DATA ASSIMILATION TO IMPROVE FORECAST QUALITY OF RIVER BASIN MODELS

ANNE KATRINE V. FALK, MICHAEL B. BUTTS, HENRIK MADSEN, JOHAN N. HARTNACK[1]

[1] *DHI – Water & Environment, Agern Allé 5, DK-2970 Hørsholm, Denmark*

ABSTRACT

This paper evaluates alternative data assimilation formulations based on the Ensemble Kalman Filter for flood forecasting. A general framework is developed for a combined river and catchment model to perform uncertainty propagation and data assimilation for hydrological forecasting and simulation. This Ensemble Kalman Filter framework allows inclusion of uncertainty in the observations, parameters and boundary conditions. Data assimilation improves forecast accuracy by using observations up to the time of forecast thereby providing the best initial conditions for a forecast. In this study two implementations of data assimilation using measured river flows are investigated. In the first only states in the hydraulic (river) part of the model are updated assuming that uncertainty arises directly from uncertainty in the catchment inflows from the rainfall-runoff model. In the second, updating is carried out on states in both the hydraulic (river) and hydrological (catchment) parts of the model, assuming uncertainty arises from the catchment rainfall. The results show that for volumetric errors, data assimilation in both the river and catchment states provides more accurate forecasts over longer lead times than updating on the river channel alone. However, in the case of phase errors the performance of the two methods are comparable.

1. INTRODUCTION

Ideally, real-time flood management decisions must be based on an understanding of the uncertainties and associated risks. It is therefore central for effective flood forecasting tools to provide reliable estimates of the forecast uncertainty [*Butts et al.*, 2004]. Only by quantifying the inherent uncertainties involved in flood forecasting, can effective real-time flood management and warning, be carried out [*Cadman et al.,* 2006]. Estimating forecast uncertainty requires the estimation of the uncertainties associated with the hydrological model inputs (e.g. observations or forecasts of precipitation), model structure, parameterisation and calibration, and methodologies that predict how the uncertainties from different sources propagate through the hydrological and hydraulic system. During a flood both lives and property are at risk and therefore any means to reduce forecast uncertainty are highly desirable. WMO has identified updating or data assimilation as an essential requirement for accurate flood forecasting [*WMO*, 1992].

Within the EU 5^{th} framework project FLOODRELIEF, an ensemble-based approach has been developed to estimate forecasting uncertainties and reduce uncertainty using data assimilation. This general stochastic framework for flood forecast modelling is based on the Ensemble Kalman Filter. The Ensemble Kalman filter was introduced by [*Evensen*, 1994] and provides a natural framework for determining how the different sources of uncertainty propagate through the hydrological and hydraulic models and to reduce forecast uncertainty via data assimilation of real-time observations. Because of its improved ability to treat non-linearity and model-independent nature, this filter is widely used in oceanography and meteorology and more recently hydrology, [*Hartnack and Madsen*, 2001; *Madsen et al.,* 2003; *Butts et al*., 2005]. This paper presents some of our investigations using this framework for real-time updating or data assimilation using measurements of river flow. An evaluation of two alternative data assimilation methods is carried out using two case studies with different hydrological regimes, the US National Weather Service study catchment (the Blue river basin), and the Waipa basin, a subbasin of the Waikato River in New Zealand.

2. DATA ASSIMILATION FRAMEWORK

While not as widely applied as one might expect, data assimilation when applied is often carried out using measurements of river flows (or water levels), correcting the simulated flows or modifying the river flows within the model. This will improve forecasts over a period corresponding to the travel time of a flood wave through the river. However the uncertainty can be traced more directly back to the uncertainty in the meteorological boundary conditions. Indeed several authors identify uncertainty in the estimated and forecasted rainfall as the main source of uncertainty in flood forecasting. Uncertainty in the rainfall generates uncertainties both directly in the runoff generated and indirectly by determining the catchment wetness prior to rainfall.

Observations of river flow contain information concerning both the current conditions in the river and also of the subcatchment conditions upstream of the measurements. To properly exploit this information, data assimilation is carried out using a combined river and catchment model. The model used for flood forecasting is the river basin model MIKE 11 [*Havnø et al.*, 1995], which includes a hydrological (rainfall-runoff) component (MIKE 11 RR) which uses subcatchments to represent, in a semi-distributed fashion, run-off into the (hydrodynamic) river component, (MIKE 11 HD). Within MIKE 11 RR, the NAM model, a lumped conceptual rainfall-runoff model, is used to represent the rainfall-runoff process in each subcatchment, [*Havnø et al.*, 1995; *Madsen*, 2000]. The MIKE 11 HD component routes the flood wave through the river system by performing a numerical integration of the Saint-Venant equations based on knowledge of the topography and the boundary conditions, [*Havnø et al.*, 1995; *DHI*, 2005].

Formulating this deterministic model as an operator in a stochastic setting we can write

$$x(t_i) = \begin{pmatrix} x_M(t_i) \\ x_\eta(t_i) \end{pmatrix} = \begin{pmatrix} M_M(x_M(t_{i-1}), u(t_{i-1}), x_\eta(t_{i-1})) & +\eta_{M,i-1} \\ M_\eta(x_\eta(t_{i-1})) & +\eta_{\eta,i-1} \end{pmatrix}$$

where

$x_M(t_i)$ is the model state vector consisting of e.g. water levels and velocities etc. in a discretized mesh

M_M is the model operator which represents the numerical scheme to solve the governing equations

$u(t_{i-1})$ is the forcing of the system. The term forcing is here used broadly for both boundaries (which usually vary with time during a simulation) and parameters (model variables that are usually not changed during the simulation)

$\eta_{M,i-1}$ is white noise perturbations of the model state

$x_\eta(t_i)$ is the forcing error model state (this is the augmented part of the total state vector $x(t_i)$)

M_η is the model operator for the forcing error model

$\eta_{\eta,i-1}$ is a white noise perturbations of the forcing error model state

The sequential estimation process in data assimilation occurs in two steps. First, the dynamical model $M = \begin{pmatrix} M_M & M_\eta \end{pmatrix}^T$ is employed to issue a forecast, and then a statistical criterion is used to meld the observed data with the dynamical forecast (measurement updating). The Ensemble Kalman filter is used here for the measurement updating.

In this study we will consider model uncertainty arising from the boundary conditions, whereas the uncertainty stemming from the choice of parameters has been minimized during a calibration process. The model structure errors are not accounted for directly. Two methods for introducing uncertainties into the model are compared:

1. Data assimilation or updating is performed only in the river channel assuming stochastic errors (coloured noise) in the runoff calculated by the catchment rainfall-runoff model. This is denoted HD-KF assimilation.
2. Data assimilation or updating is performed in both the subcatchments and the river channel assuming stochastic errors (white noise) in the meteorological input in the subcatchment model (in the present work only the precipitation boundary is considered). This is denoted HD+RR-KF assimilation.

In the HD-KF approach, the rainfall-runoff model runs deterministically and the calculated subcatchment runoff serves as boundary condition for the river (HD) model. That means that the model state only consists of state variables from the river model and the boundaries that are perturbed are the runoff provided by the subcatchment models. The perturbations of these boundary conditions (subcatchment runoff) are modelled as an autoregressive process (coloured noise) in order to represent memory in the perturbations. The state variables of the HD model are water level and discharge in all computational points.

In the HD+RR-KF approach, state variables from the subcatchment models are included in the state vector. These state variables are related to overland flow, interflow and baseflow within each subcatchment. In this case the meteorological (rainfall) boundaries to the subcatchment models are perturbed assuming they are uncorrelated in time (white noise). This implies that the augmented part of the state vector, x_η, is not included in the state space description. For both approaches the standard deviation of the perturbations is assumed to be proportional to the boundary value. For the HD+RR-KF method one drawback with this formulation is that uncertainty is present only while it is raining. When the rain stops the variation of the runoff ensemble decreases to zero with the time constant inherited from the subcatchment model. This is addressed by adding white noise perturbations to the RR state variables, i.e. a minor part of the elements in the vector $\eta_{M,i-1}$ are non-zero.

In both cases discharge measured in the river channel is used for updating. Discharge is a state variable in the river model whereas none of the RR state variables are measurable quantities. The advantage of putting the river state variables and the subcatchment state variables into one joint state vector is that it is then possible to update the subcatchment state variables even though only measurements of river state variables are available. The Kalman filter calculates the correlations between the two model spaces and the river measurements contain information on both systems.

3. CASE STUDIES

3.1 Evaluation method

To evaluate the effect of data assimilation on the accuracy of flood forecasts a number of flood peaks were selected. For each a sequence of forecasts is generated and measured flows are assimilated into the simulation model until time of forecast. The discrepancy from the measured discharge is quantified by the Root Mean Squared Error (RMSE) as a function of the forecast lead time. The lower the RMSE, the closer the group of sequential forecasts is (on average) to the observation. The calibration simulation is used as reference.

Two case studies were used to evaluate the alternative data assimilation methods. The first, the Blue River basin in USA is a semi-arid, quickly responding catchment with little or no baseflow [*Butts et al.*, 2004]. The discharge measurement is located at the downstream boundary. The Waipa basin, a subcatchment of the Waikato River, New Zealand, has a longer response time with a significant baseflow compared to the Blue River. Discharge measurements are available at 5 locations, 3 upstream, one midway in the river system and one at outlet of the subcatchment [*van Kalken et al.*, 2005].

3.2 Results

In each case, to evaluate the improvement obtained by data assimilation, two flood events were selected where the model calibration exhibited substantial deviations from the observed flood. One with volumetric errors in the flood peak simulation and one with phase errors in timing of the peak. The results for the Waipa model are shown for forecasts at one of the upstream updating points (Barton's Corner). FIGURE 1 compares the observed hydrograph to the calibration simulation for these two events. The first event consists of two successive peaks and especially for the simulation of the second peak; the predicted rise in the hydrograph is too early. The second event has a measured discharge that is of the same order of magnitude as the first, but the simulation grossly overestimates the peak value whereas the timing of the peak is correct.

FIGURE 2 shows two illustrative examples of forecasts during these two events, conducted after assimilating data with the HD-KF method (updating in river channel) and the HD+RR-KF method (updating both in catchment and river channel). The improvement obtained by updating in both the catchment and the river, as opposed to only updating in the river channel, for the event on the right with volumetric errors is clear. The HD+RR-KF assimilation successfully removes a part of the excess water, both on the rising limb and on the recession. For the event on the left with phase errors, the peak rise is also improved the most by HD+RR-KF approach for the initial peak. However in attempting to reproduce the recession prior to the main peak the HD+RR-KF approach reduces the initial wetness of the subcatchments resulting in a poor forecast, at this time, of this larger peak. This is corrected in the subsequent forecasts but illustrates one of the drawbacks of this method. By contrast the HD-KF method captures this second main peak well.

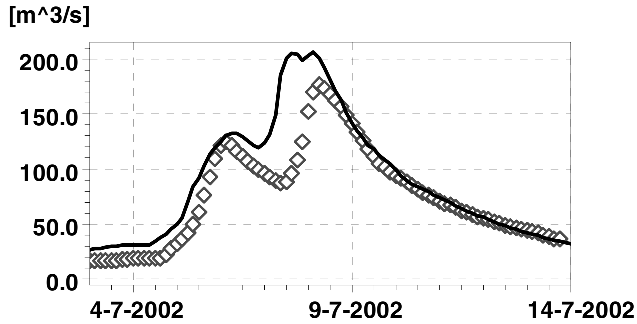

FIGURE 1. Observed hydrograph (◊) compared to that simulated by the calibration simulation (black line) for two peak events in the Waipa model

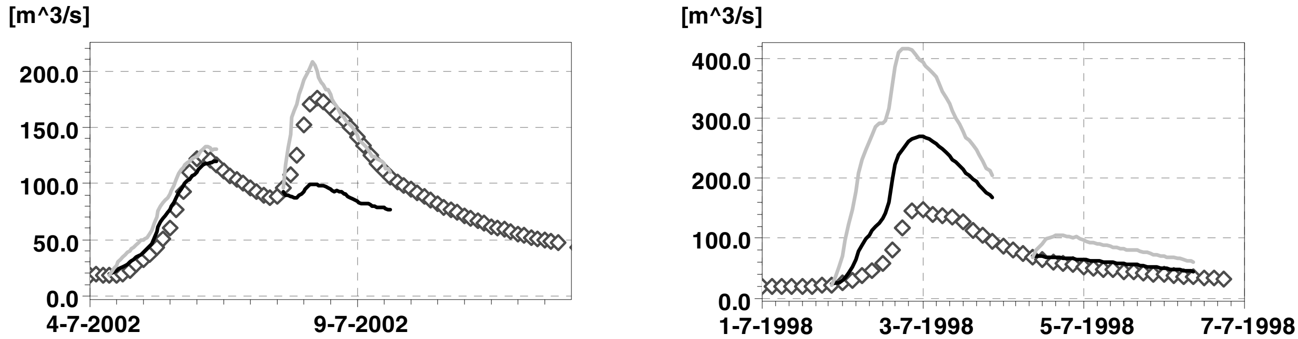

FIGURE 2. Forecasts initiated after assimilating data by the two methods of assimilation: HD-KF (grey lines), HD+RR-KF (black lines) and measurements (◊)

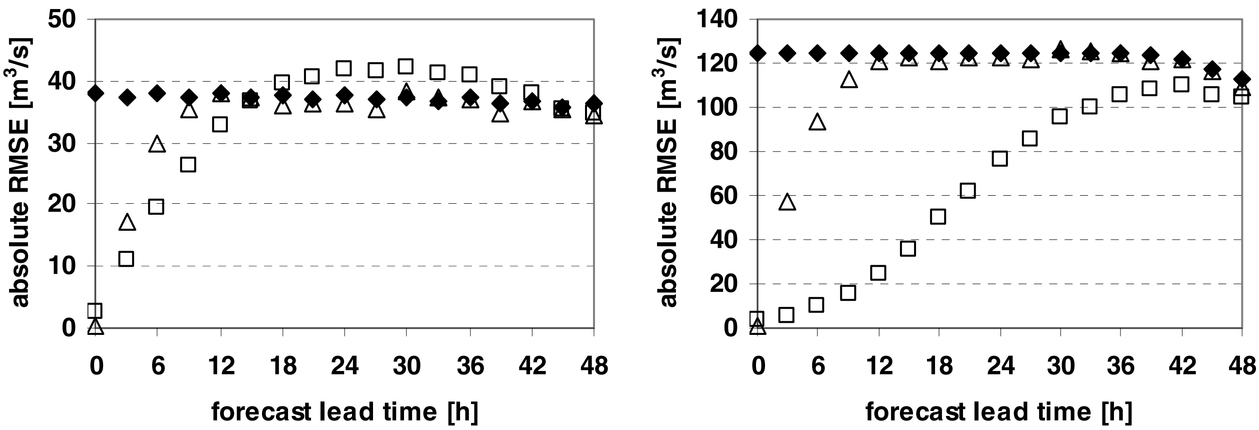

FIGURE 3. RMSE for the two Waipa events as function of forecast lead time. Calibration (♦), HD-KF(Δ) and HD+RR-KF (□)

FIGURE 3 shows the RMSE for the sequence of forecasts as a function of lead time. The RMSE for the two assimilation methods is compared to that obtained using the calibration as forecast. If the RMSE is lower than that produced by the calibration the assimilation has improved the forecasts (on average) over the event, if the RMSE is higher one would have been better off using the calibrated model without assimilation to forecast. For both events HD-KF improves the forecasts until lead times about 12 hours. For longer lead times, no benefit from data assimilation is obtained. The physical explanation is that modification to the river states will be maintained over a period corresponding to the travel time in the river, in this case about 12 hours. It is expected that by modifying the catchment states as well this period is extended by the time of concentration (time from rainfall to generation of runoff) within each subcatchment. This is dramatically illustrated in the graph on the right in FIGURE 3. Updating within both the catchment and river (HD+RR-KF) improves forecast accuracy generally and does so for lead times up to 42 hours ahead. For the event on the left where phase errors are present, this approach improves the forecast for lead times up to 12 hours, also when compared to the HD-KF method. However for larger lead times the forecast accuracy becomes poorer, apparently because the phase errors are not correctly captured. This behaviour is maintained out to lead times of 42 hours.

A similar analysis has been carried out for the Blue River basin, see FIGURE 4. The left hand event peaks too early, the right hand event underestimates the volume. Again updating both the river and subcatchment states appears to be significantly better at correcting for volumetric errors, see FIGURE 5. This is confirmed in FIGURE 6 where the HD+RR-KF improves the forecast accuracy for lead times up to 18 hours compared to around 12 hours for the HD-KF approach. For phase errors, the two methods exhibit comparable accuracy for lead times of up to 9 hours, with HD+RR-KF performing slightly better. For lead times between 9 and 18 hours the HD+RR-KF method appears to give poorer forecasts. It is not clear how important this reduction in performance is as it represents only a small fraction of the observed forecasting errors.

4. CONCLUSIONS

A stochastic framework for uncertainty estimation and data assimilation has been developed for a combined river and catchment model. River discharge measurements are used for data assimilation in order to exploit the information concerning both the river and catchment conditions embedded in these measurements. An evaluation of flood events has been carried out to evaluate the performance of two data assimilation methods where significant forecast errors are present. Two type of errors, volumetric and phase errors are considered. The results from two different catchments indicate that important improvements in forecast accuracy can be obtained where volumetric errors are present in the forecast. Data assimilation in both the river and catchment states provides more accurate forecasts over longer lead times than updating on the river channel alone. Where phase errors occur, these methods perform satisfactorily. Compared to updating only in the river system the performance is slightly better for lead times corresponding to the travel time in the river and slightly poorer beyond this time for a period corresponding roughly to the time of concentration in the subcatchments. Further work is being carried out to investigate this behaviour for other flood events and catchments.

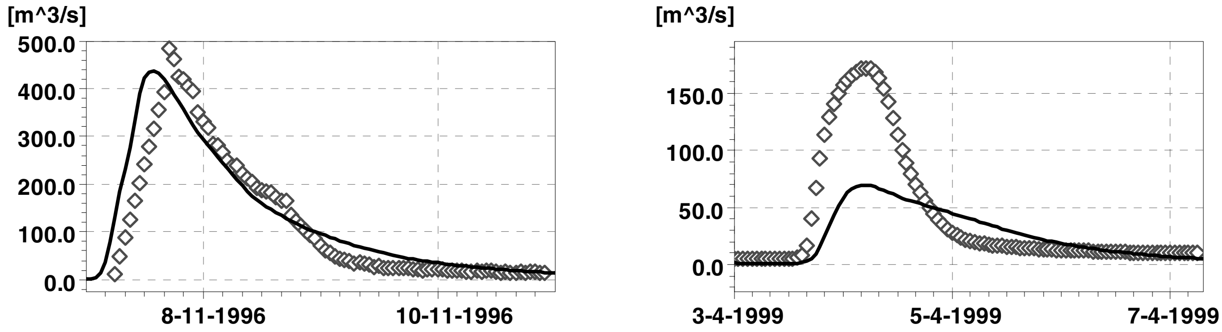

FIGURE 4. Observed hydrograph (◊) compared to that simulated by the calibration simulation (black line) for two peak events in the Blue River model.

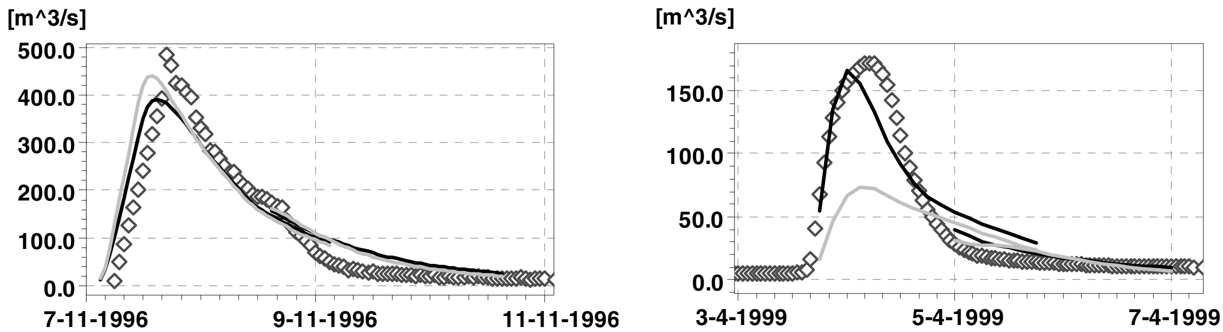

FIGURE 5. Forecasts initiated after assimilating data by the two methods of assimilation: HD-KF (grey lines), HD+RR-KF (black lines) and measurements (◊)

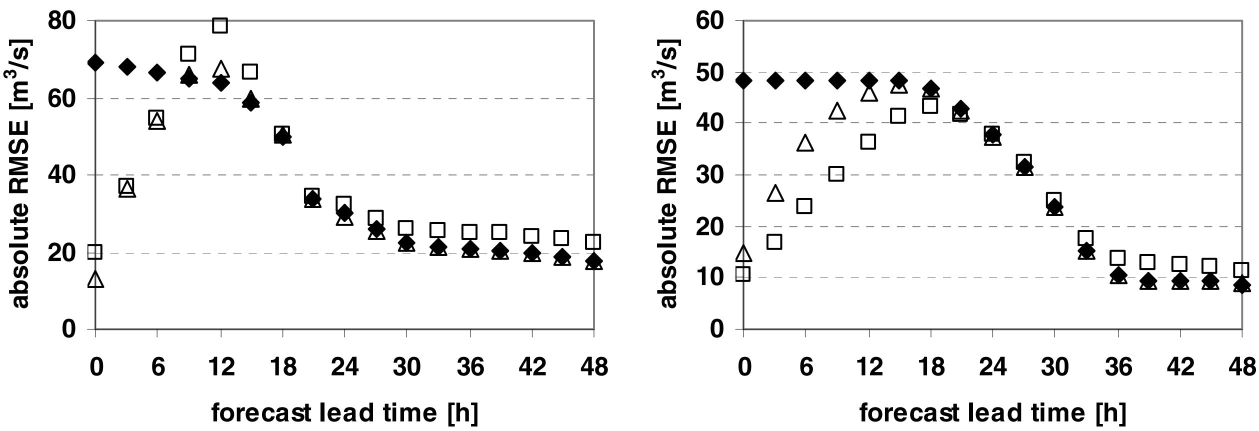

FIGURE 6. RMSE for the two Blue River events as function of forecast lead time. Calibration (♦), HD-KF(Δ) and HD+RR-KF (□)

ACKNOWLEDGEMENTS

The authors would like to acknowledge the DMIP project and the Hydrology Lab, OHD, NWS for the Blue River basin data. Environment Waikato provided the data and Terry van Kalken, DHI Water & Environment New Zealand provided the model for the Waipa basin. This study was carried out with the support of the EU 5th Framework Research Programme, FLOODRELIEF, contract EVK1-CT2002-00171, http:// projects.dhi.dk/floodrelief/.

REFERENCES

Butts, M. B., J. T. Payne, M. Kristensen and H. Madsen (2004). An evaluation of the impact of model structure and complexity on hydrological modelling uncertainty for streamflow prediction. *Journal of Hydrology, 298*, 242-266.

Butts, M. B., A. K. Falk, J. Hartnack, H. Madsen, A. Klinting, T van Kalken, D. Cadman, and D. Price (2005), Ensemble-based methods for data assimilation and uncertainty estimation in the FLOODRELIEF project. *Proc. International Conference "Innovation, advances and implementation of flood forecasting technology"*, 17-19 October 2005, Tromsø, Norway.

Cadman, D., D. Price and M. B. Butts (2006), Flood Forecasting in the Anglian Region: User-driven Development towards Forecasting Flood Risk. *Flood Risk Management in Europe: Innovation in Policy and Practice Series: Advances in Natural and Technological Hazards Research*, Vol. 25, edited by S. Begum, M. J. F Stive, J. W. Hall, ISBN: 1-4020-4199-3.

DHI (2005), MIKE 11 - A modelling system for rivers and channels, *Reference Manual* , *http://www.dhigroup.com/Software/WaterResources/MIKE11.aspx*

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res., 99*, 10143-10162.

Havnø, K., M. N. Madsen and J. Dørge (1995), MIKE 11 – A Generalized River Modelling Package, In: *Computer models of Watershed Hydrology*, edited by V. P. Singh, pp. 733-782, Water Resources Publications, Colorado, USA.

Hartnack, J and H. Madsen. (2001), Data assimilation in river flow modelling, *Proceedings of the 4th DHI Software Conference*, Helsingør, Denmark, 2001.

Madsen, H. (2000*), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives*, *Journal of Hydrology, 235(3-4),* 276-288.

Madsen, H., D. Rosbjerg, J. Damgård and F. S. Hansen (2003), Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering, in *Water Resources Systems - Hydrological Risk, Management and Development* (Proceedings of symposium HS02b held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 281, 2003. pp. 75–81.

Van Kalken, T., C. Skotner, and M. Mulholland (2005), Application of a Novel Real-Time Flood Forecast System to the Waikato River, New Zealand, *Proc. International Conference "Innovation, advances and implementation of flood forecasting technology",* 17-19 October 2005, Tromsø, Norway

WMO 1992. Simulated Real-Time Intercomparison of Hydrological Models. OHR-38. WMO n. 779.