

Real-time Flood Forecasting Using Ensemble Kalman Filter

Srikanthan, R.^{1,2}, G. Amirthanathan^{1,2} and G. Kuczera^{2,3}

¹ Hydrology Unit, The Bureau of Meteorology, Melbourne

² eWater Cooperative Research Centre for Catchment Hydrology, Canberra

³ School of Engineering, The University of Newcastle, Newcastle

Email: r.srikanthan@bom.gov.au

Keywords: Real-time flood forecasting, ensemble Kalman filter, comparison

EXTENDED ABSTRACT

There is a growing interest in knowing the uncertainty in flood forecasting and the resulting flood warnings. This is borne out of the fact that the processes involved in flood forecasting have inherent uncertainties in them. The procedure used in flood forecasting consists of a number of steps. The first step is rainfall measurement and forecasting rainfall during a flood event. The rainfall is then transformed into flow using a combined water balance and runoff-routing model. There are uncertainties associated with rainfall measurement/forecasting, model and flow measurements.

The Ensemble Kalman Filter (EnKF) and its derivatives are used widely in real time flow forecasting and it was decided to evaluate the performance of EnKF for Australian data. In an earlier study, three variations of EnKF were applied to the May 2003 flood event in Georges River in Sydney, Australia. In this study, four variations of EnKF were applied to four flood events in Gudgenby River in Canberra, Australia.

The probability distributed moisture (PDM) model was used to transform the rainfall to discharge. The resulting discharge from the PDM model was updated using the EnKF. The four variations of EnKF considered in this study were the state updating, parameter updating, dual (state-parameter) and dual (parameter-state) updating. The performance of EnKF was evaluated using the root mean square and the coefficient of efficiency for 1, 3, 6, 9 and 12-h lead time forecasts. Also, the error in peak magnitude and peak timing error were also used in the comparison.

Of the four variations considered, the parameter updating performed the best in terms of RMSE, coefficient of efficiency and peak error. The 1, 3 and 6-h lead time forecasts are shown in Figure 1 for parameter updating.

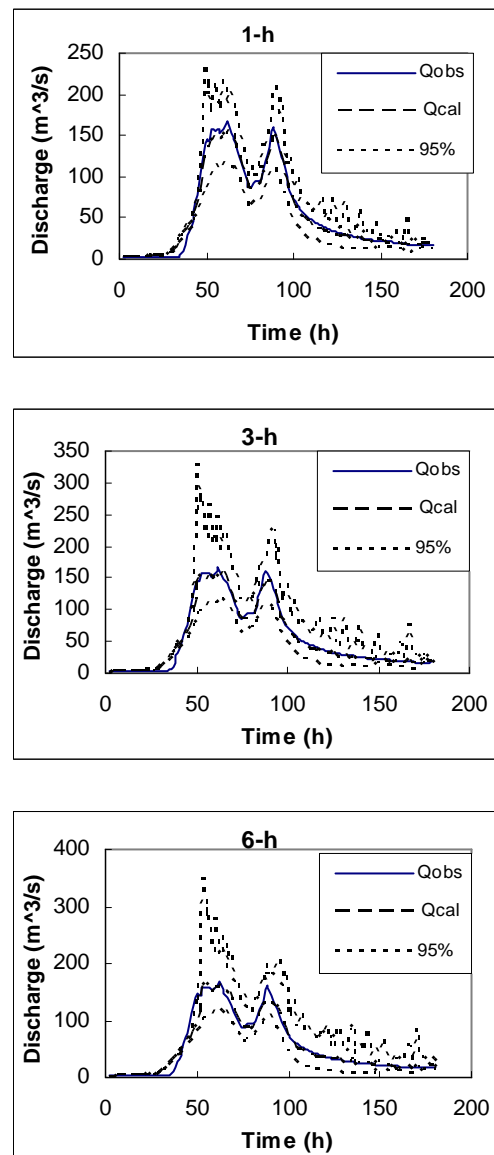


Figure 1. Comparison of flood forecast for different lead times.

1. INTRODUCTION

The Australian Government Bureau of Meteorology has a national responsibility for the provision of flood forecasting and warning services to its citizens. In Australia, the application of more recent advances in hydrologic modelling and data assimilation techniques to real-time flood forecasting has only commenced relatively recently. Through involvement in the Cooperative Research Centre (CRC) for Catchment Hydrology, the Bureau has completed a research project in which a number of different models and updating techniques were compared using data from 14 catchments (Srikanthan et al. 1997). Although individual models performed better on different catchments, a simple ranking procedure used to obtain an overall comparison showed that the Probability Distributed Moisture (PDM) model was one of the better performing model for a large number of catchments (Srikanthan et al. 1997). The uncertainty in the flood forecasting process was not considered in the above study.

Nowadays there is a growing interest in knowing the uncertainty in flood forecasting and the resulting flood warnings. This is borne out of the fact that the processes involved in flood forecasting have inherent uncertainties in them. The procedure used in flood forecasting consists of a number of steps. The first step is rainfall measurement and forecasting rainfall during a flood event. The rainfall is then transformed into flow using a combined water balance and runoff-routing model. There are uncertainties associated with rainfall measurement/forecasting, model and flow measurements. All these uncertainties contribute to the uncertainty in the resulting flood forecasts.

The interest in quantifying uncertainty in real-time flood forecasting has led to a number of publications in the treatment of uncertainty caused by model parameters. Sequential data assimilation techniques provide a means of explicitly taking account of input, model and output uncertainties. One of the earliest data assimilation techniques is the Kalman filter developed for linear systems. For use with nonlinear models, it was later extended resulting in the extended Kalman filter (EKF). These two filters have been widely used in hydrologic modelling. If the nonlinearities in the model are strong, the linearization becomes very inaccurate. This has led to the development of the ensemble Kalman filter (EnKF) where the errors are allowed to evolve with the nonlinear model equations by performing an ensemble of model runs (Burgess et al., 1998).

El Serafy and Mynett (2004) evaluated the feasibility of applying EnKF to real-time flood forecasting applications by comparing it with EKF for the Sobek River in Netherlands. The comparison showed that the EnKF gave similar results to those of the already operating EKF model with 10 or more ensemble members and they recommended the use of EnKF to other models such as Sobek Rural/Urban and Delft 3D. Weerts and El Serafy (2005) applied the EnKF and residual resampling (RR) to the HBV-96 rainfall runoff model with both synthetic and real data. The EnKF and RR algorithm performed comparably well. A number of assumptions were made on the model errors. The authors concluded that the effect of these assumptions should be investigated and quantified by systematic sensitivity analysis. Moradkhani et al. (2005) investigated the applicability and usefulness of dual state-parameter estimation of hydrologic models using ensemble Kalman filter and found that the one-day ahead forecast was consistent with the observations for the Leaf River. Weerts et al. (2006) compared sequential importance sampling (SIR), RR and EnKF with the HBV-96 rainfall runoff model for flood forecasting using synthetic and real data. The results from the real data showed that both the SIR and RR were more sensitive to the choice of model and measurement errors. This made the EnKF more robust and outperformed the other two filters. In this study, EnKF and its several variations are used with the Probability Distributed Moisture model to forecast four flood events in Gudgenby River and evaluated.

2. PROBABILITY DISTRIBUTED MOISTURE MODEL

The Probability Distributed Moisture (PDM) model is a conceptual rainfall-runoff model which transforms rainfall and evaporation data to flow at the catchment outlet (Moore, 2007). Figure 2 shows the general form of the model. The runoff production at a point in the catchment is controlled by the absorption capacity of the soil (treated together with canopy interception and surface detention) to take up water. This is conceptualised as a simple store with a given storage capacity. By considering that different points in a catchment have differing storage capacities and that spatial variation of capacity can be described by a probability distribution, it is possible to formulate a simple runoff production model which integrates the point runoffs to yield the catchment surface runoff into surface storage (S_2). The recharge from the soil moisture store (S_1) passes into subsurface storage (S_3). The outflow from surface (q_s) and subsurface (q_b) storages forms the model output

(q). A complete description of PDM model is given in Moore (2007).

variations that are used in this study are described below.

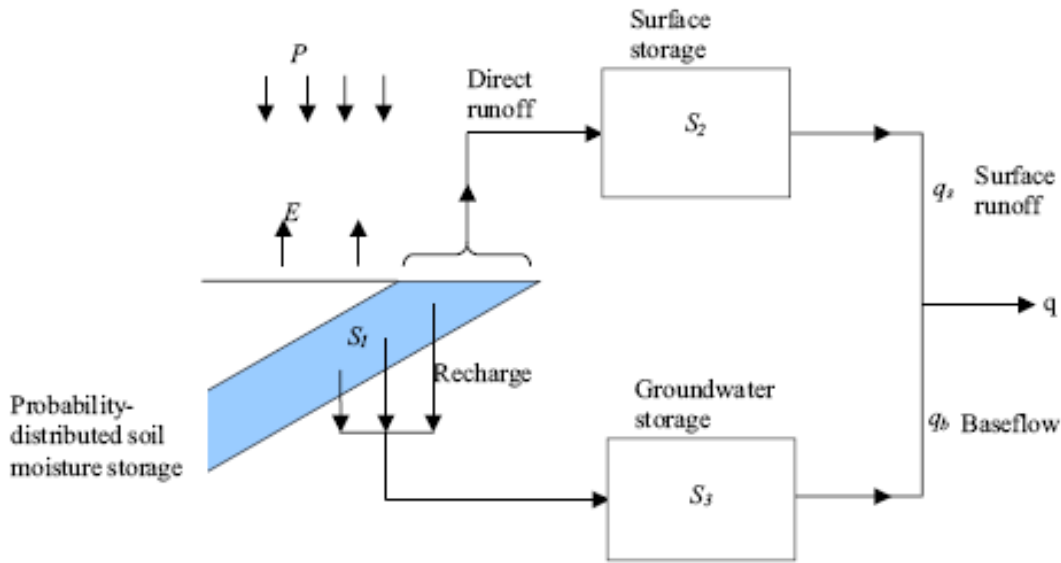


Figure 2. The PDM rainfall-runoff model (Moore, 2007)

In the PDM model formulation, the surface runoff is calculated from the previous values of the surface runoff and net rainfall. In order to update the model states using ensemble Kalman filtering, the stores in the cascade of two linear reservoirs (S_{21} and S_{22}) of the surface storage (S_2) are expressed explicitly.

3. ENSEMBLE KALMAN FILTER AND ITS VARIATIONS

The Ensemble Kalman filter (EnKF) is a suboptimal estimator, where the error statistics are predicted by using Monte Carlo integration methods. The starting point is choosing an ensemble of state estimates that captures the initial probability distribution of state estimates. These sample points are then propagated through the nonlinear system and the probability density function of the actual state is approximated by the ensemble of the estimates. The approximation of forecast state error covariance matrix is made by propagating the ensemble of model states using the updated state from the previous time step. It is necessary to generate the ensemble of observations at each update time by introducing noise drawn from a distribution with zero mean and covariance equal to the observational error covariance matrix; otherwise the updated ensemble will have a very low covariance. In this study, error variances of input rainfall and output discharge are specified a priori and not updated. The EnKF and its four

3.1. EnKF with State Updating

An ensemble of state vectors is propagated through the PDM model such that each state vector represents one realisation of generated model state. Then the state forecast is made for each ensemble member as follows:

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta) + \omega_t^i, \quad i = 1, \dots, n. \quad (1)$$

where x_{t+1}^{i-} is the i^{th} ensemble forecasted state at time $t+1$, x_t^{i+} the i^{th} updated ensemble state at time t , θ the model parameters and u_t^i the input to the model. In addition to representing the additive process noise $\omega_t^i \sim N(0, \Sigma_t^s)$, the EnKF represents the multiplicative model errors through forcing data perturbations. The input data perturbations are made by adding the ζ_t^i noise with covariance Σ_t^u to the input data at each time step:

$$u_t^i = u_t + \zeta_t^i, \quad \zeta_t^i \sim N(0, \Sigma_t^u) \quad (2)$$

It was recognized that in order for the EnKF to maintain sufficient spread in the ensemble and prevent filter divergence, the observations should be treated as random variables by generating an observation ensemble with mean equal to the actual observation at each time and a predefined covariance (Burgers et al., 1998). Thus the

forecasted states x_{t+1}^{i-} are updated using Kalman gain K_{t+1}^x as follows:

$$x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^x (y_{t+1}^i - \hat{y}_{t+1}^i) \quad (3)$$

where \hat{y}_{t+1}^i is the i^{th} predictive variable at time $t+1$ given by:

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta), \quad (4)$$

y_{t+1}^i is the i^{th} replicate of observation generated by adding the noise η_{t+1}^i with zero mean and covariance Σ_{t+1}^y to the actual observation y_{t+1} as follows:

$$y_{t+1}^i = y_{t+1} + \eta_{t+1}^i, \quad \eta_{t+1}^i \sim N(0, \Sigma_{t+1}^y) \quad (5)$$

If the measurements are a nonlinear combination of state variables, the Kalman gain in adaptation to the ensemble based approach can be shown as (Moradkhani et al., 2005):

$$K_{t+1}^x = \Sigma_{t+1}^{xy-} [\Sigma_{t+1}^{yy} + \Sigma_{t+1}^y]^{-1} \quad (6)$$

where Σ_{t+1}^{yy} is the forecast error covariance matrix of the prediction \hat{y}_{t+1}^i , and Σ_{t+1}^{xy-} is the forecast cross covariance of the state variables x_{t+1}^{i-} and prediction \hat{y}_{t+1}^i and Σ_{t+1}^y is the covariance of the measurements.

3.2. EnKF with Parameter Updating

In this formulation, the parameters are considered as state variables, where the parameter evolution is represented by a random walk, given by

$$\theta_{t+1}^{i-} = \theta_t^{i+} + \tau_t^i, \quad \tau_t^i \sim N(0, \Sigma_t^\theta) \quad (7)$$

Using the above parameter ensemble and forcing data replicates given in (4), a model state ensemble and predictions are made, respectively:

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i-}) \quad (8)$$

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta_{t+1}^{i-}) \quad (9)$$

Using the Kalman filter, the updating of the parameter ensemble members is carried out:

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^\theta (y_{t+1}^i - \hat{y}_{t+1}^i) \quad (10)$$

where K_{t+1}^θ is the Kalman gain for correcting the parameter trajectories and is obtained by:

$$K_{t+1}^\theta = \Sigma_{t+1}^{\theta y-} [\Sigma_{t+1}^{yy} + \Sigma_{t+1}^y]^{-1} \quad (11)$$

where $\Sigma_{t+1}^{\theta y-}$ is the cross covariance of parameter ensemble θ_{t+1}^{i-} and prediction ensemble \hat{y}_{t+1}^i . Now using the updated parameter ensemble, the new model state trajectories are generated:

$$x_{t+1}^{i+} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i+}) \quad (12)$$

3.3. Dual EnKF

The dual EnKF requires two separate state-space representation for the state and parameters through two parallel filters. The parameters can be updated first and then the state variables or vice versa. These two variations (state-parameter and parameter-state) of updating in dual EnKF are used in this study. The procedure for state-parameter is as follows. The steps given in Eq (1) to (6) under state updating are applied initially to obtain state updates. Then using the updated state variables given in Eq (3), the new prediction trajectories are generated:

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i+}, \theta_t^{i+}) \quad (13)$$

Using the Kalman filter, the parameter ensemble is updated:

$$\theta_{t+1}^{i+} = \theta_t^{i+} + K_{t+1}^\theta (y_{t+1}^i - \hat{y}_{t+1}^i) \quad (14)$$

where K_{t+1}^θ is the Kalman gain for correcting the parameter trajectories and is obtained by:

$$K_{t+1}^\theta = \Sigma_{t+1}^{\theta y} [\Sigma_{t+1}^{yy} + \Sigma_{t+1}^y]^{-1} \quad (15)$$

where $\Sigma_{t+1}^{\theta y}$ is the cross covariance of parameter ensemble θ_t^{i+} and prediction ensemble \hat{y}_{t+1}^i .

In dual EnKF with parameter-state updating, the steps given by Eq (7) to (11) under parameter updating (section 3.2) are applied first to obtain the parameter updates. Then using the updated parameters given by Eq (10), new model state forecasts (x_{t+1}^{i-}) and discharge forecasts (\hat{y}_{t+1}^i) are generated.

$$x_{t+1}^{i-} = f(x_t^{i+}, u_t^i, \theta_{t+1}^{i+}) \quad (16)$$

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta_{t+1}^{i+}) \quad (17)$$

The model states ensemble is updated by using

$$x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^x (y_{t+1}^i - \hat{y}_{t+1}^i) \quad (18)$$

The Kalman gain (K_{t+1}^x) for correcting the state trajectories and is obtained by

$$K_{t+1}^x = \sum_{t+1}^{xy} [\sum_{t+1}^{yy} + \sum_{t+1}^y]^{-1} \quad (19)$$

where \sum_{t+1}^{xy} is the covariance of state and prediction ensembles.

3.4. Uncertainties in Input and Output

The spread of the ensemble members is determined by the specified error in the model structure, the input forcing and discharge data. Realistic assumptions for errors in input forcing and response are essential for proper assimilation of data by filtering.

3.4.1. Input forcing errors

For areal average rainfall derived in operational flood forecasting systems with a limited number of rain gauges, the uncertainties can be up to 50% (Willems, 2001). In this study, a preliminary estimate of input error term based on Weerts and El Serafy (2006) is used:

$$P_{true} = P_{input} + \delta P \quad (20)$$

where $\delta P \sim N(0, (0.15P_{input} + 0.2)^2)$

3.4.2. Output measurement errors

The uncertainty in the discharge measurement can be obtained from the rating curve calibration data for a given gauging station. However, in this study, a preliminary estimate of discharge measurement error is chosen similar to that of Georgakakos (1986) who assumed a standard deviation of 0.1 times the measured discharge as given below.

$$Q_{true} = Q_{measured} + \delta Q \quad (21)$$

where $\delta Q \sim N(0, (0.1Q_{measured})^2)$

4. APPLICATION TO GUDGENBY RIVER

The PDM model was calibrated using daily and hourly data from Gudgenby River by the SCE algorithm. The PDM model was first run using daily rainfall data until the beginning of the flood event and then with hourly rainfall data. The calibrated parameters of the PDM model are given in Table 1. The root mean square error and the coefficient of efficiency of the calibration are 15 m³/s and 0.87 respectively using all the data.

Table 1. The parameters of the PDM model.

Parameter	Value	Parameter	Value
C _{max}	476	k ₁	1
C _{min}	0	k ₂	5.4
b	1.42	k _g	1000
b _e	5.0	S _t	3.28
b _g	1.26	t _d	0
k _b	20		

The variance of noises introduced to the input forcing and flow measurements are proportional to their magnitudes as stated in (20) and (21). The EnKF and its four variations described in Section 3 was then applied to the May 2003 flood event. For EnKF with state updating, the four stores in the PDM model (S₁, S₃, S₂₁, S₂₂) were considered as state variables and were updated sequentially as new measurements became available. The standard deviation of the four state variables was selected by sensitivity analysis and is summarized in Table 2. To obtain the lead time and peak forecasts, a perfect knowledge of the future rainfall was assumed to avoid the error in forecasting rainfall. In reality, the uncertainty in the rainfall forecasts will add to the other uncertainties in the forecasting process. Forecasts were made at 1-, 3-, 6-, 9- and 12-hour lead times at every forecast time. At each forecast time, the magnitude and time of the forecast peak were also obtained.

Table 2. State variable standard deviations.

State variable	S ₁	S ₂₁	S ₂₂	S ₃
Standard deviation	1.0	0.3	0.15	0.02

For EnKF with parameter updating, three of the PDM model parameters, namely, C_{max}, b_g and k₂ were updated. These were the most sensitive parameters of the PDM model. The updating procedure was initialised by defining prior uncertainty range associated with these three parameters as given in Table 3. As the initial ensemble of parameters had to be specified, these three parameters were randomly sampled from a normal distribution with the standard deviations given in Table 3 which were obtained by sensitivity analysis. For dual EnKF, state-parameter updating was considered where both state variables and parameters were sequentially updated.

Table 3. Standard deviation and range of the PDM parameters updated.

Parameter	Minimum	Maximum	Std deviation
C _{max}	100	450	3.5
b _g	0.5	2.0	0.175
k ₂	6.0	12.0	0.175

5. MODEL EVALUATION

The adequacy of the EnKF was evaluated by using the root mean square (RMSE) and the coefficient of efficiency for 1, 3, 6, 9 and 12 hour forecasts. The RMSE is defined as

$$RMSE_L = \sqrt{\frac{\sum_{i=1}^n \{Q_{f,L}(i) - Q_{obs}(i)\}^2}{n}} \quad (18)$$

where $Q_{f,L}(i)$ forecasted discharge for lead time L for forecast i and Q_{obs} corresponding observed discharge.

The coefficient of efficiency of a model is defined as the proportion of the variance of the observed discharge accounted by the model (Nash and Sutcliffe, 1970):

$$CoE_L = 1 - S/S_{obs} \quad (19)$$

$$\text{where } S = \sum_{i=1}^n \{Q_{f,L}(i) - Q_{obs}(i)\}^2$$

$$S_{obs} = \sum_{i=1}^n \{Q_{obs}(i) - \bar{Q}_{obs}\}^2$$

$$\bar{Q}_{obs} = \sum_{i=1}^n Q_{obs}(i)$$

In addition, the error in the peak discharge magnitude and the timing were also used in the evaluation. To make objective comparison of the performance of EnKF with different options, a perfect knowledge of future observed rainfall is assumed in obtaining lead time forecasts and peak discharges.

6. DISCUSSION

For each EnKF method applied to Gudgenby River, the average values of the root mean square of error and coefficient of efficiency for the four events for lead times of 1, 3, 6, 9, 12 hours are given in Tables 4 and 5 respectively. These statistics were determined only for the period where the observed discharge is greater than 30 m³/s to avoid the small discharge values influencing the statistics. In flood forecasting, small discharges are not important.

It can be seen from Table 4 that the EnKF with parameter updating gave the smallest RMSE for all the lead time. Likewise, it gave the largest coefficient of efficiency for all the lead time. The quality of the forecasts deteriorated with lead time. As observed in an earlier study (Srikanthan et al. 2007), the EnKF with parameter updating performs the best compared to the other three variations.

The 1, 3 and 6-h lead time forecasts with 95% forecast limits are shown in Figure 1 while Figure

2 shows the forecasts for lead times 9 and 12 hours for event 1.

Table 4. Comparison of RMSE.

Variations of EnKF	Lead time (h)				
	1	3	6	9	12
State	11.27	14.33	18.24	20.80	22.19
Par	8.61	11.31	14.72	17.33	19.12
State-Par	9.72	12.76	16.61	19.03	20.44
Par-State	12.16	15.20	18.91	21.40	22.67

Table 5. Comparison of the Coefficient of efficiency.

Variations of EnKF	Lead time (h)				
	1	3	6	9	12
State	0.901	0.841	0.745	0.671	0.630
Par	0.940	0.898	0.831	0.770	0.723
State-Par	0.925	0.875	0.792	0.730	0.689
Par-State	0.889	0.826	0.730	0.656	0.617

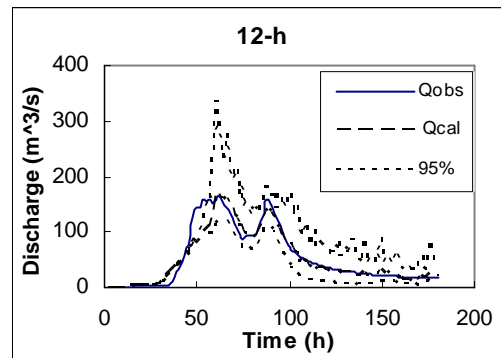
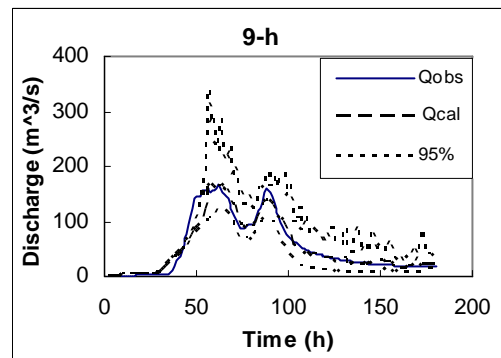


Figure 3. Comparison of flood forecast for 9- and 12-hr lead times.

The errors in the peak magnitude expressed as a ratio of the forecasted to observed peak for the four flood events were averaged and plotted as a function of time to peak in Figure 4.

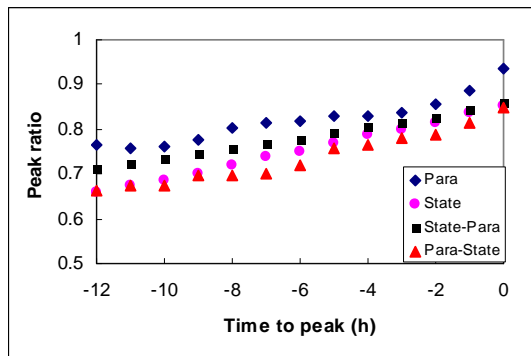


Figure 4. Variation of forecast peak error with time to peak.

This figure also shows the superior performance of parameter updating over the other variations. In terms of the timing error, there was not much difference between the different variations and error was very small (of the order of one hour or less).

7. CONCLUSION

The ability of the EnKF with the PDM model to forecast discharge was evaluated by using four flood events in the Gudgenby River. Four variations of the EnKF, namely, the state, parameter and dual (state-parameter and parameter-state) were considered. The uncertainty in the input rainfall data and the output discharge measurements were represented by Gaussian white noise with zero mean and variance obtained from the published literature. The uncertainty in the states and the parameters was obtained by sensitivity analysis. Perfect knowledge of observed rainfall was used during the forecasts. The parameter updating performed better than the other three. However, the application of EnKF gave good results only for short lead time forecasts with the results deteriorating markedly for long lead times. The inability to give good forecasts for long lead times even with perfect knowledge of observed rainfall needs further investigation. A weakness of the EnKF approach is the need to specify a priori uncertainty in states, rainfall and discharge. This makes the elucidation of the sensitivity of the forecast errors in peak flow and timing to a priori uncertainty a pressing issue. It follows that accurate specification of those sources of uncertainty deemed as sensitive is essential. Stream gauging data can be used to quantify the errors involved in using the rating curve for obtaining the discharges from the stage measurements. Rainfall data from a dense rainfall network can give reliable estimates of the error involved in the rainfall data. Further work is in progress to quantify these sources of uncertainty

8. REFERENCES

- Burgers G., P.J. van Leeuwen, G. Evensen (1998) Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review* 126, 1719-1724.
- El Serafy G. Y. and A.E. Mynett (2004), Comparison of EKF and EnKF in Sobek River: Case study Maxau-Ijssel. In: Liong S-Y, Phoon K-K and Baovic V. (Eds.) *Proc. 6th International Conference on HydroInformatics*, World Scientific, pp. 513-520.
- Georgakakos, K.P. (1986) A generalized stochastic hydrometeorological model for flood and flash flood forecasting: 2. Case studies, *Water Resources Research*, 22(13), 2096-2106.
- Koster, T., G. El Sefary, A.W. Heemink, H.F.P. van den Boogaard and A.M. Mynett (2004) Input correction in rainfall-runoff models using Ensemble Kalman Filtering (EnKF). In: Lee, J.H.W. & Lam K.M. (Eds.) *Proc. 4th Int. Symp. On Environmental Hydraulics*, Hong Kong, 15-18 December 2004, Leiden Proc, Balkema, p. 1991-1996.
- Moore, R. J. (2007) The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, 11(1), 483-499.
- Moradkhani, H.S., S. Sorooshian, H.V. Gupta and P. Houser (2005), Dual state-parameter estimation of hydrologic model using ensemble Kalman filter, *Advances in Water Resources*, 28(2), 135-147.
- Srikanthan, R., P. Sooriyakumaran and J.F. Elliott (1997), Comparison of four real-time flood forecasting models. *24th Hydrology and Water Resources Symposium*, Auckland, 185-190.
- Srikanthan, R., G.E. Amirhanathan and G. Kuczera (2007) Application of Ensemble Kalman Filter to Real-time Flood Forecasting. *2nd International Conference of GIS/RS in Hydrology*, Water Resources and Environment (ICGRHWE' 07), Guangzhou, China.
- Weerts A.H. and G El Serafy (2005) Particle filtering and ensemble Kalman filtering for input correction in rainfall runoff modelling. *International conference on innovation advances and implementation of flood forecasting technology*, Tromse, Norway.
- Weerts A.H. and G El Serafy (2006) Particle filtering and ensemble Kalman filtering for State updating with hydrological conceptual Rainfall Runoff Models, *Water Resources Research*, 42, W09403, doi,10.1029/2005WR004093, 2006.
- Willems, P. (2001) Stochastic description of the rainfall input errors in lumped hydrological. *Stochastic Environmental Research and Risk Assessment* 15, 132-152.