Assimilation of streamflow discharge into a continuous flood forecasting model

YUAN LI[1,2], DONGRYEOL RYU[1], Q. J. WANG[2], THOMAS PAGANO[2], ANDREW WESTERN[1], PRASANTHA HAPUARACHCHI[2] & PETER TOSCAS[3]

1 *Department of Civil and Environmental Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia*
y.li40@pgrad.unimelb.edu.au

2 *CSIRO Land and Water, Highett, Victoria 3190, Australia*

3 *CSIRO Mathematics Informatics and Statistics, Clayton, Victoria 3168, Australia*

Abstract Four state updating schemes are explored to integrate the observed discharge data into a flood forecasting model. Hourly streamflow discharge measured in the Ovens River catchment, Australia, is assimilated into the Probability Distributed Model (PDM) using the ensemble Kalman filter. The results show that the overall forecast accuracy improves when the discharge observations are integrated, mainly due to better initialisation of the model. Setting error covariance proportional to each state variable gives better results than setting error covariance as a constant value. Updating routing states of PDM affects discharge prediction instantly, while the effect of soil moisture updating results in a lagged response in discharge leading to a poorer update performance. However, during the forecast lead time, updating soil moisture results in slower degradation of the forecast accuracy, which is mainly because the soil moisture store is the only state influencing discharge volume, while the routing storages only describe the flow delay.

Key words discharge assimilation; flood forecasting; ensemble Kalman filter; state updating

INTRODUCTION

Flood has destructive impacts on people and their living environment. Flood forecasting, with sufficient lead time and accuracy, has great significance for effective flood warning and emergency response. However, forecasting models, the core of the quantitative flood forecasting systems, are still far from perfect, although they have improved considerably from conceptual to process-based ones to date. Moreover, useful information existing in the monitoring data collected during the antecedent periods and any response observed up to the start of the forecast period need to be utilized. Therefore, data assimilation, aiming to merge model forecasts and observation data to reduce forecasting uncertainty, receives growing attention as a way to further improve the accuracy of the forecasting system.

In general, hydrological data assimilation can be separated into three categories: error prediction, parameter estimation, and state updating (Anctil *et al.*, 2003; Sene, 2008). Error prediction, regarded as a post-processing method, is to correct the errors between model outputs and observed values (Anctil *et al.*, 2003), while parameter estimation can adjust parameters to achieve an improved forecast (Vrugt *et al.*, 2006). According to recent work, state updating is known to be more effective for the real-time forecasting process (Clark *et al.*, 2008; Komma *et al.*, 2008; Sene, 2008; Seo *et al.*, 2009; Thirel *et al.*, 2010), and there is some research showing the benefits of updating states and parameters simultaneously (Moradkhani *et al.*, 2005; Vrugt *et al.*, 2006; Salamon & Feyen, 2009).

State updating involves assimilating available observations into forecasting models to improve the overall prediction. Gauged discharge data are effective and commonly preferred for assimilation, since the purpose of real-time updating in flood forecasting is to obtain better discharge forecasts. While soil moisture and evapotranspiration assimilations are also becoming increasingly important, particularly in other areas of hydrology, they are still not easy to effectively use in operational flood forecasting. There are various kinds of state updating methods. The most widely used methods are sequential data assimilation and variational data assimilation. Variational data assimilation is essentially a smoothing method that requires a set of observations during a period of time, while sequential data assimilation, such as the Kalman filter (KF), is more suitable for real-time updating as it can adjust state variables whenever a new observation is available. The extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) are two widely-used sequential methods, adapted from the standard Kalman filter for nonlinear dynamic systems. Both require estimates of model and observation error to determine the optimal combination of model and observations. The EKF estimates model error by propagating error covariance matrix using a linearized operator (Clark *et al.*, 2008). This approach is inefficient for complex systems and unstable when applied to highly nonlinear models. On the other hand, the EnKF does not require linearization of the models to propagate ensemble states and to estimate model error covariance at updating time steps (Evensen, 1994). It has proven to be relatively efficient and suitable for complex nonlinear models.

There have been a number of streamflow forecasting studies based on the EnKF in the past decade (Moradkhani *et al.*, 2005; Vrugt *et al.*, 2005, 2006; Weerts & El Serafy, 2006; Andreadis *et al.*, 2007; Clark *et al.*, 2008; Komma *et al.*, 2008; Pauwels & De Lannoy, 2009). These studies argue that, despite the more detailed specification of physical processes in process-based models, simple models can run more efficiently and the state updating can be implemented more easily with them.

In fact, current flood forecasting systems used in most countries are mainly conceptual models (Sene, 2008). Although the conceptual models are rather simple compared to the physically-based hydrological models, integrating state updating into the models is still challenging because of some technical issues. One of the most notable problems is to quantify error covariance and select appropriate state variables to update. In this paper we test four updating schemes to demonstrate the impacts of different error quantification and choice of states to update on real-time forecasting.

DATA AND METHODS

Data

The study area is the Ovens River catchment upstream region of the Myrtleford observation station in northeast Victoria, Australia. The catchment area is 1210 km^2 and the stream channel network is about 8000 km long. The catchment drains the northern slopes of the Great Dividing Range. Data used in this paper include hourly averaged precipitation from the raingauges, monthly averaged potential evapotranspiration, and hourly discharge at the catchment outlet. With these data, model parameter calibration is carried out from 1 January 1999 to 21 July 2004 using the Shuffled Complex Evaluation (SCE) algorithm (Duan *et al.*, 1992). State updating is carried out for two continuous flood events from 22 July to 2 August 2004 using EnKF. Based on the updated state variables, the forecasting is implemented for the next two flood events from 3 to 25 August 2004. All the data used are continuous observations.

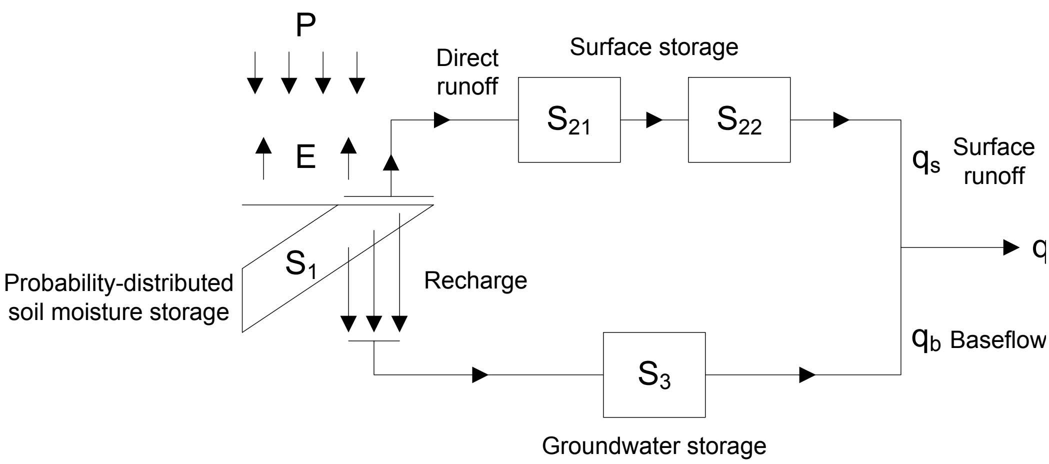

Fig. 1 The PDM model with two linear storages for surface runoff routing.

PDM model

The Probability Distributed Model (PDM) is a conceptual rainfall–runoff model which calculates the soil moisture storage by a spatial probability distribution and routes the surface flow and the subsurface flow by linear and nonlinear routing methods, respectively (Moore, 2007). In this paper, we use a cascade of two linear storages as the surface routing model and a cubic nonlinear storage as the groundwater routing model. The model structure is shown in Fig. 1. More detailed description of the PDM model can be found in Moore (2007) and Srikanthan *et al.* (2008).

EnKF

When using the ensemble Kalman filter for sequential state updating, there are two main steps: model prediction and state updating. In the model prediction step, state variable ensembles are propagated within the state space and into the observation space through the following two equations:

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

$$\hat{y}_{t+1}^i = h(x_{t+1}^{i-}, \theta) \tag{2}$$

where x_{t+1}^{i-} is the predicted state ensemble, x_t^{i+} is the updated ensemble, u_t^i is the input data vector, θ is the model parameter vector, ω_t^i is the error of model with a Gaussian probability distribution, $\hat{y}_{t+1}^i$ is the prediction vector, *t* is the time step and *i* is the ensemble member number.

In the state updating step, predicted state ensembles are updated as the following equation:

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

where y_{t+1}^i is the observation vector generated by adding Gaussian noise representing the observation error, K_{t+1} is the Kalman gain matrix which can be calculated by the following equation:

$$K_{t+1} = \Sigma_{t+1}^{xy-}[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^{y}]^{-1} \tag{4}$$

where Σ_{t+1}^{xy-} is the cross covariance matrix of the state variables and prediction, Σ_{t+1}^{yy} is the error covariance matrix of the prediction, and Σ_{t+1}^{y} is the observation error covariance matrix.

Equations (3) and (4) show the method of implementing the difference between discharge observation and prediction to update state variables in this paper. For more details about the application of EnKF in hydrological forecasting, refer to Moradkhani *et al.* (2005) and Srikanthan *et al.* (2008).

RESULTS AND DISCUSSION

As mentioned above, among the key information required for optimal state updating are the errors in model predictions. Model uncertainty comes from input data, initial states, model parameters and model structure. Errors from these parts will transfer into the state variables and then to the output data of subsequent time steps as the model runs forward. Therefore, in using the ensemble Kalman filters we can update state variables of the current time step by perturbing state variables and sometimes also input rainfall data. Updating all the state variables is possible but one thing we should notice is that the error of rainfall–runoff states, such as soil moisture, will transfer into, and may be the main source of catchment routing states. Thus, as mentioned in the PDM model (Moore, 2007), many applications only focus on updating soil moisture or catchment routing states.

The forecasting model used in this paper has four state variables: the soil moisture storage S1; the surface routing storage S21 and S22; and the subsurface routing storage S3. In order to demonstrate the impact of the different state variables and state error quantification on updating performance, we designed four updating schemes: (1) update S22 and S3 with a fixed magnitude of error; (2) update S22 and S3 with varying errors as a specified fraction of the state variables; (3) update S1 with a fixed magnitude of error; and (4) update S1 with varying errors as a specified fraction of the variable. The standard deviations are shown in Table 1.

Table 1 Standard deviations of state variables for perturbation.

Standard deviation	Scheme 1	Scheme 2	Scheme 3	Scheme 4
S1 (mm)	–	–	10	S1×10%
S22 (mm)	0.8	S22×10%	–	–
S3 (mm)	0.1	S3×10%	–	–

S1, soil moisture; S22, the second surface routing storage, S3, subsurface routing storage. The values of the states used in Scheme 2 and Scheme 4 are the mean of ensemble states.

Model performance during updating periods

Figure 2 demonstrates the results of four continuous updating schemes. Within less than 20 hours, the predictions with updating of all the four schemes reach a high precision compared with the observations. Schemes 1 and 2 outperform schemes 3 and 4. However, this is not surprising as S22 and S3 are the state variables that directly determine the surface discharge and baseflow discharge, while S1 only affects discharge through routing storages. This is also why the predictions of updating scheme 1 and scheme 2 improve faster than scheme 3 and scheme 4 and why updating soil moisture results in lags during peak periods. From the left panel of Fig. 2, we can see that scheme 2 performs well during the whole period, whilst scheme 1 is worse during peak periods. This is mainly because of the assumed constant error of S22 and S3 in scheme 1. During the peak periods, the perturbations of states are not sufficient for the prediction to be adjusted fully to the observations in scheme 1. The predications of scheme 3 and scheme 4 have little difference which is mostly because the soil moisture does not vary much during the whole updating periods.

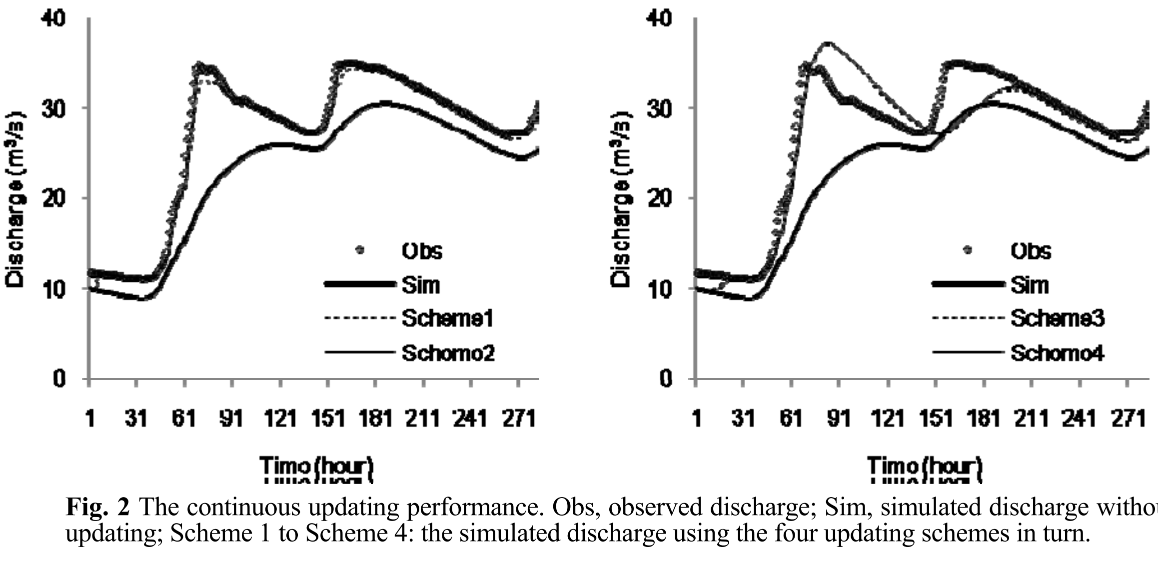

Fig. 2 The continuous updating performance. Obs, observed discharge; Sim, simulated discharge without updating; Scheme 1 to Scheme 4: the simulated discharge using the four updating schemes in turn.

Forecast based on state updating

As noted before, quantifying state errors as certain ratios of state variables is more rational and gives better results in the updating period. Consequently, using the updated state variables obtained from updating scheme 2 (forecast 1) and scheme 4 (forecast 2), we ran forecasts for 24 days that include the next two runoff events (Fig. 3). Observed rainfall was used to represent the forecast rainfall, which assumes the rainfall forecasting is as accurate as the observations of rainfall.

The results show that during the subsequent days, the improvement from updated initial states is significant but the benefits decrease as lead time increases. The benefits from routing storages updating decrease more rapidly compared with the soil moisture updating, even though updating routing storages affects the predicted discharge more directly and gives better analysis results than updating soil moisture (Fig. 2). This is probably partly because there is shorter memory in the routing storages, particularly the surface routing, which is important during events, compared with the soil moisture storage. It is also partly due to the soil moisture having a direct impact on the volume of water generated during forecast events, whilst routing storages are only used to describe the delay of water flow and the effects could be direct but short. Thus updating routing states without updating soil moisture will not provide much benefit at longer forecast lead times. The

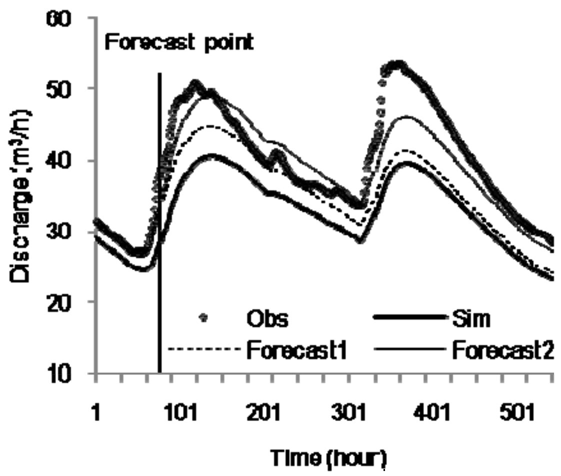

Fig. 3 Short time forecasts based on state updating. Obs, observed discharge; Sim, simulated discharge without updating; Forecast 1, forecast discharge based on updating scheme 2; Forecast 2, forecast discharge based on updating scheme 4.

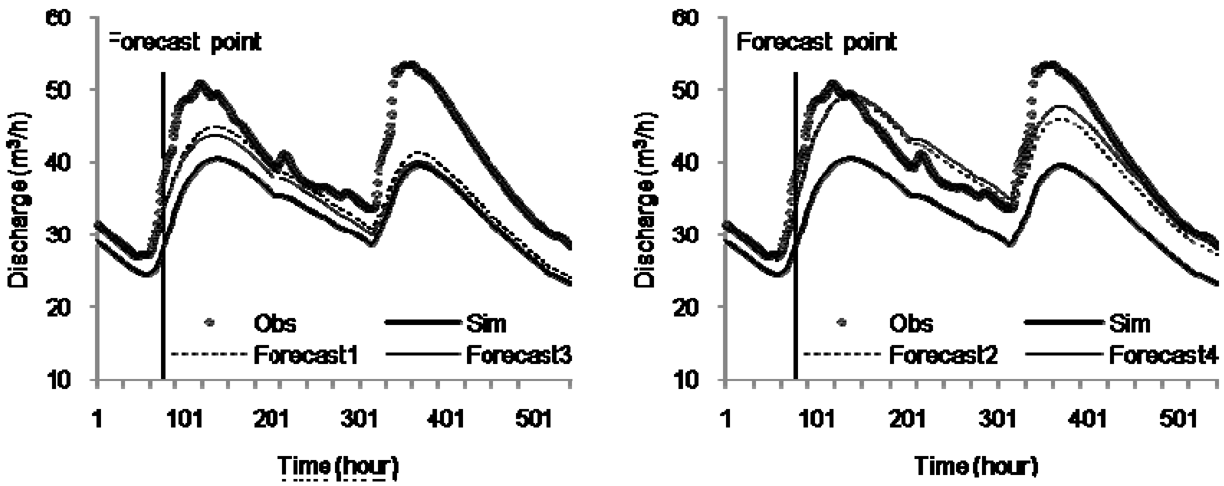

Fig. 4 Short time forecasts based on state updating using perturbed rainfall observations. Forecast 3 (left panel): forecast discharge based on updating scheme 2 using perturbed rainfall; Forecast 4 (right panel): forecast discharge based on updating scheme 4 using perturbed rainfall.

model forecasting performance indicates that for flood forecasting purposes, updating soil water states is more important than updating routing states.

In order to simulate the real-time forecasting as realistically as possible, we used the perturbed rainfall to forecast discharge based on state updating schemes 2 and 4, resulting in forecast 3 and forecast 4, respectively. The standard deviations were set to be 30% of rainfall observations and a zero mean temporally independent normally distributed error was added to the observed hourly rainfall. As shown in Fig. 4, the degradation of the input data influences the forecasts to some extent. The degradation of the rainfall input had surprising little impact on the quality of the forecast in both cases. The results indicate that the choice of update scheme has more impact than the rainfall forecase error; however, this may not hold with more realistic forecast errors that could have a bias over a whole event. The slight improvement in forecast 4 compared with forecast 2 is probably a chance outcome associated with the particular realisation of rainfall errors.

CONCLUSION

With the continuous catchment observation data we calibrated the PDM model using SCE. Using EnKF, we tested four continuous state updating schemes. Based on the updated state variables, we ran the model for streamflow forecasting and analysed the impacts of different updating schemes and rainfall perturbation schemes on forecasting results.

The results show that continuous state updating can significantly improve the model performance. Updating routing states affect discharge prediction directly, while the effect of soil moisture updating has a time lag, which implies that updating the soil moisture may only not result in the prompt improvements of discharge as much as it is seen for the routing states updating. Setting error covariance as a ratio of each state variable for state perturbation gives better results than setting error covariance as a constant value. Despite the time lag, soil moisture updating gives better forecasts after the updating period than routing states updating. That is mainly because the soil moisture store is the only state influencing runoff volume, while the routing storages only describe the flow delay. When using the perturbed precipitation for forecasting, the error of rainfall will add more uncertainty to forecasted discharge but state updating still makes sense for the forecasting within the lead time of a few days.

REFERENCES

Anctil, F., Perrin, C. & Andreassian, V. (2003) Ann output updating of lumped conceptual rainfall/runoff forecasting models. *J. Am. Water Resour. Assoc.* **39**(5), 1269–1279.

Andreadis, K. M., Clark, E. A., Lettenmaier, D. P. & Alsdorf, D. E. (2007) Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophys. Res. Lett.* **34**(10), L10403, doi:10.1029/2007GL029721.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J. & Uddstrom, M. J. (2008) Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.* **31**(10), 1309–1324.

Duan, Q., Sorooshian, S. & Gupta, V. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models, *Water Resour. Res.* **28**(4), 1015–1031.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143–10162.

Komma, J., Bloschl, G., & Reszler, C. (2008) Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting. *J. Hydrol.* **357**(3–4), 228–242.

Moore, R. J. (2007) The PDM rainfall–runoff model. *Hydrol. Earth System Sci.* **11**(1), 483–499.

Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. R. (2005) Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* **28**(2), 135–147.

Pauwels, V. R. N. & De Lannoy, G. J. M. (2009) Ensemble-based assimilation of discharge into rainfall–runoff models: A comparison of approaches to mapping observational information to state space. *Water Resour. Res.* **45**, W08428, doi:10.1029/2008WR007590.

Salamon, P. & Feyen, L. (2009) Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *J. Hydrol.* **376**(3–4), 428–442.

Sene, K. (2008) *Flood Warning, Forecasting and Emergency Response*. Springer, Berlin, Germany.

Seo, D. J., Cajina, L., Corby, R. & Howieson, T. (2009) Automatic state updating for operational streamflow forecasting via variational data assimilation. *J. Hydrol.* **367**(3–4), 255–275.

Srikanthan, R., Amirthanathan, G. & Kuczera, G. (2008) Application of ensemble Kalman filter for flood forecasting in Australian rivers. *Australian J. Water Resour.* **12**(3), 245–255.

Thirel, G., Martin, E., Mahfouf, J. F., Massart, S., Ricci, S. & Habets, F. (2010) A past discharges assimilation system for ensemble streamflow forecasts over France – Part 1: Description and validation of the assimilation system. *Hydrol. Earth System Sci.* **14**(8), 1623–1637.

Vrugt, J. A., Gupta, H. V. & Nuallain, B. O. (2006) Real-time data assimilation for operational ensemble streamflow forecasting. *J. Hydromet.* **7**(3), 548–565.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. & Verstraten, J. M. (2005) Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.* **41**(1), W01017, doi:10.1029/2004WR003059.

Weerts, A. H. & El Serafy, G. Y. H. (2006) Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall–runoff models. *Water Resour. Res.* **42**(9), W09403, doi:10.1029/2005WR004093.