Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering

HENRIK MADSEN
DHI Water & Environment, Agern Allé 11, DK-2970 Horsholm, Denmark
hem@dhi.dk

DAN ROSBJERG, JESPER DAMGÅRD & FRANDS SØBJERG HANSEN
Environment & Resources DTU, Technical University of Denmark, Building 115, DK-2800 Kongens Lyngby, Denmark

Abstract A procedure is presented for assimilation of water levels and fluxes in the MIKE 11 Flood Forecasting (FF) system. The procedure implemented is based on the ensemble Kalman filter that provides a cost-effective and efficient updating and uncertainty propagation scheme for real-time applications. Up to the time of forecast, the model is updated according to the Kalman filter algorithm using the available measurements. In forecast mode, the Kalman filter provides an ensemble forecast that is used for estimation of water levels and fluxes in the river system and the associated uncertainties. A test example is presented where the MIKE 11 FF system is applied for flood forecasting in the Piedmont region in the northwestern part of Italy. Application of the ensemble Kalman filter significantly improves the forecast skills as compared to forecasting without data assimilation.

Key words data assimilation; ensemble Kalman filter; flood forecasting; Kalman filter; updating

INTRODUCTION

State-of-the-art real-time river flow and flood forecasting systems, such as the MIKE 11 Flood Forecasting (FF) system (DHI, 2002), include data assimilation (DA) facilities in which measured water levels or fluxes in the river system are used on-line to update the initial conditions of the hydrodynamic and hydrological forecast model. Different DA techniques exist and can be classified according to the variables that are modified in the assimilation process, i.e. input variables, model states, model parameters, and output variables (Refsgaard, 1997). For real-time applications, sequential or filtering procedures are often applied for state updating. One of the more advanced techniques in this group is the Kalman filter (KF).

The main advantage of the KF as compared to other DA techniques is that it explicitly takes model and data uncertainties into account in the updating process and provides an estimate of the model prediction uncertainty. However, when applied to high dimensional modelling systems as considered herein, the standard KF (or extended KF for nonlinear models) imposes unacceptable computational burdens for real-time applications. In recent years, several KF schemes have been formulated that use different approximations of the covariance modelling to reduce the computational costs. State-of-the-art procedures which have been implemented in hydrodynamic and
hydrological modelling systems include the reduced rank square root filter and the ensemble Kalman filter (e.g. Evensen, 1994; Verlaan & Heemink, 1997; Madsen & Cañizares, 1999).

In this paper a DA procedure implemented in the MIKE 11 FF system based on the ensemble Kalman filter is presented. Results are shown from an application of the procedure in a simulated real-time forecasting test of the Sesia catchment in the Piedmont region in the northwestern part of Italy.

KALMAN FILTER DATA ASSIMILATION PROCEDURE

The basis for the Kalman filter is a state-space formulation of the MIKE 11 model:

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

where \( \Phi(.) \) is the model operator representing the numerical scheme used to solve the governing equations, \( x_k \) is the state vector at time step \( k \) representing water levels and fluxes in the numerical grid, and \( u_k \) is the forcing of the system representing all boundary conditions. The model equation in (1) is given a stochastic interpretation:

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

where \( \epsilon_k \) is a stochastic element representing the uncertainty of the modelled system. In this case the uncertainties are assumed to originate from errors in the boundary conditions.

It is assumed that measurements of water levels or fluxes are available at different locations in the river system. This is formulated in the measurement equation:

\[ z_k = C_k x_k + \eta_k \]  

where \( C_k \) is a matrix that describes the relation between measurements and state variables (i.e. a mapping of state space to measurement space), and \( \eta_k \) is a random noise, representing the observation errors. Observation errors arise from a number of sources, including the measurement equipment, the sampling procedure, and the interpretation of sensor measurements as state variables (i.e. the use of point measurements to represent grid averages in the numerical model).

An updated (or analysed) state \( x^u_k \) is obtained by a linear combination of the model forecast \( x^f_k \) cf. equation (1) and the measurements, equation (3):

\[ x^u_k = x^f_k + K_k (z_k - C_k x^f_k) \]  

where \( K_k \) is the Kalman gain:

\[ K_k = P_k^f C_k^T \left[ C_k P_k^f C_k^T + R_k \right]^{-1} \]  

In equation (5), \( P_k^f \) is the covariance of the model forecast, and \( R_k \) is the covariance of the measurements. The updated covariance reads:

\[ P_k^u = P_k^f - K_k C_k P_k^f \]
In the ensemble KF, the statistical properties of the state vector are represented by an ensemble of possible state vectors. Each of these vectors is propagated according to the model operator accounting for model errors, equation (2), and the resulting ensemble then provides estimates of the forecast state vector and the corresponding covariance. The algorithm can be summarized as follows (Madsen & Cañizares, 1999):

(a) Each member of the ensemble of $M$ state vectors is propagated forward in time according to the dynamics of the system and the specified model error, i.e.:

$$x^f_{i,k} = \Phi(x^u_{i,k-1}, u_k + \varepsilon_{i,k}) \quad , \quad i = 1,2,..., M$$

where the model error $\varepsilon_{i,k}$ is randomly drawn from a predefined error model. In this case a first order autoregressive, AR(1), model is adopted.

(b) The forecast of the state vector is calculated as the mean value of the ensemble, i.e.:

$$\bar{x}^f_k = \frac{1}{M} \sum_{i=1}^{M} x^f_{i,k}$$

The covariance matrix of the forecast can be estimated from the ensemble as:

$$P^f_k = S^f_k (S^f_k)^T, \quad s^f_{i,k} = \frac{1}{\sqrt{M-1}} (x^f_{i,k} - \bar{x}^f_k)$$

where $s^f_{i,k}$ is the $i$-th column in $S^f_k$.

(c) An ensemble of size $M$ of possible measurements is generated:

$$z_{i,k} = z_k + \eta_{i,k}, \quad i = 1,2,..., M$$

where $z_k$ is the actual measurement vector, and $\eta_{i,k}$ is the measurement error that is randomly generated from a normal distribution with zero mean and predefined covariance $R_k$.

(d) Each ensemble member is updated using equation (4), and based on the updated ensemble the updated state vector and corresponding covariance are estimated using equations (8)-(9). In the implementation $P^f_k$ is not calculated, and all operations are done using $S^f_k$ directly.

**CASE STUDY**

**Model setup**

The Kalman filter data assimilation procedure has been tested using the flood forecasting system that has been developed for the Piedmont region in the north-western part of Italy. The system covers the Upper Po River basin, an area of approximately 37,000 km². Detailed MIKE 11 rainfall–runoff and river hydrodynamic models have been setup and connected to MIKE FloodWatch, a generic real-time flood forecasting system. The system is linked to a telemetric system that provides measured data from more than 200 meteorological stations (precipitation and temperature) and about 70 water level gauging stations. In addition, quantitative precipitation and temperature forecasts are obtained from a local area meteorological model. The present operational system includes a data assimilation system based on an error correction procedure (Rungø et al., 1989). For more information about the system see Barbero et al. (2001).
For testing the Kalman filter a sub-basin was selected, the Sesia River basin, which covers an area of about 3450 km$^2$ (Fig. 1). The basin ranges in altitude from about 90 m at the outlet into the Po River to about 4500 m in the upper alpine part of the basin. The MIKE 11 setup includes the Sesia River and the three main tributaries: Sessera, Cervo and Elvo. The basin includes 14 sub-catchments for modelling the rainfall–runoff process, which provide lateral inflow to the MIKE 11 model. Using the available precipitation measurements, catchment average hourly precipitation is estimated for the 14 catchments. Similarly, temperature data for the snow module are estimated from the available temperature measurements and referenced to an altitude of 1000 m. An altitude dependent snow model is applied to individual elevation zones in the catchment by using a temperature lapse rate. For evapotranspiration calculations, monthly potential evapotranspiration data have been compiled for different elevations.

**Kalman filter setup**

For the data assimilation three water-level measurement stations are available (Fig. 1). Measurement errors are described by a normal distribution with zero mean and a standard deviation of 0.1 m. To specify the uncertainties in the rainfall–runoff boundaries, a Monte Carlo simulation study was carried out to evaluate the uncertainty in basin runoff as a function of uncertainty in input data (precipitation and temperature) and model parameters. In general, the response to input and parameter uncertainty is more pronounced for mountainous catchments than lowland catchments. The relative uncertainty in runoff from high-, middle- and low-latitude catchments were set to, respectively: 35%, 30% and 25%. In addition, coloured noise was assumed for all catchments using a time constant of 12 h for the AR(1) process.

When coloured noise is used, the boundary error terms are updated along with the update of the water levels and fluxes in the river system. These errors are propagated in time according to the defined AR(1) model. In forecast simulation, the boundary errors at the time of forecast are then reduced according to an exponential decay with the specified time constant. Thus, the filter also allows correction of the rainfall–runoff boundaries in the forecast period.

An important parameter of the ensemble KF is the ensemble size $M$. To reduce the computational time required for the filter one should keep the ensemble size as small as possible (the required computational time corresponds to $M$ times a normal model run). On the other hand, the larger the ensemble size, the better is the approximation of the covariance modelling. Preliminary test runs for the Sesia catchment showed that an ensemble size of $M = 50$ gave a reasonable compromise between the computational costs and the accuracy of the covariance approximation.

**RESULTS**

The MIKE 11 FF system has been applied for simulated real-time forecasting of a high flow event that occurred in the Sesia catchment during the end of April and beginning of May 2000. For this event, forecasts with a lead-time of 48 h were performed every 6 h. Observed water level data were assimilated at the three gauging stations up to the time of forecast. In the forecast period the ensemble was propagated to provide a
Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering

Fig. 1 Sesia flood forecast model. The 14 rainfall-runoff sub-catchments are numbered 1001–1013 and 8008. The setup includes three water level gauging stations: 188, 200 and 413.

Fig. 2 Water level forecasts at station 413 for two different times of forecast (TOF) compared to observations.

Forecast of water levels and fluxes in the river system (ensemble average) and the associated confidence limits (ensemble quantiles).

Examples of water level forecasts are shown in Fig. 2. For this event the rainfall is underestimated, which results in an underprediction of the water level in the river system when data assimilation is not applied. When water level measurements at the
three stations are assimilated, the initial conditions at the time of forecast are updated and the forecasts are significantly improved, although an underprediction is still observed. The quality of the forecast is explicitly quantified by the predicted confidence limits.

To evaluate the performance of the DA procedure the root mean square error (RMSE) between the forecast and observed water levels at the three gauging stations have been calculated for different lead times. For increasing lead time, the forecast with updating will move towards the forecast without updating, i.e. the forecast RMSE will tend to the RMSE of the forecast water levels when no data assimilation is applied. In Fig. 3 the relative reduction in forecast RMSE compared to the RMSE without updating is shown for the three stations. As expected, for short lead times the forecast RMSE is significantly reduced. Significant RMSE reductions are also observed for longer lead times, especially at station 413 (upstream Sesia) and station 200 (downstream Sesia). At station 413 the RMSE is reduced by up to 30–35%, even for forecasts with a 48-hour lead time.

CONCLUSIONS

A DA procedure has been presented for assimilation of observed water levels and fluxes in the MIKE 11 FF system. The DA procedure is based on the ensemble Kalman filter that allows a cost-effective and efficient approximation of the covariance
modelling that can be used for real-time applications. In the forecast period, the ensemble forecast provides an estimate of the prediction uncertainty consistent with the assumed uncertainties in the boundary conditions. Application of the updating procedure for simulated real-time forecasting in the Sesia catchment showed significantly improved forecast skills for lead times of up to 48 hours as compared to forecasts without updating.

Acknowledgements This work was in part funded by the Danish Technical Research Council under the Talent Project no. 9901671: Data Assimilation in Hydrodynamic and Hydrological Modelling. Data for the case study was provided by Regione Piemonte.

REFERENCES


