Data Assimilation and Nested Hydrodynamic Modelling in Storm Surge Forecasting

Rafael Cañizares$^1$ and Henrik Madsen$^1$

Abstract

A data assimilation method for state updating in a hydrodynamic model is presented. The method is based on the extended Kalman filter in which the error covariance matrix is approximated by a matrix of lower rank using a square-root factorisation (reduced rank square-root filter). Results from a test of the Kalman filter in a regional model of the North Sea and Baltic Sea are presented. In this respect, the influence of using nested hydrodynamic models together with data assimilation techniques is illustrated and discussed. The test reveals that assimilation of water level measurements from coastal stations significantly improves the model results.

Introduction

During the last decades, the interest of predicting water level rising as a consequence of storm surges has grown considerably. In countries that are affected by this phenomenon substantial efforts are made to predict storm surges so far ahead in time that appropriate actions can be taken.

The combination of numerical weather prediction models and hydrodynamic models forms the main frame of an operational storm surge forecast system (Bode and Hardy, 1997). The hydrodynamic model uses the predicted meteorological wind and pressure data to provide a prediction of the water level field. The storm surge prediction, however, is not always as accurate as desired which can mainly be ascribed to

1. Simplifications of the description of the physical processes in the numerical hydrodynamic model.
2. Bias in the meteorological forcing prediction.

---

$^1$International Research Centre for Computational Hydrodynamics and Danish Hydraulic Institute, Agern Allé 5, DK-2970 Hørsholm, Denmark. rct@dhi.dk; hem@dhi.dk
3. Inaccurate open boundary conditions.

The main errors in the hydrodynamic model are, respectively, simplifications of the physical processes in the numerical model equations and application of a too coarse spatial resolution to adequately resolve the dynamics of the system. To reduce the errors caused by the former problem more complex models should be considered, whereas for the latter problem the model has to be applied in a finer grid which can be defined locally by using nested hydrodynamic models (Vested et al, 1995). Nested models allow to have a finer resolution in areas where required while a coarser resolution can be maintained in the rest of the model domain.

The uncertainties of the meteorological forcing and the open boundary conditions as well as uncertainties of the physical model parameters can be accounted for by using data assimilation. In data assimilation, model and measurements of the system are combined in order to obtain a better estimate of the state of the system. Data assimilation algorithms, however, can become prohibitive when applied in large scale models due to the huge computational cost associated with this kind of applications. New algorithms, such as the reduced rank square-root (RRSQRT) Kalman filter (Verlaan and Heemink, 1995) have solved this problem. The RRSQRT filter is a suboptimal scheme of the extended Kalman filter that uses a square-root algorithm as well as a lower rank approximation of the error covariance matrix.

The RRSQRT filter has been integrated into an existing hydrodynamic modelling system that solves the vertically integrated equations of continuity and momentum in two horizontal directions (Cañizares et al, 1998). When measurements are available, the Kalman filter is adopted for updating the state of the system. In storm surge forecasting, the updated state is then used as initial conditions for the forecast simulation.

The paper is organised as follows. In the first section the applied numerical hydrodynamic model is described. In sections two and three the general Kalman filter updating scheme and the RRSQRT filter are presented and the implementation in the hydrodynamic model is described. Sections four and five outline the specific features of the nested version of the hydrodynamic model and the implementation of the filter in this type of model. Finally, a test case is presented where the Kalman filter is applied in a regional model of the North Sea and Baltic Sea.

The deterministic numerical model

In the present study, the data assimilation method has been implemented in the hydrodynamic module of the MIKE 21 modelling system which solves the vertically integrated equations of continuity and conservation of momentum (shallow water equations) in two horizontal directions (DHI, 1995). The equations that are solved are those of:
Continuity:

\[
\frac{\partial \xi}{\partial t} + \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = S_0 - e
\]  

(1)

\[x\text{-momentum:}\]

\[
\frac{\partial p}{\partial t} + \frac{\partial}{\partial x} \left( \frac{p^2}{h} \right) + \frac{\partial}{\partial y} \left( \frac{pq}{h} \right) + \frac{gh}{C^2} \left( \frac{p^2 + q^2}{h^2} \right) - fVV_x \frac{h}{\rho_w} \frac{\partial}{\partial x} (P_a) - \Omega q - E \left( \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \right) = S_{ix}
\]  

(2)

and \(y\)-momentum:

\[
\frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left( \frac{q^2}{h} \right) + \frac{\partial}{\partial y} \left( \frac{pq}{h} \right) + \frac{gh}{C^2} \left( \frac{p^2 + q^2}{h^2} \right) - fVV_y \frac{h}{\rho_w} \frac{\partial}{\partial y} (P_a) - \Omega p - E \left( \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} \right) = S_{iy}
\]  

(3)

where:

- \(x, y\): Horizontal coordinates [m].
- \(t\): Time [s].
- \(h\): Water depth [m].
- \(\xi\): Surface elevation [m].
- \(p, q\): Flux densities in \(x\) and \(y\) directions \([m^3/s/m]\). \((p,q) = (v_x h, v_y h)\) where \(v_x\) and \(v_y\) are the depth averaged \(x\) and \(y\) velocities \([m/s]\).
- \(S_0\): Source magnitude per unit horizontal area \([m/s]\).
- \(S_{ix}, S_{iy}\): Source impulse in \(x\) and \(y\) directions \([m^2/s]\).
- \(e\): Evaporation rate \([m/s]\).
- \(E\): Eddy viscosity coefficient \([m^2/s]\).
- \(g\): Acceleration due to gravity \([m/s^2]\).
- \(C\): Chezy bed resistance coefficient \([m^n/s]\).
- \(\Omega\): Coriolis parameter \([s^{-1}]\).
- \(f\): Wind friction factor [-].
- \(V_x, V_y\): Wind speed and wind speed components in \(x\) and \(y\) directions \([m/s]\).
- \(P_a\): Atmospheric pressure \([kg/m/s^2]\).
- \(\rho_w\): Density of water \([kg/m^3]\).
At closed boundaries the flow perpendicular to the boundary is set to zero. At open boundaries the surface elevation is prescribed. With these boundary conditions and with prescribed initial values of surface elevations and flux densities, (1)-(3) form a well-posed boundary value problem.

MIKE 21 uses a finite difference approximation to solve the partial differential equations where the variables are defined on a space-staggered rectangular grid with surface elevations at grid points and fluxes midway between grid points (Leendertse, 1964). A time-centered alternating direction implicit (ADI) scheme is adopted. The equations are solved in one-dimensional sweeps, alternating between $x$ and $y$ directions. In the $x$-sweep, the continuity equation and the momentum equation in the $x$ direction are solved with respect to $\xi$ at time step $k+1/2$ and $p$ at time step $k+1$ using the known variables $\xi_k$, $p_k$, $q_{k-1/2}$, and $q_{k+1/2}$. In the $y$-sweep, the continuity equation and the momentum equation in the $y$ direction are solved with respect to $\xi_{k+1}$ and $q_{k+3/2}$ using $\xi_{k+1/2}$, $q_{k+1/2}$, $p_k$, and $p_{k+1}$.

Implementation of the Kalman filter in MIKE 21

For implementation of the Kalman filter in MIKE 21, the numerical model has to be formulated in a state-space form. The state variables to be considered are surface elevations and depth averaged $x$ and $y$ velocities in every point of the horizontal grid. The Kalman filter algorithm is based on a recursive two-time step formulation. The numerical scheme in MIKE 21, however, involves the $y$ velocity at three time steps. To express this scheme using only two time steps, the $y$ velocity at time steps $k+1/2$ and $k-1/2$ are included in the state vector. The numerical scheme based on (1)-(3) can then be written

$$x_k = \Phi(x_{k-1}, u_k)$$

where $x_k = (\xi_k, v_{k,x}, v_{y,k+1/2}, v_{y,k-1/2})$ is the state vector, and $u_k$ is the forcing of the system in terms of the surface elevations at open boundaries and the meteorological forcing components in the momentum equations (wind stress and pressure gradient).

For modelling the uncertainty of the system, it is assumed that model errors are mainly related to errors in the forcing components. The error processes are assumed to be less spatially variable than the water flow process (Heemink, 1990), and the discrete error processes can thus be defined on a grid $G_2$ that is coarser than the model grid $G_1$. A stochastic representation of the system equation (4) can then be written

$$x_k = \Phi(x_{k-1}, u_k + \Lambda \epsilon_k)$$

where $\epsilon_k$ contains the model error in every grid point of $G_2$, and $\Lambda$ is a matrix that represents the sequence of linear interpolations between $G_2$ and $G_1$.
For the system, measurements $z_k$ of the state are assumed to be available at certain points in the model grid $G_l$. The stochastic representation of the measurement equation reads

$$z_k = C_k x_k + \eta_k$$

where $C_k$ is a matrix that describes the relation between measurements and state variables, and $\eta_k$ is a random measurement error with zero mean and known covariance matrix $R_k$.

When measurements are available, cf. (6), the model forecast and the measurements can be combined to obtain an updated estimate of the state of the system. The Kalman filter update of the state vector and the error covariance matrix $P_k$ is given by

$$x_k^a = x_k^f + K_k (z_k - C_k x_k^f)$$
$$P_k^a = P_k^f - K_k C_k P_k^f$$

where $K_k$ is the Kalman gain matrix

$$K_k = P_k^f C_k' [C_k P_k^f C_k' - R_k]^{-1}$$

which serves as a weighting function of model forecast and measurements and depends on the associated errors $P_k^f$ and $R_k$. In (7)-(9) superscripts $f$ and $a$ refer to, respectively, forecast and analysis (or update).

For large systems, the propagation of the error covariance matrix (determination of $P_k^f$) is the main bottleneck. This step requires $2n$ as much computing effort as is required to advance the deterministic model itself ($n$ being the dimension of the state vector). Applications to large systems are prohibitive under such conditions, and hence approximations of the Kalman filter, reducing the computational effort, have to be used. The technique described below, the reduced rank square-root filter, introduced in (Verlaan and Heemink, 1995) and (Verlaan, 1998), is based on an approximation of the error covariance matrix.

The Reduced Rank Square Root (RRSQRT) filter

For non-linear model dynamics an extended Kalman filter can be formulated in which the propagation of the error covariance matrix is based on a statistical linearisation of the model equation. Considering a white noise process for the model errors, the forecast step is given by
where $Q_k$ is the covariance matrix of the system noise, defined on grid $G_2$.

The RRSQRT approximation of the extended Kalman filter uses a square-root algorithm as well as a lower rank approximation of the error covariance matrix (Verlaan and Heemink, 1995). Denote by $S$ the approximation of rank $M$ of the square root of the error covariance matrix. The propagation of the error covariance matrix is then given by

$$S_k' = \left[ F_k S_{k-1}' \right] G_k \Lambda Q_k^{1/2}$$

(12)

where $Q_k^{1/2}$ is the square-root of $Q_k$. The matrix $S_{k-1}'$ has $M$ columns where $M$ is chosen much smaller than the dimension $n$ of the state vector. To calculate the derivatives in $F_k$ and $G_k$ a finite difference approximation of $\Phi(\cdot)$ is adopted. Thus, the propagation of the error covariance matrix requires $M + p$ (the total number of noise points) model integrations, which is much smaller than the $2n$ integrations required in (11).

The propagation step in (12) increases the number of columns in the error covariance matrix from $M$ to $M + p$. To reduce the number of columns, and hence keep the rank of the matrix constant throughout the simulation, a lower rank approximation of $S_k'$ in (12) is applied by keeping only the $M$ leading eigenvectors of the error covariance matrix. The reduction can be achieved either by a singular value decomposition of $S_k'$ or by an eigenvalue decomposition of the matrix $(S_k')^T S_k'$ (Cañizares et al., 1998).

For uncorrelated measurement errors, a sequential updating scheme can be applied (Maybeck, 1979). This algorithm avoids the expensive calculation and storage of $P_k'$ as well as the matrix inversion in (9) for the calculation of the Kalman gain. The procedure used in the present study follows (Potter, 1967), i.e.

$$a_{k,i} = \left[ S_k' \right]^T C_i$$

(13)

$$\gamma_{k,i} = \frac{1}{a_{k,i}^2 + \sigma_{k,i}^2}$$

(14)

$$K_{k,i} = S_{k,i}^T a_{k,i} \gamma_{k,i}$$

(15)
\[ S_{k,i} = S_{k,i-1} - K_{k,i} a_{k,i}^T \frac{1}{1 + \sqrt{Y_{k,i} \sigma_{k,i}^2}}, \quad S_{k,0} = \sigma \]

\[ x_{k,i} = x_{k,i-1} + K_{k,i} \left[ z_{k,i} - C_i x_{k,i-1} \right], \quad x_{k,0} = x \]

where \( C_i \) is the \( i \)'th column of matrix \( C \), \( z_{k,i} \) is the \( i \)'th measurement, and \( \sigma \) is the standard deviation of the measurement noise.

The filter can use time-coloured noise. In this case the state vector is augmented with variables that represent the estimated value of the noise. For the propagation of the state vector and the error covariance matrix new equations have to be defined for the RRSQRT filter (see Madsen and Cañizares, 1998). Moreover, the error covariance matrix has to be normalised prior to the eigenvalue decomposition. For further details on the RRSQRT algorithm see (Cañizares, 1998).

**Special features of the MIKE 21 nested model**

The nested version of the MIKE 21 model solves the hydrodynamic equations simultaneously in a number of dynamically nested grids. An important difference between nesting and boundary transfer from a coarse model to a finer one is that in nesting the information between grids travel in two directions, i.e. from the coarse to the fine grid and vice versa. On the other hand, in a boundary transfer model information travel only from the coarse to the fine grid. The two techniques are also denoted two-way and one-way nesting, respectively.

In order to ensure model stability and smooth transition between areas, certain constraints are imposed. The most important are:

- Open boundaries can only be defined in the coarsest grid.
- The spatial resolution from one level to another is reduced by a fixed factor (grid reduction factor), which is equal to 3.
- The water depths in common grid points along the internal boundaries must be equal in both the coarse and the fine grid. Between the common points along the internal boundary, the water depths in the fine grid are linearly interpolated using the values at the common points.
- The water depth in the coarse grid has to be identical in three points orthogonal to the internal boundary (at the border and one point at each side). Therefore the first four points orthogonal to the internal boundary in the fine grid have the same water depth. The intention of these corrections is to avoid instabilities in the internal boundaries.

Further details about the nested model can be found in (DHI, 1995).
Implementation of the RRSQRT filter in the nested model

The main features of the implementation of the RRSQRT filter in the nested model are:

- The model is propagated using the complete nested model.
- The error covariance matrix is represented in the coarsest grid (the main area). Hence, the number of variables considered in this matrix is the same as if only the coarsest grid is considered.
- The error covariance matrix is propagated using the complete nested model. The model error is interpolated from the main grid to the internal grids.
- If the measurement position is located in a grid point of an internal area, the value is extrapolated to the main area. Thus, vector $C$, represents the relation between the position of the measurement in the fine grid and the surrounding positions in the coarsest grid.
- The Kalman gains $K_t$ correspond to variables in the main area. The gains are interpolated from the main grid to the internal grids.

Under these assumptions, the associated cost of the data assimilation scheme in the nested model is comparable to application in the model defined in the coarse area. The main difference is that the time associated with a run of the nested model is larger than for the model defined in the coarse area. For further details on the application of the RRSQRT in the nested hydrodynamic model see (Cañizares, 1998).

Application: a regional model of the North Sea and Baltic Sea

The Kalman filter has been applied to a regional model covering the North Sea and the Baltic Sea. Two open boundaries are defined in the North Sea between, respectively, Stavanger and Orkney Island (northern boundary) and Dover and Calais (southern boundary). A coarse model is defined with a grid size of 9 nautical miles (16670 m) in both directions. A nested area of the inner Danish waters has been defined in order to obtain a better and more detailed description of the water level and current fields in this area. The local model is defined in a grid with origin in (48,18) of the coarse grid and a grid size of 3 nautical miles (5667 m), i.e. one third of the grid size of the coarse model. The model setup and bathymetry are shown in Figure 1.

At the two open boundaries, the water level is specified. For the simulation period, wind velocity fields and pressure fields are available every three hours and they are linearly interpolated at every model time step (set equal to 10 min.). The flow resistance is defined with a Manning number equal to 32 $m^{1/3}/s$ in the entire model domain. The model is initialised on 01/10/97 at 00:00 with water level and velocity fields obtained from a spin-up simulation of 48 hours.

The performance of the RRSQRT filter in both the standard hydrodynamic model (HD) and the nested hydrodynamic model (NHD) is tested. In the simulations, water level data from 14 coastal stations are assimilated and the results are validated
against data from another 7 available coastal stations. The positions of the, in total, 21 stations are shown in Figure 1.

![Figure 1. North Sea and Baltic Sea model setup and bathymetry (depth in meters). A nested area is defined for the inner Danish waters. Water level stations are represented with circles (measurement stations) and squares (validation stations).](image)

A simulation period of three days from 01/10/97 00:00 to 03/10/97 00:00 was applied where continuous measurement of water levels were available at the 21 stations. Based on an initial sensitivity test the following parameters for the Kalman were used:

- The rank of the error covariance matrix is set equal to 100.
- The grid reduction factor between the noise grid and the coarse model grid is set equal to 8.
- Time-coloured noise is defined using a first order autoregressive model with a lag-one autocorrelation coefficient of 0.97.
- The noise in the meteorological forcing components of the momentum equation is defined using an exponential spatial correlation model with a correlation coefficient of 0.9 and a standard deviation that varies in space. The magnitude of the standard deviation varies from 0.0005 m²/s² in the North Sea to 0.0001 m²/s² in the Baltic Sea.
- At the northern boundary noise is defined using a spatial correlation coefficient of 0.95, standard deviation of 0.1 m, and a grid reduction factor of 3. At the southern boundary the same parameters are used except for the grid reduction factor, which is set to 1, i.e. it coincides with the model grid.
- The standard deviation of the measurement noise is set equal to 0.05 m.
Measurements were available every 20 minutes, and hence the updating step of the RRSQRT filter takes place every second time step.

In order to evaluate the performance of the filter the root mean square error (RMSE) between the observed and updated water levels are calculated and compared with the RMSE of the deterministic model simulation for both models. The RMSE has been calculated using the last 36 hours of simulation in order to reduce the influence from the initialisation of the filter. Figures 2 and 3 present the RMSE for the measurement and the validation stations, respectively, obtained from the deterministic and the updated HD and NHD models. Table 1 shows the global (spatial average) values of the RMSE for the different models.

Figure 2. RMSE for the deterministic (Det) and the updated (KF) hydrodynamic (HD) and nested hydrodynamic (NHD) models at measurement stations.

Figure 3. RMSE for the deterministic (Det) and the updated (KF) hydrodynamic (HD) and nested hydrodynamic (NHD) models at validation stations.
Table 1. Global (spatial average) RMSE values for the deterministic and the updated hydrodynamic (HD) and nested hydrodynamic (NHD) models.

<table>
<thead>
<tr>
<th></th>
<th>Deterministic RMSE (m)</th>
<th>Kalman filter RMSE (m)</th>
<th>Deterministic RMSE (m)</th>
<th>Kalman filter RMSE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HD</td>
<td>HD</td>
<td>NHD</td>
<td>NHD</td>
</tr>
<tr>
<td>Measurement stations</td>
<td>0.218</td>
<td>0.072</td>
<td>0.216</td>
<td>0.077</td>
</tr>
<tr>
<td>Validation stations</td>
<td>0.240</td>
<td>0.126</td>
<td>0.240</td>
<td>0.127</td>
</tr>
</tbody>
</table>

Figure 4. Water level and velocity field for the inner Danish waters at 02/10/97 19:00 calculated from the deterministic model (top) and the Kalman filter (bottom) using the hydrodynamic model (HD).
The HD and NHD deterministic models yield basically the same results. The use of a detailed description in the inner Danish waters improves the simulation only in some of the stations while providing worse results in others. This is a direct consequence of a poor schematisation of the area between Denmark and Sweden in the finer grid, resulting in worse results at the stations located at the southern part of the fine grid (Gedser and Klagshamn). The simulation at the other stations located in the nested area is improved using the finer resolution. No further efforts for calibrating the models have been done.

The Kalman filter for the HD as well as for the NHD model efficiently corrects the water levels in all parts of the regional model. At measurement points a global value of the RMSE equal to 0.072 m for the HD and 0.077 for the NHD have been obtained which is only slightly higher than the assumed standard deviation of the measurement noise (0.05 m). At validation points the corrections are not that significant but they still present a marked improvement (about 50% reduction of the RMSE) in all regions when compared with the deterministic models. In general, the results obtained in this case, in terms of reducing errors in water levels, are not improved by using a finer grid. Since the state vector is defined in the coarse grid, corrections in the fine area are obtained by interpolating the error estimated by the filter in the coarse area. The introduction of this new approximation causes a slightly worse performance of the filter in the finer grid. The use of a finer resolution is more important for corrections of the velocity field as illustrated below.

Figure 5. Water level and velocity field for the detailed area of the inner Danish waters at 02/10/97 19:00 calculated from the deterministic model (left) and the Kalman filter (right) using the nested hydrodynamic model (NHD).
Water level and velocity fields for the deterministic and updated models at 02/10/1997 19:00 are presented in Figures 4-5. Important differences in the velocity field can be observed in these figures. The flux entering the Danish waters from the North Sea is not well represented in the deterministic solution. The updated model provides higher water levels in the entrance to the Baltic Sea than in the deterministic model. In general, the updated model is able to reproduce the high water level elevation in the inner Danish waters and the northern coasts of Germany and Poland. When using a finer resolution, the velocity field is better represented. In this case, larger currents are obtained in the updated model, especially in areas with strong water level gradients. The type of global improvements presented in Figures 4-5 is maintained during the entire simulation.

Conclusions

A data assimilation method based on the RRSQRT filter has been implemented in a hydrodynamic model that solves the shallow water equations simultaneously in a number of dynamically nested areas with different resolutions. The use of a finer resolution grid in some areas of the model rapidly increases the number of computational points in the model. The state vector has been defined on the main (coarsest) grid in order to reduce the computational cost and the storage requirements of the Kalman filter. This implementation, however, did not improve the corrections of water levels in the detailed area as compared with the corrections obtained using only one area with a coarse resolution. Although this result may be partly caused by a poor schematisation of the nested model, more accurate results are expected if variables in all points of the nested grids are defined in the state vector, and hence avoiding the interpolation of the Kalman gain from the coarse to the fine grid.

The application example reveals that the use of data assimilation significantly improves the global model results. In storm surge forecasting, the improved estimate of the state of the system is then used as initial conditions for the forecast simulation. In the cases where the main interest is on water level predictions, as in the case in storm surge forecasting, the implementation using only one area of the regional model with a coarse resolution can provide sufficient accuracy. On the other hand, for prediction of currents more detailed information is required, and in this case nested modelling combined with data assimilation should be applied.

Acknowledgement

This study was funded by the Danish National Research Foundation through the International Research Centre for Computational Hydrodynamics (ICCH). Their support is greatly appreciated.

Appendix. References


