gaussian probability distributions, $n+\kappa = 3$ is a good rule of thumb. $(\cdot)_i$ is the ith column of the matrix inside the brackets, and W_i the weight associated with the ith point.

The propagation of the set of points through the model M allows for better accuracy in the propagation of the mean and covariance of a distribution even through nonlinear models as compared to the EKF (cf. appendices in [28]). The statistics of the propagated points are calculated as follows.

The resulting random variable $\mathbf{x}_{k+1} = M(\mathbf{x}_k)$ is represented by the propagation of the ensemble members:

$$\mathcal{X}_i' = M(\mathcal{X}_i) \tag{4.4}$$

The mean of the resulting ensemble is approximated by:

$$\mathrm{E}\left\{\mathbf{x}_{k+1}\right\} \approx \sum_{i=0}^{2n} W_i \mathcal{X}_i' \tag{4.5}$$

And the covariance matrix by:

$$\mathbf{P}_{k+1} \approx \sum_{i=0}^{2n} W_i \left(\mathcal{X}_i' - \mathrm{E}\left\{\mathbf{x}_{k+1}\right\}\right)\left(\mathcal{X}_i' - \mathrm{E}\left\{\mathbf{x}_{k+1}\right\}\right)^{\mathrm{T}} \tag{4.6}$$

4.2.2 The unscented Kalman filter

The unscented Kalman filter relies on the unscented transformation to propagate the uncertainty instead of calculating the Jacobian of the operators. The calculation of the Kalman gain and the updates relies on the estimates of the mean and the covariance as explained in equations 4.5 and 4.6. The algorithm is described in detail in [29] and [48]. They show that nonlinearities are better taken into account, compared to the extended Kalman filter.

An improvement of the UKF is the square root UKF developed in [48]. The main advantage of the square root algorithm is to provide a better numerical stability and ensure that the state covariance matrices are positive, which was not necessarily the case in the standard UKF. The square root version of the filter performs as well as the standard filter.

4.3 Some additional comments

The two implementation of the filter for nonlinear system assume that n is small:

- There is a need to store n^2 elements for the covariance matrix. For problems where n reaches the value of $10^4 - 10^6$, it surpasses the storage capacity of most computers.

- There is a need for $2n + 1$ model evaluations. In case of real life models, the time necessary to evaluate the model is the limiting factor compared to the time necessary to propagate the filter.

For complex models and large dimensional systems, there is a need for algorithms that reduce the computational time and somehow compact the representation of the covariance matrix.

Some other extensions of the EKF are available, where the nonlinearities are better taken into account. They are normally based on a different sampling of the space to avoid derivation of the model operator. For example, [36] uses Stirling's interpolation formula to approximate the model operator. In general, those methods are as complex as UKF and EKF and present little interest in their original form for complex models and highly dimensional systems.

Chapter 5

Computational issues: sub-optimal schemes

In this chapter, methods are described that allow for high dimensional systems and reduce the number of model runs. These approaches are all sub-optimal, i.e. they do not lead to the optimal state estimate. But they have the advantage to work easily with large problems and complex models.

The methods can be divided into two categories:

- The covariance reduction methods, where the covariance matrix is approximated to reduce the number of model evaluation and the storage requirements.
- The model reduction methods, where the model operator is simplified to reduce the computational cost of the covariance matrix propagation.

5.1 Covariance reduction methods

5.1.1 Ensemble Kalman filter

Evensen in [14] provided for the first time a version of the Kalman filter that did not need any derivation of the model operator and that could virtually be used with any probability distribution.

The Ensemble Kalman filter (EnKF) is based on a Monte Carlo simulation of the probability distribution of the state. A set of realizations (the ensemble) is generated and propagated through the model operator. The statistics necessary for the Kalman filter are derived from the analysis of the ensemble.

The approach is very popular as its implementation is very simple, and the number of members of the ensemble necessary to get acceptable behavior of

the Kalman filter is small compared to the dimension of the system.

The main drawback of this method is the random sampling. The fact that the population is not drawn to represent some of its statistical features makes the convergence very slow (as fast as $n^{-0.5}$) [20].

Evensen in [15] reviews what has been done with the EnKF and gives a better mathematical framework for the filter as well as some implementation advices. In the case of non-linear model dynamics, the filter is not presented as a variance minimizer (in agreement with van Leeuwen in [49]). The stress is put on the simplicity of implementation and the ability of the filter to handle the non-linearities because it does not use tangent-linear models or adjoints.

The algorithm can be summarized as follows.

The ensemble of m members is propagated through the model:

$$\mathbf{x}^{\mathrm{f}}_{k,i} = M(\mathbf{x}^{\mathrm{a}}_{k-1,i}, \mathbf{u}_k, \eta_{k,i}) \tag{5.1}$$

where $i = 1, 2, ..., m$ is the subscript designing the ensemble member and the set of $\eta_{k,i}$ are realizations of the model noise. The forecast of the state vector is calculated as the mean of the ensemble.

The error covariance matrix of the forecast is estimated from the ensemble as:

$$\mathbf{P}^{\mathrm{f}}_k = \Delta^{\mathrm{f}}_k (\Delta^{\mathrm{f}}_k)^{\mathrm{T}} \tag{5.2}$$

Each column $\Delta^{\mathrm{f}}_{k,i}$ of the matrix Δ is defined as:

$$\Delta^{\mathrm{f}}_{k,i} = \frac{1}{\sqrt{m-1}} \left(\mathbf{x}^{\mathrm{f}}_{k,i} - \mathrm{E}\left\{ \mathbf{x}^{\mathrm{f}}_k \right\} \right) \tag{5.3}$$

The covariance matrix is used to calculate the Kalman gain. The analysis is done by generating an ensemble of possible observations (assuming additive noise):

$$\mathbf{y}^{\mathrm{o}}_{k,i} = \mathbf{y}^{\mathrm{o}}_k + \varepsilon_{k,i} \tag{5.4}$$

and apply the analysis equation to each member of the ensemble:

$$\mathbf{x}^{\mathrm{a}}_{k,i} = \mathbf{x}^{\mathrm{f}}_{k,i} + \mathbf{K}_k \left(\mathbf{y}^{\mathrm{o}}_{k,i} - \mathbf{H}_k \mathbf{x}^{\mathrm{f}}_{k,i} \right) \tag{5.5}$$

The propagation of the covariance matrix is implicit in the propagation of the ensemble and does not require the linearization of the model nor the calculation of tangent linear model.

5.1.2 Reduced rank square root Kalman filter

The reduced rank square root Kalman filter (RRSQRT KF) is also a population based algorithm. The choice of the population is here deterministic. It is based on the square root filter (described for example in [6], pp 103-105).

The square root filter

One of the numerical issues of the Kalman filter is the inversion of large covariance matrices. The square root filter is based on the Kalman filter but operates with a lower triangular square root of the covariance matrix. It has the advantage of simplifying the inversion of the matrices as there exist specific numerical methods for triangular matrix inversion.

Moreover, the covariance matrices can contain very large and very small entries, making their manipulation numerically unstable. Manipulating square root of those matrices reduces the difference between the larger and the smaller entry of the matrix and ensures the positive-definiteness of the resulting covariance matrix.

The reduced rank square root filter

Verlaan and Heemink introduced in [51] the RRSQRT KF that limited the computation needed to obtain an estimate of the square root of the state covariance matrix. The population to evaluate the covariance matrix is chosen to represent the main eigenvalues of the covariance matrix. It is then assumed that the model is linear enough so that the propagation of the main eigendirections is sufficient to represent the state probability distribution.

The algorithm is summarized as follows.

The state vector is propagated through the model:

$$\mathbf{x}_k^{\mathrm{f}} = M(\mathbf{x}_{k-1}^{\mathrm{a}}, \mathbf{u}_k, 0) \tag{5.6}$$

Instead of calculating the whole covariance matrix, only a reduced rank square root approximation $\mathbf{L}$ is calculated. The number of column of $\mathbf{L}$ is $p < n$. Each column $\mathbf{L}_{k,i}$ $(i = 1, ..., p)$ is propagated by:

$$\mathbf{L}_{k,i}^{\mathrm{f}} = M(\mathbf{x}_{k-1}^{\mathrm{a}} + \mathbf{L}_{k-1,i}^{\mathrm{a}}, \mathbf{u}_k, 0) - M(\mathbf{x}_{k-1}^{\mathrm{a}}, \mathbf{u}_k, 0) \tag{5.7}$$

The system noise is propagated by adding extra columns:

$$\mathbf{L}_{k,p+i}^{\mathrm{f}} = M(\mathbf{x}_{k-1}^{\mathrm{a}}, \mathbf{u}_k, \eta_{k,i}) - M(\mathbf{x}_{k-1}^{\mathrm{a}}, \mathbf{u}_k, 0) \tag{5.8}$$

where $\eta_{k,i}$ $(i = 1, ..., p')$ are the modes of the square root of the system noise covariance matrix $\mathbf{Q}$ that are kept for approximation. In case of additive noise, the operation is equivalent to the concatenation of the system covariance matrix to the first p columns of the matrix $\mathbf{L}$.

The number of columns of the matrix $\mathbf{L}_k^{\mathrm{f}}$ is reduced from $p + p'$ to p by eigenvalue decomposition of the matrix $(\mathbf{L}_k^{\mathrm{f}})^{\mathrm{T}}\mathbf{L}_k^{\mathrm{f}}$.

The Kalman gain and the updates are calculated using the approximation $\mathbf{P}_k^{\mathrm{f}} \approx \mathbf{L}_k^{\mathrm{f}}(\mathbf{L}_k^{\mathrm{f}})^{\mathrm{T}}$.

This filter is performing well on cases that are close to linear but has some drawbacks:

- The implementation works best in case of additive noise. If the model noise is not additive, extra model computations are needed to calculate the extra columns in the equation 5.8.
- The selection of the main eigenvalues as representing the covariance matrix requires that the model is linear or close to linear. For nonlinear systems, improvements of the filter are needed and explored in section 5.1.3.
- The reduction of the covariance matrix, based on a singular value decomposition requires that the state vector contains values of the same order of magnitude. In his thesis [50], Verlaan proposes to normalize the state vector in case its values are too far apart. By experience, the normalization is not easy to implement and remains one of the major drawbacks of the method. A new implementation of the filter, called TRUE RRSQRT avoids the normalization part by using a different algorithm to reduce the matrix $\mathbf{L}$ [46].
- The implementation of the filter is less straightforward than the EnKF. The points mentioned above can be difficult to tackle.

If the filter can apply, then similar performances can be attained using less model evaluations than the EnKF. [32].

5.1.3 Hybrids of ensemble and reduced rank Kalman filter

The creators of the RRSQRT KF propose in [20] two filters that are hybrids of the EnKF and the RRSQRT KF. They aim essentially at reducing the computing power needed while better taking into account the nonlinearities of the system. In both techniques, the main variations of the state are captured by a reduced square root filter. An ensemble filter is used to capture the additional variations left out by the square root filter.

The Partial Orthogonal Ensemble Kalman Filter, POEnKF

In this filter, q_1 elements are propagated using the RRSQRT KF algorithm, while q_2 elements are propagated using the EnKF. During the measurement update, the q_2 elements of the ensemble filter are taken into account only for their contribution in the space orthogonal to the q_1 elements of the square filter.

The Complementary Orthogonal subspace Filter For Efficient Ensembles, COFFEE

The COFFEE filter uses the same principle as the POEnKF. The only difference is in the method of sampling the q_2 elements of the ensemble filter. Instead of sampling over the whole state space, only elements that are orthogonal to the q_1 elements are chosen.

The results show that those filters are efficient for small q_2. They are much faster than the EnKF as most of the covariance evaluation is done following a RRSQRT KF scheme.

5.1.4 Singular Evolutive Kalman filters

Different implementations of the singular evolutive extended Kalman filter (SEEK) have been provided in a series of papers [37, 38, 22, 23]. The original filter consists in the approximation of the covariance matrix using an Empirical Orthogonal Function (EOF) analysis of its evolution in time. The filter then makes the correction only in the directions that are significant.

This filter has the drawback on relying on the linearization of the model operator. The Singular Evolutive Interpolated Kalman filter (SEIK) do not rely on such a linearization. It uses instead an ensemble of realizations that it propagates in the same way as the EnKF. The difference lies in the way the ensemble is generated. Instead of being drawn completely at random, the members of the ensemble are drawn in the main orthogonal directions that have been defined by the EOF analysis of the model. By definition, the resulting ensemble has the mean and covariance matrix estimated by the analysis step of the Kalman filter.

5.2 Model reduction methods

5.2.1 Simplification of the model dynamics

The simplification of the model dynamics allow for a larger amount of model runs, without additional computational costs. A simple model is generally

used to compute the propagation of the covariance matrix, while the full model is still used for the propagation of the state.

Whereas the covariance reduction methods are general, the model reduction methods are related to the type of problem solved. Todling and Cohn in [45] propose to classify the suboptimal filters in six categories. In reality most methods are hybrids.

1. Covariance modelling. (See section 5.1)
2. Dynamics simplification. Dee [9] used a simplification of a wind model as a forcing for a one-dimensional shallow water model. He showed that the analysis remains nearly optimal. In [21], Hoeben and Troch use a linearization of their unsaturated zone model to assimilate microwave data.
3. Reduced resolution. Sørensen et al. in [43] used a two-dimensional approximation of their three-dimensional hydrodynamic model to estimate the state covariance matrix. EOF analysis of the model operator or of the covariance matrix (like in [7]) can be included in these methods as they are only taking in account the main modes of the dynamics.
4. Local approximation. Only points that are close enough to each other are considered in the estimation of the covariance structure (c.f. [45]). This method can be considered either as model reduction or a covariance reduction.
5. Limiting filtering. (c.f. the steady Kalman filter in section 5.2.2)
6. Monte Carlo methods (c.f. EnKF in section 5.1.1)

The study by Fukumori and Malanotte-Rizzoli [18] is an hybrid of different methods. They use an approximation of the covariance matrix, combined with a model linearization for the error propagation and a reduction of the model dimension.

5.2.2 The steady Kalman filter

The steady Kalman filter is an extreme case of model reduction. The Kalman gain is assumed to be constant, which implies that the model does not affect the covariance matrix of the state. Usually, the Kalman gain is estimated during a warm-up period and then is kept constant for the rest of the simulation. This technique has the advantage of not requiring any additional model run when the Kalman gain is estimated. In [43], the Kalman

gain is estimated using an EnKF run before being used in other experiments. Cañizares et al. [5] showed that in an operational 2D model of the Danish waters a few days of assimilation were sufficient to reach a quasi-steady state.

5.3 Further than the Kalman filter: a variance minimizing filter

The variance minimizing filter has been proposed by van Leeuwen in [49] as an alternative to the EnKF in situations where non-linearities are important. In this case, the linear analysis step of the Kalman filter does not allow for taking into account probability distributions that are not defined by their first two moments. The EnKF is not a variance minimizing filter anymore.

The filter is derived from Bayes's theorem. The main interest of the filter is that it does not require any inversion or linearization of the model operator or the observation operator. The representation of the probability distribution is based on an ensemble where the members are weighted according to their probability.

The original implementation needs a large amount of ensemble members to be able to represent the realization with low probability. A technique similar to importance resampling has been used to tackle the problem. It consists in checking the weight of each member of the ensemble and duplicate it according to its weight. The number of copies is calculated by the weight multiplied by the size of the ensemble. The fractional remainders defines a new probability distribution where the last ensemble members are randomly drawn from. This approach prevents the ensemble to diverge by removing ensemble members that are unlikely to occur.

Chapter 6

Application of the Kalman filter to hydrological modelling

6.1 The Kalman filter in practice

6.1.1 Discussion of the usual assumptions

To apply data assimilation, and more specifically the Kalman filter, to hydrological modelling, assumptions have to be made. Even though it is most of the time necessary to make those assumptions to be able to apply the algorithms, one has to be aware of the limitations they impose on the systems and the way the data are handled.

Ergodicity

The ergodic hypothesis was first stated in 1850 by Maxwell, as explained in [3]. It states that for a system with a very large number of degrees of freedom, the statistical mean is equal to the temporal mean.

This assumption is needed in hydrology when real-life systems are studied, because the only data available are historical time series. The statistics have to be drawn from temporal sampling instead of ensemble sampling. The problems of bias in relation with the Kalman filter are one of the examples where the assumption is used: the difference between the bias and the random error is quantified by the time variability. For example, Dee and Todling in [11] use the hydrological time series to identify the parameter γ defined in section 3.5.6.

Stationarity

There are different levels of stationarity. Here are described those that are the most commonly considered in hydrology

First order stationarity The function $X(t)$ is stationary to the first order if it satisfies the condition [3]:

$$\mathrm{E}\{X(t)\} = m \tag{6.1}$$

for all t, with m independent of t.

Weak second order stationarity The function $X(t)$ is stationary to the second order in the weak sense if it is first order stationary and satisfies the condition [3]:

$$cov\{X(t), X(t-\tau)\} = cov\{X(t+h), X(t-\tau+h)\} \tag{6.2}$$

for any h.

A weaker stationarity definition includes only the stationarity of the covariance.

In both definitions, one can replace the temporal variable by a spatial variable. For the first order stationarity, it means that the expected value of a field is constant over space. In the case of the weak second order stationarity, it means that the covariance between two points in space only depends on the distance between the two points. The construction of semi-variograms relies on this assumption.

Stationarity and Kalman filtering Because most of the error covariance models are stationary, it is sometimes assumed that the Kalman filter needs some stationarity. In fact, the filter itself does not require any stationarity assumption.

Gaussian noise Most of the applications quoted in the following sections still assume that the noise structures are Gaussian and additive. These are the original assumptions of the linear Kalman filter. Those assumptions, as questionable as they are, still provide decent results. In any case, the analysis equation is linear and therefore corrects only the first two moments of the probability distribution of the state vector.

6.1.2 Data assimilation in hydrology

As mentioned in [47], data assimilation in hydrology uses algorithms that have been developed in other domains, like oceanography or meteorology. With the availability of new data sources, both from remote sensing and cheaper sensors on the ground, it became possible to validate and correct hydrological models. There is still a need for improvement and adaptation of the methods to the specificities of hydrological modelling, such as the problems of heterogeneity and the difficulties to observe subsurface processes.

McLaughlin in [33] goes in the same way when he points out that most of the data assimilation techniques used in hydrology are "natural extensions of classical linear Gaussian estimation methods". Most of hydrological problems are non-linear and would demand more advanced techniques.

Nevertheless the following sections show that the techniques derived from the Kalman filter have potentials to improve the performance of hydrological models.

6.2 Groundwater modelling

Most of the applications of Kalman filtering to groundwater modelling before 1995 has been studied in [12]. This part will focus on the recent developments.

During the past decade, the main improvement in data collection comes from the availability of remote sensing data, mostly microwave observations. Therefore, the research has focused on the unsaturated zone and the possibility to propagate in depth information given by the moisture observations of the upper part of the soil.

6.2.1 State estimation

The state estimation consists in the update of the moisture contents or water heads at observation points and reconstruction of the state at unobserved points.

In the unsaturated zone, the work of Walker et al. [52] shows the ability of the Kalman filter to correct field simulations of the soil moisture content. As they work with a three-dimensional soil-moisture model that is non-linear and quite expensive in terms of computation, they also introduce a simplified propagation of the covariance matrix, using an autoregressive approximation of the model operator.

Reichle et al. [41] use the Ensemble Kalman filter (EnKF) to assimilate microwave radiobrightness observations into a land surface model. The filter

is compared with the reference performance of an optimal smoother. They studied the possibility of the EnKF to work with non-Gaussian distributions. They show that a minimum ensemble size of 500 is necessary to get an error reduction similar to the smoother and a correct evaluation of the error variance. The study also shows the existence of a residual error compared to the optimal estimate of the smoother even when the ensemble size is increased up to 10000. The study in [8] confirms the performance of the EnKF without making any comparison with an optimal scheme. A previous study by Hoeben and Troch [21] used a linearization of the unsaturated zone model instead of a Monte Carlo approach to assimilate radiobrightness observations.

In some applications the Kalman filter, used to get an optimal estimate of the water heads or soil moisture, is combined with a maximum likelihood estimation of the parameters. In the saturated zone, Bierkens et al. in [2] use a geostatistical model instead of a physical model of the water heads combined with an ensemble Kalman filter. In the unsaturated zone, Cahill et al. [4] use the extended Kalman filter combined with a physical model to estimate the hydraulic conductivities in an experimental field.

6.2.2 Parameter estimation

Parameter estimation is often called inverse modelling in the groundwater modelling world. The main parameters estimated are the hydraulic conductivities in the saturated zone.

Even though McLaughlin and Townley in [34] consider that the extended Kalman filter should not be used as parameter estimator because it is suboptimal and the convergence is not well understood, there have been different successful attempts in using the filter. One of the possible causes of success is the lack of extreme non-linearities. In the case when the uncertainties are assumed to be Gaussian, the Kalman filter becomes a maximum a-posteriori estimate (cf. section 2.1.3).

In [16], a linear Darcy flow model is used to estimate conductivities on a real case-study. The authors have derived the linear system to be solved using the Kalman filter formulation.

One step further, Eppstein and Dougherty in [13] simplify the extended Kalman filter by simplifying the covariance update. They also use a technique to create zones in the hydraulic conductivity field using a clustering algorithm. This approach turns the normally ill-posed problem of parameter estimation into a well-posed problem by reducing the number of unknown variables. Unfortunately, they do not apply their technique to a real case.

6.3 Rainfall-runoff and surface water modelling

A series of articles have been written on the NAM model and the DHI model MIKE 11

Refsgaard et al. in [40] have reformulated the NAM (A lumped rainfall-runoff model) into a state space form to allow the use of the Kalman filter. The main interest of this paper is the resulting uncertainty analysis showing that the uncertainty on the output is mainly driven by the uncertainty on the rainfall.

Reefsgaard in [39] compared the combination of the NAM model and the Kalman filter as in [40] to the NAM/MIKE 11 model coupled with an error prediction technique and concludes that the Kalman filter performs better than an error prediction model only when the basic hydrological model is well-calibrated.

Finally, Hartnak and Madsen provide in [19] a thorough analysis of the use of the ensemble Kalman filter in river modelling with MIKE 11. The Kalman filter was able to correct different perturbations applied to the model, including white noise on the boundary conditions, trend in the boundary conditions, phase shift. Moreover the filter could reconstruct lost information and be used in a flood forecasting implementation.

Others have applied the Kalman filter to rainfal-runoff models:

In [54], Wood and O'Connell derive the Kalman filter and the extended Kalman filter for real-time forecasting and apply it for simultaneous parameter and state estimation in the Sacramento Soil Moisture model and the National Weather Services River Forecasting model.

In [31], Lee and Singh apply successfully the Kalman filter for parameter estimation in a tank model. The filter has been used to correct the parameters of the model through time and update the uncertainty on the runoff.

Chapter 7

Conclusions

This review provides an overview of the uses of the Kalman filter in a domain that is far from its original use: from a linear filter used in signal processing, it is now adapted and used in high-dimensional models with non-linearities.

The methods have been developed to a level where they are operational, and proved that they can be used in hydrology. Now the focus should shift to issues that have been neglected up till now. The main issue is the way the uncertainties are actually represented. In most application, the noise is mostly described as Gaussian and additive. This representation is of course far from reality but well suited for the linear implementation of the filter.

With the non-linear implementation of the filter, and the development of more accurate filters (like the variance minimizing filter of van Leeuwen [49]), a more complex and more detailed representation of the uncertainties becomes meaningful. Whether the improvement of the noise representation will improve the filter performances remains to be seen, but there is definitely knowledge to gain by questioning ourselves on the real origins of uncertainty.

Bibliography

[1] M. G. Anderson and T. P. Burt, editors. *Hydrological Forecasting*. John Wiley and Sons Ltd, 1985.

[2] M. F. P. Bierkens, M. Knotters, and T Hoogland. Space-time modeling of water table depth using a regionalized time series model and the Kalman filter. *Water Resources Research*, 37(5):1277–1290, 2001.

[3] A. Blanc-Lapierre and R. Fortet. *Theory of Random Functions*. Gordon and Breach Science Publishers, 1968.

[4] A. T. Cahill, F. Ungaro, M. B. Parlange, M. Mata, and D. R. Nielsen. Combined spatial and Kalman filter estimation of optimal soil hydraulic properties. *Water Resources Research*, 35(4):1079–1088, 1999.

[5] R. Cañizares, H. Madsen, H. R. Jensen, and H. J. Vested. Developments in operational shelf sea modelling in danish waters. *Estuarine and Coastal Shelf Science*, 53:595–605, 2001.

[6] C. K. Chui and G. Chen. *Kalman Filtering with Real Time Applications*. Springer-Verlag, third edition, 1999.

[7] S. E. Cohn and R. Todling. Approximate data assimilation schemes for stable and unstable dynamics. *Journal of the Meteorological Society of Japan*, 74(1), 1996.

[8] W. T. Crow and E. F. Wood. The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, 26:137–149, 2003.

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

[10] D. P. Dee and A. M. Da Silva. Data assimilation in the presence of forecast bias. *Quart. J. R. Meteorol. Soc.*, 124:269–295, 1998.

[11] D. P. Dee and R. Todling. Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Monthly Weather Review*, 128(9):3268–3282, 2000.

[12] U. Eigbe, M. B. Beck, and H. S. Wheater Adn F. Hirano. Kalman filtering in groundwater flow modelling: Problems and prospects. *Stochastic Hydrology and Hydraulics*, 12:15–32, 1998.

[13] M. J. Eppstein and D. Dougherty. Simultaneous estimation of transmissivity values and zonation. *Water Resources Research*, 32(11):3321–3336, 1996.

[14] G. Evensen. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162, 1994.

[15] G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *submitted to Ocean Dynamics*, 2002.

[16] M. Ferraresi, E. Todini, and R. Vognoli. A solution to the inverse problem in groundwater hydrology based on Kalman filtering. *Journal of Hydrology*, 175:567–581, 1996.

[17] B. Friedland. Treatment of bias in recursive filtering. *IEEE Trans. Aut. Contr.*, AC-14(4):359–367, 1969.

[18] I. Fukumori and P. Malanotte-Rizzoli. An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *Journal of Geophysical Research*, 100(C4):6777–6793, 1995.

[19] J. Hartnak and H. Madsen. Data assimilation in river flow modelling. In *Proceedings of the 4^{t}h DHI Software Conference*, Helsingør, Denmark, 2001.

[20] A. W. Heemink, M. Verlaan, and A. J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129:1718–1728, 2001.

[21] R. Hoeben and P. A. Troch. Assimilation of active microwave observation for soil moisture profile estimation. *Water Resources Research*, 36(10):2805–2819, 2000.

[22] I. Hoteit, D. T. Pham, and J. Blum. A semi-evolutive partially local filter for data assimilation. *Marine Pollution Bulletin*, 43(7-12):164–174, 2001.

[23] I. Hoteit, D. T. Pham, and J. Blum. A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine systems*, 36:101–127, 2002.

[24] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc. Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75(1B):181–189, 1997.

[25] M. B. Ignagni. An alternate derivation and extension of Friedland's two-stage Kalman estimator. *IEEE Trans. Aut. Contr.*, AC-26(3):746–750, 1981.

[26] M. B. Ignagni. Separate-bias Kalman estimator with bias state noise. *IEEE Transactions on Automatic Control*, 35(3):338–341, 1990.

[27] M. B. Ignagni. Optimal and suboptimal separate-bias Kalman estimators for a stochastic bias. *IEEE Transactions on Automatic Control*, 45(3):547–551, 2000.

[28] S. J. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. *Internet publication*, 1996.

[29] S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Aerosense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, Florida, 1997.

[30] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, pages 35–45, March 1960.

[31] Y. H. Lee and V. P. Singh. Tank model using Kalman filter. *Journal of hydrologic engineering*, pages 344–349, October 1999.

[32] H. Madsen and R. Cañizares. Comparison of extended and ensemble kalman fitlers for data assimilation in coastal area modelling. *International Journal For Numerical Methods in Fluids*, 31:961–981, 1999.

[33] D. McLaughlin. An integrated approach to hydrologic data assimilation: Interpolation, smoothing and filtering. *Advances in Water Resources*, 25:1275–1286, 2002.

[34] D. McLaughlin and L. R. Townley. A reassessment of the groundwater inverse problem. *Water Resources Research*, 32(5):1131–1161, 1996.

[35] J. M. Mendel. Extension of Friedland's bias filtering technique to a class of nonlinear systems. *IEEE Trans. Aut. Contr.*, pages 296–298, April 1976.

[36] M. Nørgaard, N. K. Poulsen, and O. Ravn. New developments in state estimation for nonlinear systems. *Automatica*, 36:1627–1638, 2000.

[37] D. T. Pham, J. Verron, and L. Gourdeau. Singular evolutive Kalman filters for data assimilation in oceanography. *C. R. Acad. Sci. Paris*, 326:255–260, 1998.

[38] D. T. Pham, J. Verron, and M. C. Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16:323–340, 1998.

[39] J. C. Refsgaard. Validation and intercomparison of different updating procedures for real time forecasting. *Nordic Hydrology*, 28:65–84, 1997.

[40] J. C. Refsgaard, D. Rosbjerg, and L. M. Markussen. Application of the Kalman filter to real time operation and to uncertainty analyses in hydrological modelling. *Scientific Procedures Applied to the Planning, Desing and Management of Water Resources Systems. IAHS Publ. no. 147*, pages 273–282, 1983.

[41] R. H. Reichle, D. B. McLaughlin, and D. Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130:103–114, 2002.

[42] J. Sénégas, H. Wackernagel, W. Rosenthal, and T. Wolf. Error covariance modeling in sequential data assimilation. *Stocahstic environmental research and risk assessment*, 15:65 – 86, 2001.

[43] J. V. T. Sørensen, H. Madsen, and H. Madsen. Towards an operational data assimilation system for a three-dimensional hydrodynamic model. In I. D. Cluckie, D. Han, J. P. Davis, and S. Heslop, editors, *Hydroinformatics 2002, Proceedings of the Fifth International Conference on Hydroinformatics*, pages 1204–1209, Cardiff, UK, 2002.

[44] R. Todling. Estimation theory and foundations of atmospheric data assimilation. Technical Report DAO Office Note 1999-01, Data Assimilation Office, Goddard Space Flight Center, 1999.

[45] R. Todling and S. E. Cohn. Suboptimal schemes for atmospherical data assimilation based on the Kalman filter. *Monthly Weather Review*, 122:2530–2557, 1994.

[46] D. Treebushny and H. Madsen. A new reduced rank square root Kalman filter for data assimilation in mathematical models. In *Proceedings of the 3rd International Conference on Computational Science*, Melbourne (Australia) and St. Petersburg (Russian Federation), 2003.

[47] P. A. Troch, C. Paniconi, and D. McLaughlin. Catchment-scale hydrological modeling and data assimilation. *Advances in Water Resources*, 26:131–135, 2003.

[48] R. van der Merwe and E. A. Wan. The square-root unscented Kalman filter for state and parameter-estimation. In *Proceedings of the International Conference on Acoustics, Speech & Signal processing*, Salt Lake City, Utah, 2001.

[49] P. J. van Leeuwen. A variance-minimizing filter for large-scale applications. *submitted to Monthly Weather Review*, 2002.

[50] M. Verlaan. *Efficient Kalman Filtering Algorithms for Hydrodynamic Models*. PhD thesis, Delft Technical University, 1998.

[51] M. Verlaan and A. W. Heemink. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11:346–368, 1997.

[52] J. P. Walker, G. R. Willgoose, and J. D. Kalma. Three-dimensional soil moisture profile retrival by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application. *Water Resources Research*, 38(12):1301–1313, 2002.

[53] E. A. Wan and R. van der Merwe. The unscented Kalman filter for nonlinear estimation. Technical report, Oregon Graduate Instute of Science & Technology, 2000.

[54] E. F. Wood and P. E. O'Connell. *Real-Time Forecasting*, chapter 15, pages 505–558. In Anderson and Burt [1], 1985.

[55] D. H. Zhou, Y. X. Sun, Y. G. Xi, and Z. J. Zhang. Extension of Friedland's separate-bias estimation to randomly time-varying bias of nonlinear systems. *IEEE Trans. Auto. Contr.*, 38(8):1270–1273, 1993.