Stochastic Hydrology and Hydraulics 11: 349-368 © Springer-Verlag 1997

Tidal flow forecasting using reduced rank square root filters

M. Verlaan and **A.W. Heemink**
Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, Delft, The Netherlands

Abstract: The Kalman filter algorithm can be used for many data assimilation problems. For large systems, that arise from discretizing partial differential equations, the standard algorithm has huge computational and storage requirements. This makes direct use infeasible for many applications. In addition numerical difficulties may arise if due to finite precision computations or approximations of the error covariance the requirement that the error covariance should be positive semi-definite is violated.

In this paper an approximation to the Kalman filter algorithm is suggested that solves these problems for many applications. The algorithm is based on a reduced rank approximation of the error covariance using a square root factorization. The use of the factorization ensures that the error covariance matrix remains positive semi-definite at all times, while the smaller rank reduces the number of computations and storage requirements. The number of computations and storage required depend on the problem at hand, but will typically be orders of magnitude smaller than for the full Kalman filter without significant loss of accuracy.

The algorithm is applied to a model based on a linearized version of the two-dimensional shallow water equations for the prediction of tides and storm surges.

For non-linear models the reduced rank square root algorithm can be extended in a similar way as the extended Kalman filter approach. Moreover, by introducing a finite difference approximation to the Reduced Rank Square Root algorithm it is possible to prevent the use of a tangent linear model for the propagation of the error covariance, which poses a large implementational effort in case an extended kalman filter is used.

Key words: Data assimilation, Kalman filter, Square root filter.

I Introduction

In the Netherlands large areas of land lie below or just above mean sea level. To protect these densely populated areas from the sea many dikes and barriers were constructed. For the large barriers in the Eastern Scheldt and 'Nieuwe Waterweg' accurate predictions for waterlevels are needed 6 hours in advance to decide whether these barriers have to be closed or not. Also for the protection of the dikes and for the ships entering the harbor at Rotterdam these predictions are needed.

At the moment the predictions for waterlevels during storm surges are computed using a two-dimensional shallow water flow model of the North Sea and a steady state Kalman filter algorithm to assimilate waterlevel measurements into the model to improve the model forecasts [Heemink, 1986, Heemink and Kloosterhuis, 1990]. To obtain a filter algorithm suitable for implementation either a distributed parameter filter for the shallow water equations is derived and after that discretized [Curi et al., 1995], or first the partial differential equations are discretized and a 'lumped parameter' Kalman filter is employed. The latter approach will be followed in the sequel.

The number of additional computations needed for data assimilation with the steady state Kalman filter is very small. Provided that the model is time invariant and approximately linear this procedure works quite well and has been used on a routine basis for some years. To speed up of the (off-line) steady state gain computations a Chandrasekhar type algorithm [Morf et al., 1974, Heemink, 1986, Bolding, 1995] or a doubling algorithm [Anderson and Moore, 1979] can be applied. When there are no irregular boundaries a coarser grid combined with an interpolation scheme can be used for the gain computations [Fukumori and Melanotte-Rizzoli, 1995].

For many data assimilation problems a steady state approach is not possible and a full Kalman filter has to be used. For storm surge prediction the errors in the wind forcing are non-stationary and the wind friction coefficient depends on the mean wave-height which varies during a storm. As a result a time varying Kalman filter would improve the model forecasts considerably. For problems with one spatial dimension a full extended Kalman filter can sometimes be used [Heemink, 1986, Budgell, 1986]. However for two or more spatial dimensions the computational burden is usually too large. And even if the computation were possible numerical difficulties can be expected because of the high condition number of the error covariance matrix [Boggs et al., 1995]. As a result approximations of the Kalman filter equations are needed. Following Todling and Cohn we will refer to these approximations as sub-optimal schemes or SOS's [Todling and Cohn, 1994].

Most of these methods are aimed at an approximation of the model dynamics or of the error covariance matrix, because the main part of the computations is needed for the propagation of the error covariance. The model is often simplified by removing less important terms from the equations, or by introducing other simplifying assumptions. The simplified model is then used for time propagation of the error covariance and the full model for the time propagation of the estimate. Cohn recently proposed to approximate the state transition matrix by one of a lower rank [Cohn and Todling, 1995]. The partial singular value decomposition can be used in this case to reduce the computations.

Various methods have been proposed for the approximation of the error covariance matrix. Setting correlations for large distances to zero can be exploited to speed the algorithm up considerably [Parrish and Cohn, 1985]. However, due to the generally large condition number of the error covariance matrix negative eigenvalues may appear. A solution to this problem is to use a square root filter [Boggs et al., 1995], but in this approach it is more difficult to exploit the sparse structure of the matrices. Often the error covariance matrix has only a few large eigenvalues, which can be used for approximation. The resulting partial eigenvalue decomposition can be used for fast propagation of the error covariance. Todling and Cohn used this idea together with a Lanczos algorithm for the eigenvalue computations to obtain an efficient and

general algorithm, the Partial Eigen decomposition Kalman Filter or PEKF [Cohn and Todling, 1995].

In this paper the Reduced Rank Square Root (RRSQRT) algorithm is presented. Preliminary results were presented at the Second International Symposium on Assimilation of Observations in Meteorology and Oceanography in Tokyo [Verlaan and Heemink, 1995]. The algorithm uses a square root approach together with an approximation of the error covariance matrix by one of a lower rank. The optimal choice for this low rank approximation results in the use of the eigenvalues and eigenvectors of the error covariance matrix. The .algorithm is applied to storm surge forecasting in the North Sea. For a test model the results of the suboptimal filter are compared with the exact Kalman filter results.

Finally, for use with non-linear models, a method is proposed to propagate the error covariance matrix, that does not use a tangent linear model but only the full state transition function. The method is based on finite differences, and simplifies the use of the RRSQRT algorithm considerably.

2 A deterministic model for storm surge prediction

In order to obtain estimates that are consistent with physical laws like conservation of mass and momentum the stochastic model is based on a deterministic model that reflects these laws. For storm surge prediction the shallow water equations can be used for the deterministic model [Stelling, 1984].

$$\frac{\partial \mathrm{h}}{\partial \mathrm{t}} + \frac{\partial \mathrm{H}\nu_\xi}{\partial \mathrm{t}} + \frac{\partial \mathrm{H}\nu_\eta}{\partial \eta} = 0 \tag{1}$$

$$\frac{\partial \mathrm{H}\nu_\xi}{\partial \mathrm{t}} + \nu_\xi \frac{\partial \nu_\xi}{\partial \xi} + \nu_\eta \frac{\partial \nu_\xi}{\partial \eta} + \mathrm{g} \frac{\partial \mathrm{h}}{\partial \xi} - \mathrm{f}\nu_\eta + \frac{\mathrm{g}\nu_\xi\sqrt{\nu_\xi^2+\nu_\eta^2}}{\mathrm{C}^2\mathrm{H}} - \mathrm{C_d}\frac{\rho_\mathrm{a}}{\rho_\mathrm{w}}\frac{\mathrm{V}^2\cos\psi}{\mathrm{H}} + \frac{1}{\rho_\mathrm{w}}\frac{\partial \mathrm{p_a}}{\partial \xi} = 0 \tag{2}$$

$$\frac{\partial \mathrm{H}\nu_\eta}{\partial \mathrm{t}} + \nu_\xi \frac{\partial \nu_\eta}{\partial \xi} + \nu_\eta \frac{\partial \nu_\eta}{\partial \eta} + \mathrm{g} \frac{\partial \mathrm{h}}{\partial \eta} + \mathrm{f}\nu_\xi + \frac{\mathrm{g}\nu_\eta\sqrt{\nu_\xi^2+\nu_\eta^2}}{\mathrm{C}^2\mathrm{H}} - \mathrm{C_d}\frac{\rho_\mathrm{a}}{\rho_\mathrm{w}}\frac{\mathrm{V}^2\sin\psi}{\mathrm{H}} + \frac{1}{\rho_\mathrm{w}}\frac{\partial \mathrm{p_a}}{\partial \eta} = 0 \tag{3}$$

where:

ξ, η	=	coordinates in the horizontal plane
ν_ξ	=	depth-averaged current in ξ direction
ν_η	=	depth-averaged current in η direction
h	=	water level above the reference plane
D	=	water depth below the reference plane
H=h+D	=	total water depth
g	=	gravity acceleration

f	=	coefficient for the Corriolis force
C	=	Chezy coefficient
V	=	wind velocity
ψ	=	wind angle with respect to the positive ξ-axis
C_d	=	wind friction coefficient
p_a	=	air pressure at the surface
ρ_w	=	density of sea water
ρ_a	=	density air at the surface

A model usually has two or three types of boundaries. At land-water boundaries the normal flow is set to zero. At the open boundaries no physical boundaries exist and thus artificial ones will have to be specified. Often the surface level or the flow is prescribed at these boundaries, but also non-reflecting boundaries are used [Stelling, 1984]. The waterlevels are often specified by their harmonic constituents.

When meteorological activity is low and there are no external surges, this deterministic model is quite accurate. In these cases the Root Mean Square (RMS) error in the waterlevels is approximately 15 cm. In case of storm surges the accuracy is less and it is believed that the main sources of error are the open boundary condition and the wind input as provided by the meteorological model.

Although the detailed non-linear model described can be used for the RRSQRT algorithm and this would most likely result in more accurate predictions, a simplified linear time-independent model will be used in the sequel since this will make a more detailed analysis of the results possible. Since the model used is linear it is possible to separate the astronomical tide and the so-called set-up. For storm surge prediction we are mainly interested in the set-up since the astronomical tide is known much more accurately. The simplified model for the set-up is given by:

$$\frac{\partial h}{\partial t} + D\,\frac{\partial \nu_\xi}{\partial \xi} + D\,\frac{\partial \nu_\eta}{\partial \eta} = 0 \tag{4}$$

$$\frac{\partial \nu_\xi}{\partial t} + g\,\frac{\partial h}{\partial \xi} - f\nu_\eta + \frac{\lambda \nu_\xi}{D} - \frac{\tau_\xi}{D} = 0 \tag{5}$$

$$\frac{\partial \nu_\eta}{\partial t} + g\,\frac{\partial h}{\partial \eta} + f\nu_\xi + \frac{\lambda \nu_\eta}{D} - \frac{\tau_\eta}{D} = 0 \tag{6}$$

where λ is the coefficient for the linearized friction and τ_ξ, τ_η are stresses due to wind. These equations are discretized using an Alternating Directions Implicit (ADI) method and a staggered grid that is based on a simplification of the method by Leenderstse and Stelling [Stelling, 1984] for the equations 4-6 (see [Brummelhuis, 1992]).

3 Stochastic extension of a deterministic model and measurements

Before a Kalman filter can be applied a description of the errors in the model and the measurements are needed, since the covariances of the errors determine how the model predictions and the measurements will be weighted.

An important tool for the description of the errors in the model and measurements are ARMA processes [Box et al., 1994]. An ARMA model maps a white noise process to an error process with the desired shape of the autocorrelation function.

For storm surge forecasting in the North Sea it is assumed that errors in the forecast are mainly caused by the uncertainty at the open boundary and in the meteorological forcing, i.e. wind stress and pressure gradients. The covariances of these errors will be modelled using ARMA models.

To obtain a general notation for a stochastic system the waterlevels h and the current speeds u, v at all the grid points (i,j) are put together in a large vector x(k) together with the state variables of the ARMA models. Using this state vector the discretized shallow water equations can formally be written as:

$$x(k+1) = f(k, x(k), u(k), w(k)) \tag{7}$$

In a similar way, with the measurements at time k stacked in y(k), the measurement relation can be denoted as:

$$y(k) = g(k, x(k), v(k)) \tag{8}$$

where $x(k) \in R^n$, $w(k) \in R^m$ and $v(k), y(k) \in R^p$, $u(k) \in R^l$. The u(k) contain the astronomical part of the waterlevels on the boundary and the prediction of wind-stresses by a meteorological model. The system noise w and the measurement noise v are white Gaussian and zero mean. The covariances are given by $E[w(k)w(k)'] = \sum_s k)$, $E[v(k)v(k)'] = \sum_0(k)$ and $E[w(k)v(l)'] = 0$ for all k, l. The initial condition is given by

$$E[x(0)] = x_0 \tag{9}$$

$$E[(x(0) - x_0)^2] = P_0 \tag{10}$$

For the linearized shallow water equations 7 and 8 can be written as

$$x(k+1) = A(k)x(k) + B(k)u(k) + F(k)w(k) \tag{11}$$

$$y(k) = C(k)x(k) + v(k) \tag{12}$$

This formal notation will be used in the sequel to describe the various algorithms.

4 The Kalman filter

When a stochastic description of model and measurements is available it is possible to combine these sources of information to obtain an optimal estimate of the state. When the model is linear with Gaussian noise Kalman and Bucy showed that a recursive update of the estimate can be found [Kalman, 1960, Kalman and Bucy, 1961]. Under these assumptions the estimate is optimal for several criteria, such as minimum variance and maximum likelihood. For the estimate of x(k) given $Y(k) = \{y(l), l=0, \cdots, k\}$ denoted by $\hat{x}(k|k)$ the Kalman filter equations are given by

$$\hat{x}(k+1|k) = A(k)\hat{x}(k|k) + B(k)u(k) \tag{13}$$

$$P(k+1|k) = A(k)P(k|k)A(k)' + F(k)\sum_s(k)F(k)' \tag{14}$$

$$\begin{aligned}\hat{x}(k+1|k+1) = {} & \hat{x}(k+1|k) \\ & + K(k+1)(y(k+1) - C(k+1)\hat{x}(k+1|k))\end{aligned} \tag{15}$$

$$K(k+1) = P(k+1|k)C(k+1)' \\ (C(k+1)P(k+1|k)C(k+1)' + \sum_0(k+1))^{-1} \quad (16)$$

$$P(k+1|k+1) = P(k+1|k) - K(k+1)C(k+1)P(k+1|k) \quad (17)$$

$$\hat{x}(0|-1) = x_0 \quad (18)$$

$$P(0|-1) = P_0 \quad (19)$$

Although these equations, at least in principle, provide a solution to many data assimilation problems, a straight forward application is not possible for the problem of reconstructing storm surges because the number of computations becomes infeasible for large systems. Also storage requirements grow fast with increasing model dimension.

It can be seen from the structure of equations 13-19, that K(k) does not depend on the measurements y and therefore can be computed in advance and stored. During the actual filtering the stored values can then be used. If the model is time invariant and stable, it can be shown that the Kalman gain K(k) converges to a limit value K. When this steady state Kalman gain is used for all measurement times the estimate converges to the optimal estimate for large k.

To compute the steady state Kalman gain also a Chandrasekhar type algorithm [Morf et al., 1974] or a doubling algorithm [Anderson and Moore, 1979] can be used instead of the equations 13-19. Both algorithms can reduce computation times considerably. The Chandrasekhar type filter is based on a recursion for $\Delta P(k):=P(k|k)-P(k-1|k-1)$. The advantage is that for a time invariant model with $P_0=0$ the rank of these matrices is m. The doubling algorithm performs steps from time k to 2k instead of to k+1.

The steady state approach has been used successfully for large, two dimensional, models (eg. [Heemink, 1986, Bolding, 1995, Fukumori and Melanotte-Rizzoli, 1995]). Compared to a more traditional prediction by a deterministic model only, the number of additional computations is small while the reduction in errors can be large. A disadvantage is that the steady state approach can not be used for many applications because in many applications the model is not nearly linear or the measurements are irregular in time or space.

The product $A(k)P(k|k)A(k)'$ in equation 14 requires order n^3 computations if A(k) is a full matrix. For most finite difference methods the sparse structure can be used to reduce this to order n^2 computations. Even then this part of the equations remains the bottleneck. For this reason most approximate algorithms introduced in literature are aimed at reducing the number of computations in this part.

One possibility is to represent P(k|k), P(k+1|k) on a coarser grid and perform the covariance updates on this coarse grid. Let $x(k)=\Gamma x_c(k)$ represent an interpolation from the coarse grid to the fine grid. The equation 15 is changed to

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) \\ + \Gamma K_c(k+1)(y(k+1) - C(k+1)\hat{x}(k+1|k)) \quad (20)$$

Equation 13 remains unchanged and 14 and 16 are performed on the coarse grid. Several successful applications of this approach have been reported (eg. [Fukumori

and Melanotte-Rizzoli, 1995, Cohn and Todling, 1995]). For storm surge prediction the complicated patterns of the closed boundaries make it difficult to define an interpolation scheme that results in physically acceptable solutions. If for instance a bilinear interpolation scheme is used near the boundaries the component of the flow normal to the boundary is in general nonzero.

Recently Todling and Cohn [Cohn and Todling, 1995] introduced an approximate algorithm based on a singular value decomposition of the matrices A(k). It is well known that the best rank q approximation in the Frobenius norm as well as the spectral radius is given by setting all singular values from q+1 on to 0 [Golub and Van Loan, 1989]. Let A=UDV′ be the singular value decomposition of a matrix $A \in R^{n \times n}$, where U, V are orthogonal and D diagonal with elements $[D]_{i,i} = \sigma_i^2$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. The optimal rank q approximation is given by

$$A_{\text{approximate}} = [U]_{1:n,1:q}[D]_{1:q,1:q}[V]'_{1:n,1:q} \tag{21}$$

where $[]_{i_1:i_2,j_1:j_2}$ denotes the submatrix with rows i_1 through i_2 and columns j_1 through j_2. The leading singular vectors and singular values can be computed efficiently using a Lanczos algorithm [Golub and Van Loan, 1989]. If there are only a few relatively large singular values, the matrix can be approximated quite well with q<<n. Although the method works well it was outperformed by another method (PEKF [Cohn and Todling, 1995]) introduced in the same article.

Instead of approximating A(k) one can also try to approximate P(k|k), P(k+1|k). In addition to reducing the number of computations it is often also possible to reduce storage requirements in this case.

One way to approximate P(k|k), P(k+1|k) is to set correlations over large distances to zero [Parrish and Cohn, 1985]. This is also called a banded approximation since in one dimension the resulting matrix becomes a band matrix. The sparse structure of the approximate P(k|k), P(k+1|k) can be used to reduce the number of computations. Since the matrices P(k+1|k) and P(k|k) represent covariance matrices, they should be positive semidefinite. If due to approximations this is not true for the computed matrices, this may cause divergence of the solution. The square root algorithms as introduced by Potter (see [Maybeck, 1979, Bierman, 1977] for an introduction) avoid this problem by using the square root of the error covariance matrix, P(k|l)=L(k|l)L(k|l)′, with L(k|l) a lower triangular matrix. Because the factors L(k|l) have a much smaller range of the eigenvalues these algorithms are also numerically better conditioned then the original Kalman filter algorithm.

The time update of the L matrix is given by

$$L(k+1|k) = [A(k)L(k|k), F(k)\textstyle\sum_s^{1/2}]U(k) \tag{22}$$

where U(k) is a unitary matrix such that the last m rows of the first factor on the right hand side become zero. Usually Householder reflections or the Modified Gramm Schmidt procedure is used for this. It is easily shown that the multiplication with a unitary U(k) does not change P(=LL′). $\sum_s^{1/2}$ is a square root factor of $\sum_s$, i.e. $\sum_s^{1/2}(\sum_s^{1/2})' = \sum_s$. The notation [,] means that a larger block matrix is built from the two submatrices A(k)L(k|k) and $F(k)\sum_s^{1/2}$.

The measurement update for scalar measurements p=1 is given by

$$H(k+1) = L(k+1|k)'C(k+1)' \tag{23}$$

$$\gamma(k+1) = (H(k+1)'H(k+1)\sum_0(k+1))^{-1} \tag{24}$$

$$K(k+1) = L(k+1|k)H(k+1)\gamma(k+1) \tag{25}$$

$$\begin{aligned} L(k+1|k+1) &= L(k+1|k) - K(k+1)H(k+1)' \\ &\quad (1+(\gamma(k+1)\sum_0(k+1))^{1/2})^{-1} \end{aligned} \tag{26}$$

If there is more than one measurement at one time then the measurements are transformed using $\sum_0^{\frac{1}{2}}$. The resulting independent measurements can be processed one at a time.

Though more robust square root algorithms are in general not more efficient than the standard Kalman filter algorithm and therefore the square root filter equations described above can not be used directly for large scale models.

Recently Boggs [Boggs et al., 1995] proposed a banded approach on the square root of the error covariance. In one dimension the Cholesky factor of a banded error covariance matrix is also banded and can be used to reduce the computational burden. In two dimensions L contains more nonzero elements than P if the error covariance P is banded.

Another method is to approximate P(k|k), P(k+1|k) by one of a lower rank. Since the matrices P(k|k), P(k+1|k) are symmetric the singular value decomposition reduces to the eigen decomposition. The error covariance matrix is approximated using the largest eigenvectors and eigenvalues. Todling and Cohn [Cohn and Todling, 1995] and Verlaan and Heemink [Verlaan and Heemink, 1995] proposed approximate Kalman filter algorithms based on this idea. The Partial Eigendecomposition Kalman Filter (PEKF) by Todling and Cohn uses a Lanczos type algorithm to efficiently compute the largest eigenvalues and the corresponding vectors [Anderson and Moore, 1979]. The Reduced Rank Square Root Filter (RRSQRT) uses a square root like algorithm to update the decomposition. The reduced rank structure allows the eigenvalues and vectors to be updated efficiently. The RRSQRT algorithm will be explained in detail in the sequel.

5 The reduced rank square root filter

The square root factorization of a positive semidefinite matrix is not unique. The lower triangular form or Cholesky decomposition used in most square root filtering algorithms was chosen because of computational efficiency. Unfortunately this form does not allow for easy approximation. For the Reduced Rank Square Root algorithm the square root factors are based on the eigendecomposition. If P=UDU′ is the eigendecomposition of the error covariance matrix P then $L=UD^{1/2}$ is a square root factor of P. The error covariance matrix is now approximated by using q leading eigenvalues only. If the eigenvalues are ordered, i.e. $[D]_{1,1} \geq \cdots \geq [D]_{n,n} \geq 0$, then the approximation can be accomplished by truncating after the first q columns of the square root factor L.

The steps of the RRSQRT algorithm resemble those of the square root filtering algorithm (22-26). The three main steps are the "time-step" the "reduction-step" and the "measurement-step".

5.1 "Time-step"

The "time-step" performs the time propagation of the estimate and error covariance and is equivalent to equations 13,14 of the Kalman filter equations. The equations are

$$\hat{x}(k+1|k) = A(k)\hat{x}(k|k) + B(k)u(k) \tag{27}$$

$$L(k+1|k) = [A(k)L(k|k), F(k)\textstyle\sum_s(k)^{1/2}] \tag{28}$$

where L(k|k) is the n by q estimate square root of the error covariance P(k|k). The multiplication A(k)L(k|k) in equation 28 is much faster than the one in equation 22, since in the RRSQRT algorithm the matrix L(k|k) contains q columns instead of n and q<<n. The connection between equations 28 and 14 can be seen from

$$\begin{aligned} P(k+1|k) &= L(k+1|k)L(k+1|k)' && (29)\\ &= [A(k)L(k|k), F(k)\textstyle\sum_s(k)^{1/2}][A(k)L(k|k), F(k)\sum_s(k)^{1/2}]' && (30)\\ &= A(k)L(k|k)L(k|k)'A(k)' + F(k)\textstyle\sum_s(k)F(k)' && (31)\\ &= A(k)P(k|k)A(k)' + F(k)\textstyle\sum_s(k)F(k)' && (32) \end{aligned}$$

5.2 "Reduction-step"

The addition of rows, $F(k)\sum_s(k)^{1/2}$ in equation 28, for the system noise every time-step would quickly increase computation times. Therefore the number of columns is reduced to q after every "time-step". The concept of this approximation is to use only the first q leading eigenvalues and eigenvectors of the error covariance matrix L(k+1|k)L(k+1|k)'. In order to compute this efficiently first the eigendecomposition of the matrix L(k+1|k)'L(k+1|k) is determined:

$$L(k+1|k)'L(k+1|k) = V(k+1)E(k+1)V(k+1)' \tag{33}$$

It can easily be shown that

$$(L(k+1|k)V(k+1)E^{-1/2}(k+1))(E(k+1))(L(k+1|k)V(k+1)E^{-1/2}(k+1))' \tag{34}$$

is the eigendecomposition of L(k+1|k)L(k+1|k)' and thus

$$L^*(k+1|k) = [L(k+1|k)V(k+1)]_{1:n,1:q} \tag{35}$$

is the square root of the optimal rank q approximation of L(k+1|k)L(k+1|k)'.

The above procedure is much faster than eigenvalue computations on the matrix L(k+1|k)L(k+1|k)' or singular value computations on L(k+1|k), which could also accomplish the task of reduction. This is caused by the fact that the matrix L(k+1|k)L(k+1|k)' is a q+m by q+ m matrix and q<<n, m<<n.

5.3 "Measurement-step"

The measurement-update equations (23-26) of the square root algorithm do not depend on the specific type of square root factor used and can thus also be used for the RRSQRT algorithm, although the dimensions of some of the matrices are now different.

$$H(k+1) = L(k+1|k)'C(k+1)' \tag{36}$$

$$\gamma(k+1) \;=\; (H(k+1)'H(k+1) + \textstyle\sum_0(k+1))^{-1} \tag{37}$$

$$K(k+1) \;=\; L(k+1|k)H(k+1)\gamma(k+1) \tag{38}$$

$$\begin{aligned} L(k+1|k+1) \;&= L(k+1|k) - K(k+1)H(k+1)' \\ &\quad (1 + (\gamma(k+1)\textstyle\sum_0(k+1))^{1/2})^{-1} \end{aligned} \tag{39}$$

Independent measurements can be processed one at a time. If the measurements are correlated, ie. $\sum_0$ is not diagonal then these measurements can be transformed. Let $\tilde{y}(k)$ be defined by

$$\tilde{y}(k) : = \; \textstyle\sum_0^{-\frac{1}{2}} y(k) \tag{40}$$

where $\sum_0^{-\frac{1}{2}}$ is the matrix inverse of the Cholesky factor of $\sum_0$. Then

$$\tilde{y}(k) \;=\; \tilde{C}(k)x(k) + \tilde{v}(k) \tag{41}$$

where

$$\tilde{C}(k) : = \; \textstyle\sum_0^{-\frac{1}{2}} C(k) \tag{42}$$

$$\tilde{v}(k) : = \; \textstyle\sum_0^{-\frac{1}{2}} v(k) \tag{43}$$

These transformed measurements are equivalent to the original measurements, but the covariance matrix of the errors of $\tilde{v}(k)$ is the identity matrix.

5.4 Initialization

For many applications the initial transient of the estimate is not important and $P(0|0)$ can be set to 0. In this case $L(0|0)$ also becomes 0. If this is not the case then $P(0|0)$ or $P(0|-1)$ can be approximated using the q leading eigenvectors and eigenvalues, for which a Lanczos type algorithm [Golub and Van Loan, 1989] can be used.

The columns of L can be interpreted as error vectors in the state space. In some respects these columns are a generalization of the 'modes' of a system and will therefore also be called modes.

The number of computations required in the time propagation of the error in the covariance, which is a major fraction of the total number, is reduced by a factor $\frac{n}{q}$ with respect to the original Kalman filter algorithm. It can be shown that for q=n the RRSQRT algorithm is exact in the sense that it is equivalent to the Kalman filter equations. The parameter q controls the accuracy of the approximation. The price for greater accuracy is as always a larger computational burden.

6 Experiments

To evaluate the performance of the RRSQRT algorithm some experiments were performed. The measurements were generated using the same linear model as for the Kalman filter. Contrary to Using field data this allows for comparison between the true state and the estimate of the state. Although the RRSQRT algorithm is especially suited for nonlinear and time-varying models, a linear time-invariant model

was used, so that results can be compared with the optimal estimate, which in this case can be obtained using a Chandrasekhar filter algorithm. Comparisons using field data and a non-linear model will be performed in the future.

For simplicity the effects of wind and errors in the wind are neglected in all but the last experiment. The remaining uncertainty on the open boundary is modelled using an AR(1) model. The system noise is inserted only at a few points on the boundary. For points in between the values are interpolated.

The area of the North Sea covered by the model is shown in Figure 1. The grid size used is Δx=6400[m] and Δy=6400[m]. This results in a 95 by 60 grid with 2669 computational gridpoints. The time-step is ΔT=1800[s]. The friction coefficient is μ=2.4.10^{-3}. The system noise at the northern boundary is generated by an AR(1) process with correlation α_1=0.9 over one time-step and white noise with standard deviation $\sigma=\sqrt{0.2}$[m]. The noise is generated at four points (m=2, 30, 60, 93, n=60) with linear interpolation in between. The measurements are generated at A (m,n)=(8,55), B (m,n)=(10,40) and C(m,n)=(40,18). The measurement errors are assumed to be independent with standard deviation σ=0.1[m]. The initial estimate of the state is 0. The initial error covariance is also 0. The number of modes is q=20.

As a first test a periodic Kelvin wave is generated at the western part of the boundary. Snap-shots of the true and estimate waterlevels are shown in Figure 2. Only measurement station A was used for the assimilation. The figure shows that without noise the filter is able to track the state with only one measurement station.

For the next experiment system noise and measurement noise were generated using a random number generator, with variance according to the values above. Where possible this realization of the noise is used for the other experiments, too. Again only measurements from station A were used. Figures 3, 4 shows the RMS error for various settings of the number of modes q, as well as the optimal Kalman filter estimate, which in this special case of a time invariant linear model with zero initial noise can be computed efficiently using the Chandrasekhar algorithm. The RMS values are computed using

$$e(k) = \sqrt{(\hat{x}(k|k) - x(k))'(\hat{x}(k|k) - x(k))} \tag{44}$$

This shows how well the filter is doing. How well the filter 'thinks' it is doing can be seen from the computed error covariance P(k|k). The error e can be compared with the estimate e' given by

$$e'(k) = \sqrt{\text{trace}P(k|k)} \tag{45}$$

which can be computed from the sum of the eigenvalues in the RRSQRT algorithm. In Figures 3, 4 the true and estimated RMS values are shown for several number of modes.

The results show that for 5 modes the system with filter is probably unstable, for 10 modes the filter works well and for 20 modes the results are almost the same as for the full Kalman filter. The RRSQRT algorithm systematically underestimates the errors it is making, but with more modes the estimated error variance grows to the value of the full Kalman filter.

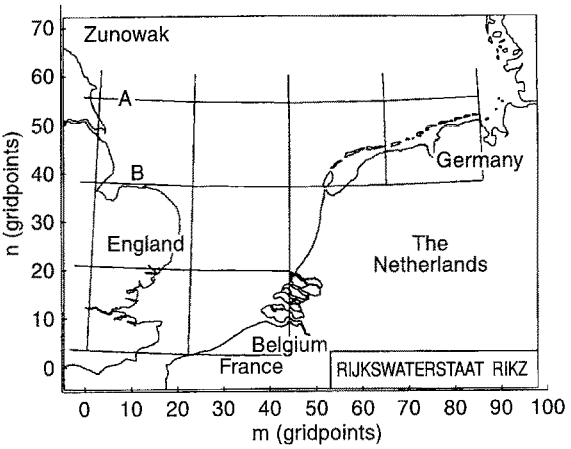

Figure 1. Area covered by the model.

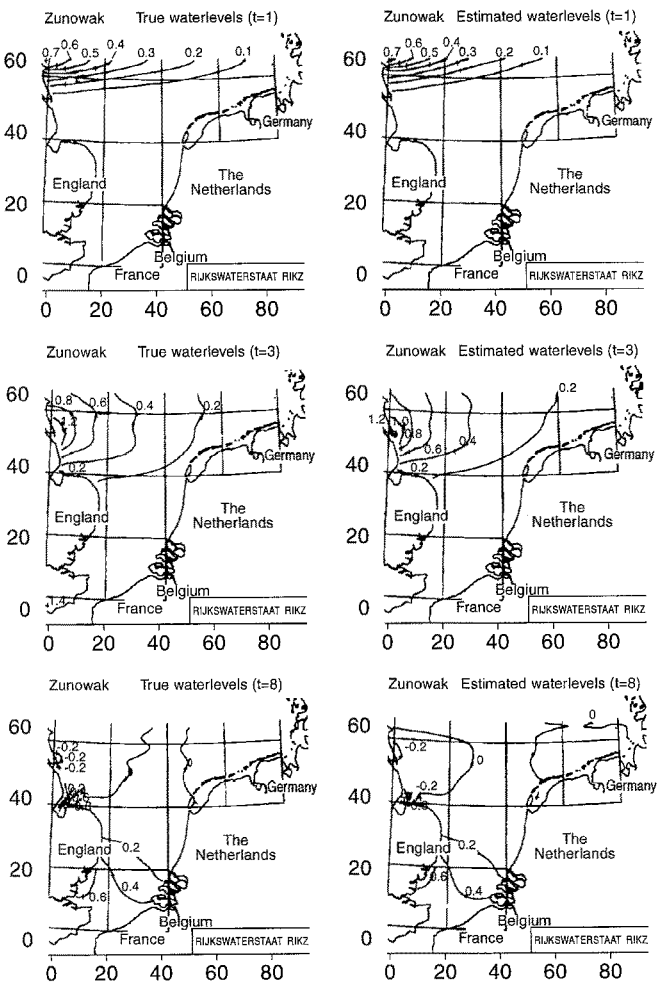

Figure 2. True and estimated waterlevels.

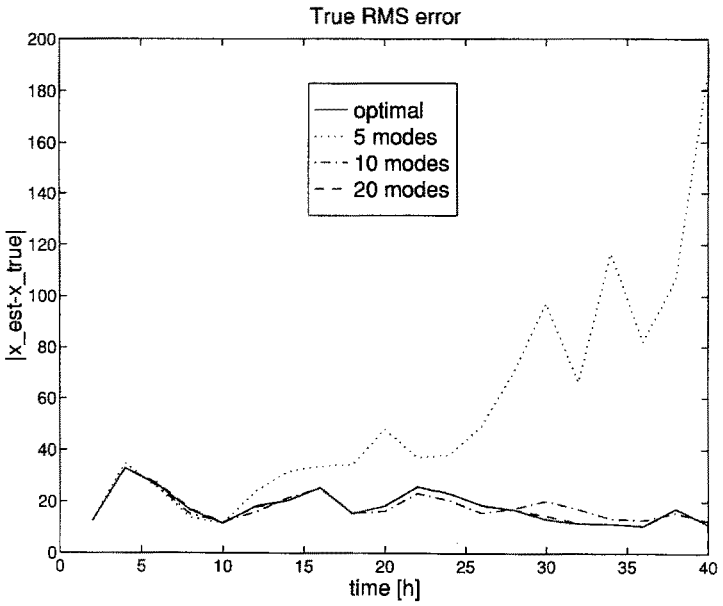

Figure 3. True RMS.

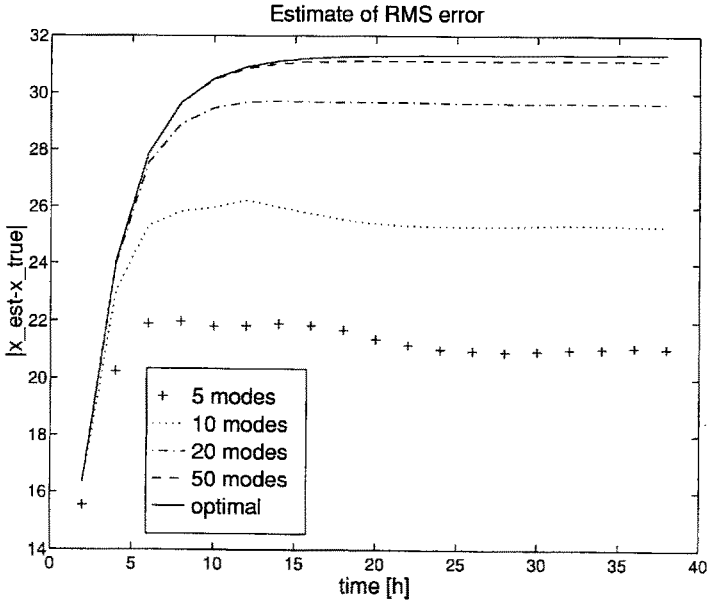

Figure 4. Estimated RMS.

In Figure 5 the computed eigenvalues are shown for various numbers of modes. These plots show that the RRSQRT algorithm mainly underestimates the errors in the smaller (faster) modes. The large range in eigenvalues is the reason accurate low order approximations of the error covariance can be made.

Since the optimal gain can be computed in this special case it is possible to determine the relative error in the gain. The Frobenius norm was used for these computations. The results are shown in Figure 6.

An important aspect of the RRSQRT algorithm is the number of modes needed to get a good approximation since the computation time is proportional to this. The number of modes can be used as a trade off between the number of computations needed and the approximation error of the algorithm. The number of modes needed will also depend on the model used, the values of the parameters, the position and number of measurements and the system noise and measurement noise. This number is not known advance, but has to be determined from experiments. Therefore the remaining experiments will study the sensitivity of the RRSQRT algorithm.

When the measurement position is changed this has little influence on the magnitude of the approximation errors. Figure 8 shows the relative error in the gain when only the measurements in B are used. There is almost no difference with Figure 6. The estimates however do change.

If all the measurements from A,B and C are used the relative error in the gain grows, as can be seen in Figure 9. This can be expected since the complexity of the data assimilation increases. By increasing the number of modes this error can be decreased again.

In the next experiment the system noise on the open boundary is correlated in space. In time the same AR(1) model as before is used. An exponential correlation model in space was used with a decorrelation length of 60 gridpoints. The correlation reduces the effective number of degrees of freedom for the system noise. Because of this it is expected that the truncation error of the RRSQRT algorithm will decrease. In Figure 10 the relative error in the gain for this experiment is shown. It indeed shows a small decrease in error, but the change is very small while quite large correlations were used.

In the last experiment additional uncertainty in the wind-stress was introduced. An AR(1) model was used to model correlation in time (α=0.9) and an exponential correlation function was used for correlation in space with decorrelation length 19.5 gridpoints and driven by white noise with standard deviation 0.001. To reduce the number of noise input variables the uncertainty was introduced on a subgrid (m=1,24,47,60, 83, n=1,20,39,58). Measurements from stations A,B and C were used for assimilation.

Figure 11 shows the true RMS and estimated RMS values for 50 modes. For fewer than 50 modes (20,30,40) the filter is unstable. For 100 modes the RMS values are very close to the optimal values. The additional uncertainty introduced by the wind requires the use of more modes. Much of the additional uncertainty can not be estimated from three waterlevel stations, but this uncertainty does occupy storage in the estimate of the error covariance.

From these experiments is seems that given a model the number of modes needed depends mainly on the number of independent noise inputs. The size of the errors introduced by the approximations in the filter is not very sensitive to changes of parameters.

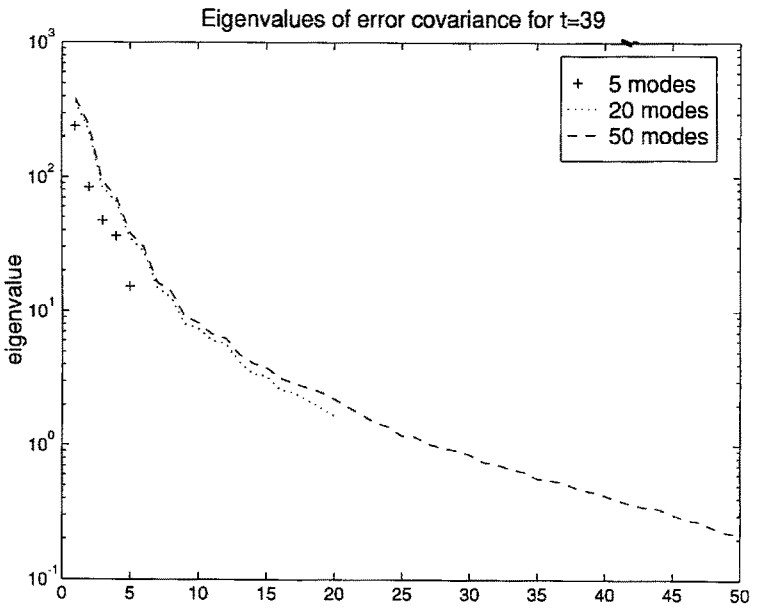

Figure 5. Eigenvalues of the approximate error covariance matrix.

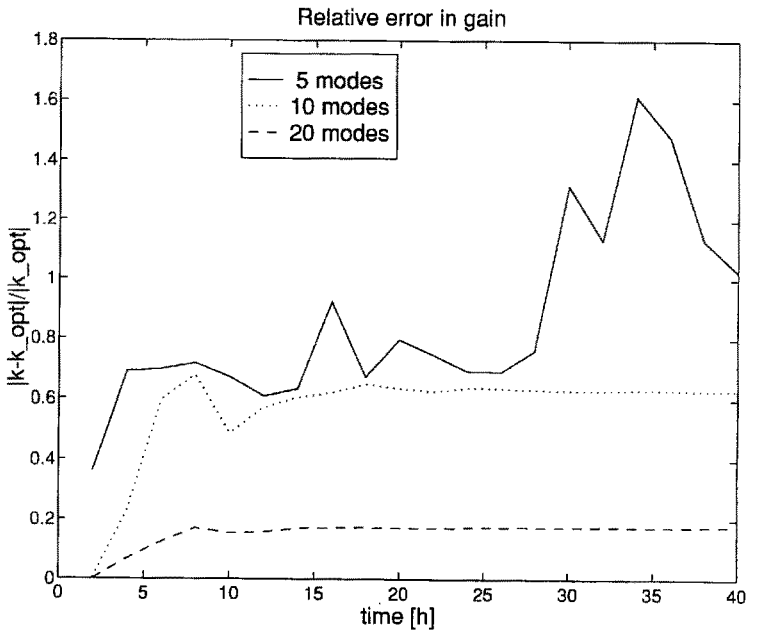

Figure 6. Relative error in the gain matrix.

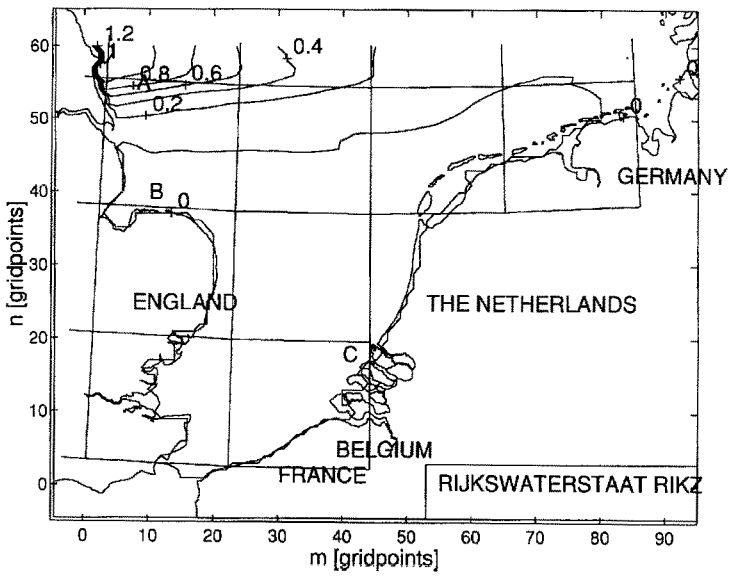

Figure 7. Part of the gain associated with the waterlevels.

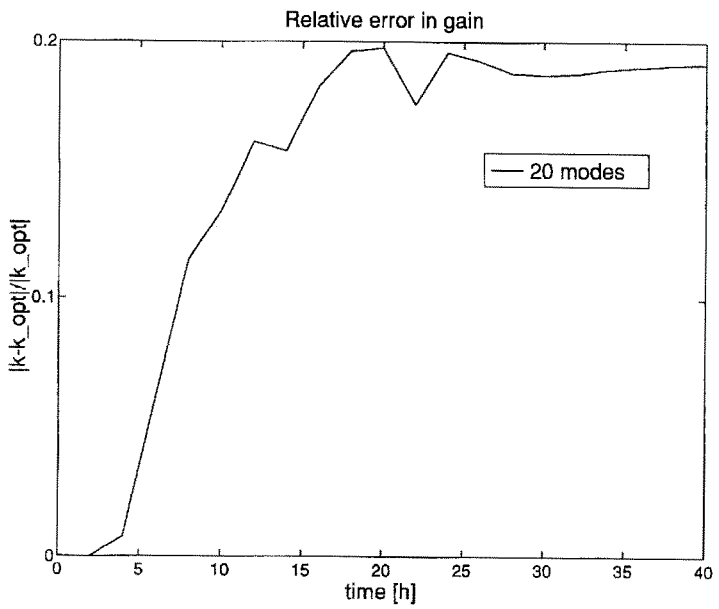

Figure 8. Relative error in gain when using only measurements from B.

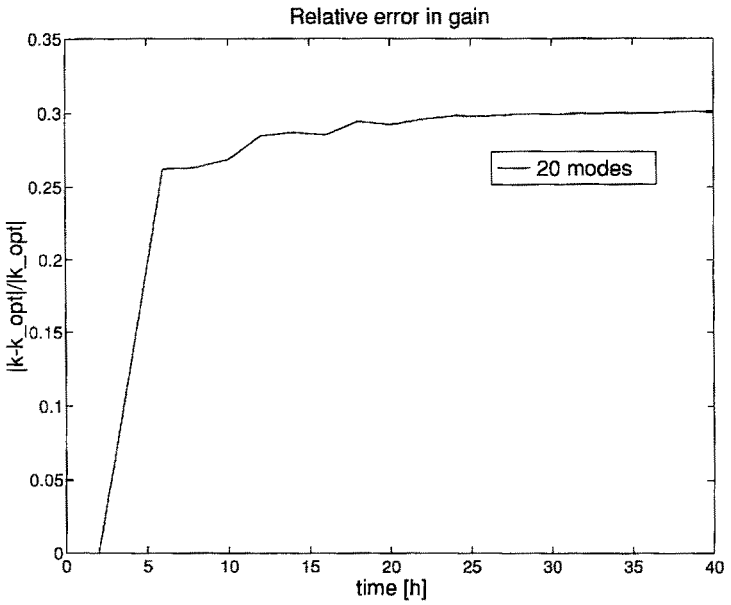

Figure 9. Relative error in gain when using only measurements from B using measurements from A,B and C.

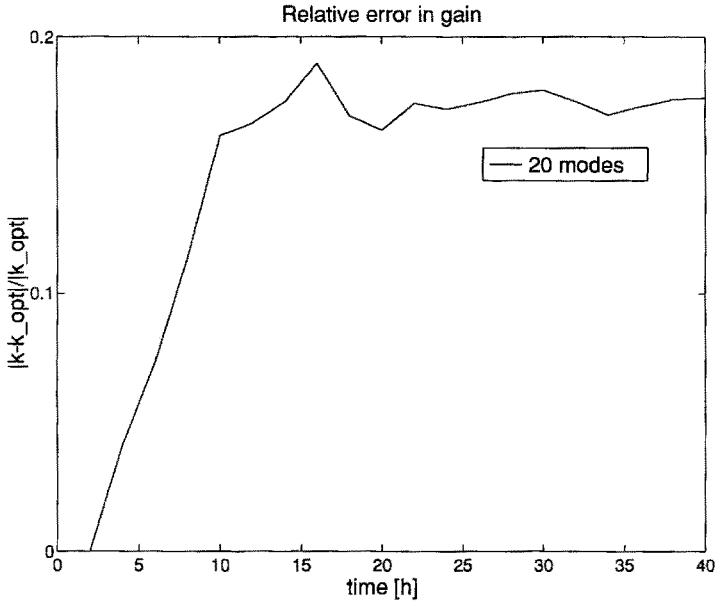

Figure 10. Relative error in gain when using correlated noise on the open boundary.

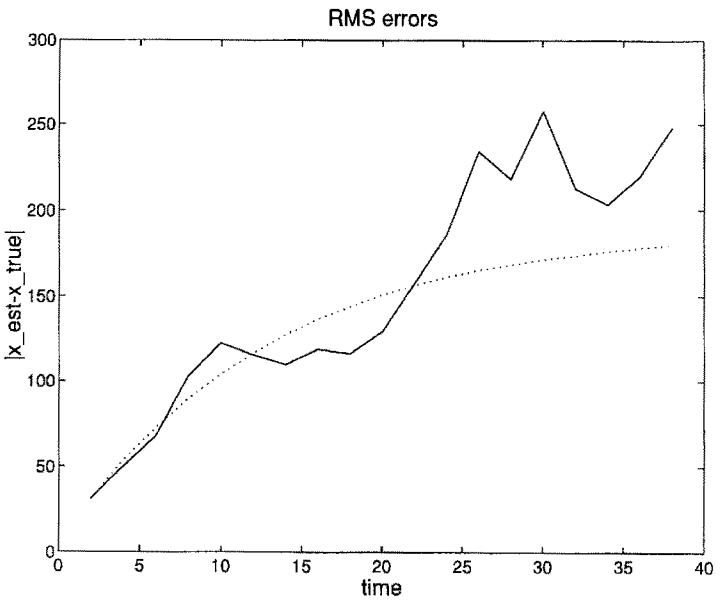

Figure 11. True and esatimated RMS error for situation with errors in wind.

7 The RRSQRT filter as extended Kalman filter

So far, this work dealt with estimation for linear models. For non-linear models the RRSQRT algorithm can be adapted to approximate the extended Kalman filter. The changes needed are conceptually not very difficult. The time propagation of the estimate is performed using the non-linear model. The time propagation of the error covariance estimate uses the tangent model $\frac{\partial f}{\partial x}$, which is linearization of the model around the current estimate. The main difficulty in practice is that the derivation and implementation of the tangent linear model is a lot of work. To avoid the use of a tangent linear model a method based on finite differences is proposed here.

For the extended Kalman filter the matrix A(k) of equation 28 is replaced by $\frac{\partial f}{\partial x}$ (see eqn. 7) evaluated at the latest estimate $\hat{x}(k|k)$. Let the i'th column of L(k|k) be denoted by $l_i(k|k)$ then

$$\begin{aligned} \frac{\partial f}{\partial x} L(k|k) &= \frac{\partial f}{\partial x} [l_1(k|k), \cdots, l_q(k|k)] \\ &= \left[\frac{\partial f}{\partial x} [l_1(k|k), \cdots, \frac{\partial f}{\partial x} l_q(k|k) \right] \end{aligned} \tag{46}$$

The column vector $\frac{\partial f}{\partial x} l_i(k|k)$ can be approximated by

$$\frac{\partial f}{\partial x} l_i(k|k) \approx \frac{f(\hat{x}(k|k) + \varepsilon l_i(k|k)) - f(\hat{x}(k|k))}{\varepsilon} \tag{47}$$

where ε is small. And thus

$$\frac{\partial f}{\partial x} L(k|k) \approx \frac{f(\hat{x}(k|k) + \varepsilon l_1(k|k)) - f(\hat{x}(k|k))}{\varepsilon}, \cdots, \frac{f(\hat{x}(k|k) + \varepsilon l_q(k|k)) - f(\hat{x}(k|k))}{\varepsilon} \tag{48}$$

For the computation of $\frac{\partial f}{\partial x} L(k|k)$ q+1 evaluations of f are needed, but $f(\hat{x}(k|k))$ is also needed for equation 27. Usually the number of computations needed for an evaluation like $\frac{\partial f}{\partial x} l_i(k|k)$ is close to the number of computations needed for an evaluation of f(.). In this case the proposed method requires approximately the same number of computations as the extended RRSQRT kalman filter using a tangent linear model, the effort needed for implementation is however considerably less.

For 'small' non-linearities the extended Kalman filter, either using a tangent linear model or finite differences is expected to yield good results. For 'strong' non-linearities or discontinuities however this may fail. An important discontinuity in tidal flow models is when area's with a height close to the mean sea level are flooded. Future research will address the problems posed by non-linear aspects of tidal-flow models.

8 Conclusions

In this paper we introduced a new filter algorithm for data assimilation for large scale systems. The algorithm is based on a reduced rank approximation of the error covariance matrix using a square root factorization. The use of the factorization ensures that the error covariance matrix remains positive semi-definite at all times, while the smaller rank reduces the number of computations and storage requirements.

The algorithm performs very well in the experiments shown. In these experiments a large reduction in computations could obtained. Although more analysis and experiments are needed the algorithm seems very promising. The methods used are generic and can be applied to various types of data assimilation problems. It is straight forward to use the algorithm as a modified version of an Extended Kalman Filter. Moreover, also a method was suggested to avoid the use of a tangent linear system in this case.

In the future the RRSQRT algorithm will be applied for storm surge forecasting using a non-linear model of the North-Sea based on the shallow water equations combined with the proposed finite-difference approach. Research will focus on non-linear effects and application using real-life data.

Acknowledgements

This work has been carried out in cooperation with and with financial support from the RIKZ.

References

Anderson, B.; Moore, J. 1979: Optimal Filtering. Prentice-Hall, Englewood Cliffs

Bierman, G. 1977: Factorization methods for discrete sequential estimation, Mathematics in Science and Engineering 128. Academic Press, Academic Press, NY

Boggs, D.; Ghil, M.; Keppenne, C. 1995: A stabilized sparsematrix u-d square-root implementation of a large-state extended kalman filter. In Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, 219-224. World Meteorological Organization, WMO

Bolding, K. 1995: Using a kalman filter in operational storm surge prediction. In Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, 379-383. World Meteorological Organization, WMO

Box, G.; Jenkins, G.; Reinsel, G. 1994: Time Series Analysis. Prentice Hall, Englewood Cliffs, 3 edition

Brummelhuis, P.T. 1992: Parameter Estimation in Tidal Models with Uncertain Boundary Conditions. PhD thesis, Technical University Delft, Delft

Budgell, W. 1986: Nonlinear data assimilation for shallow water equations in branched channels. J. Geophys. Res. 10, 633-644

Cohn, S.; Todling, R. 1995: Approximate kalman filters for unstable dynamics. In Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, 241-246. World Meteorological Organization, WMO

Curi, R.; Unny, T.; Hipel, K.; Ponnambalam, K. 1995: Application of the distributed parameter filter to predict simulated tidal induced shallow water flow. Stochastic Hydrology and Hydraulics 9, 13-32

Fukumori, I.; Melanotte-Rizzoli, P. 1995: An approximate kalman filter for ocean data assimilation; an example with an idealized gulf stream model. Journal of Geophysical Research 100, 6777-6793

Golub, G.; Van Loan, C. 1989: Matrix Computations. John Hopkins University Press, 2nd ed. edition

Heemink, A. 1986: Storm Surge Prediction Using Kalman Filtering. PhD thesis, Twente University of Technology

Heemink, A.; Kloosterhuis, H. 1990: Data assimilation for non-linear tidal models. International Journal for Numerical Methods in Fluids 11, 1097-1112

Kalman, R. 1960: A new approach to linear filter and prediction theory. J. Basic. Engr. 82D, 35-45

Kalman, R.; Bucy, R. 1961: New results in linear filtering and prediction theory. J. Basic. Engr. 83D, 95-108

Maybeck, P. 1979: Stochastic models, estimation, and control, Mathematics in Science and Engineering 141-1. Academic Press, New York, NY

Morf, M.; Sidhu, G.; Kailath, T. 1974: Some new algorithms for recursive estimation in constant, linear, discrete-time systems. IEEE Transactions on Automatic Control AC-19(4), 315-323

Parrish, D.; Cohn, S. 1985: A kalman filter for a twodimensional shallow-water model: Formulation and preliminary experiments. office note 304, New York University

Stelling, G. 1984: On the Construction of Computational Methods for Shallow Water Flow Problems. PhD thesis, Delft University of Technology. Rijkswaterstaat Communications no.35

Todling, R.; Cohn, S. 1994: Suboptimal schemes for atmospheric data assimilation based on the kalman filter. Monthly Weather Review. in press

Verlaan, M.; Heemink, A. 1995: Reduced rank square root filters for large scale data assimilation problems. In Second International Symposium on Assimilation of Observations in Meteorology and Oceanography, 247-252. World Meteorological Organization, WMO