Hamid Moradkhani, Soroosh Sorooshian, Hoshin V. Gupta, Paul R. Houser, Kuo-lin Hsu (2004), **Sequential data assimilation framework for hydrologic state-parameter estimation and ensemble forecasting**, in *Proceedings of the 2nd international CAHMDA workshop on: The Terrestrial Water Cycle: Modelling and Data Assimilation Across Catchment Scales*, edited by A.J. Teuling, H. Leijnse, P.A. Troch, J. Sheffield and E.F. Wood, pp. 191–195, Princeton, NJ, October 25–27

Sequential data assimilation framework for hydrologic state-parameter estimation and ensemble forecasting

Hamid Moradkhani[1], **Soroosh Sorooshian**[1], **Hoshin V. Gupta**[2], **Paul R. Houser**[3], **Kuo-lin Hsu**[1]

[1]*Civil and environmental Engineering Department, University of California, Irvine, California, USA*
[2]*Hydrology and Water Resources Department, University of Arizona, Tucson, Arizona, USA*
[3]*Hydrological Sciences Branch, NASA-GSFC, Greenbelt, Maryland, USA*

One of the primary purposes of hydrologic models is to perform prediction using physical relationships bounded by parameters and state variables. Much of the efforts in simulation-based hydrologic-systems analysis have been primarily focused on (1) improved parameter estimation methods that do not include state variables or (2) improved time-varying state estimation with predetermined parameters. Although the parameters of a hydrologic model can be estimated in a batch-processing scheme, there is no guarantee that model behavior does not change over time; therefore model adjustment over time may be required. Additionally, due to the multiplicative nature of errors in forcing data and observation, it is prudent to assemble the parameter adaption in the state evolution and forecasting system. The need for the real time state-parameter estimation of hydrological models has been reported in several studies (*Todini et al.*, 1976; *Kitanidis and Bras*, 1980; *Bras and Rodriguez-Iturbe*, 1985; *Young*, 2002; *Moradkhani et al.*, 2004). In this paper we extend the applicability of ensemble Kalman filter (EnKF), a recursive Data Assimilation (DA) technique, with Monte Carlo parameter smoothing to sequentially estimate model parameters and state variables. The applicability and usefulness of the current algorithm is demonstrated for the streamflow forecasting in Leaf River Watershed located north of Colins, Mississippi, using a conceptual hydrologic model, HyMOD (*Boyle*, 2000). This methodology offers two additional features: (1) the various sources of uncertainties can be properly addressed, including input, output and parameter uncertainties, (2) unlike the batch calibration procedures; the algorithm is recursive and therefore does not require storage of all past information.

Rcursive state-parameter estimation using EnKF Data assimilation techniques have garnered hydrologist's attention with the potential to use real time observations to produce more accurate hydrological forecasts. The basic objective of data assimilation is to characterize the system state at some future time given initial state knowledge. EnKF, a Monte Carlo approach of Kalman filter proposed by *Evensen* (1994) and later clarified by *Burgers et al.* (1998), is a DA algorithm suitable for nonlinear dynamic systems which uses a forecast model to integrate an ensemble of model states from one update time to the next and employs ensemble-based covariances in the update step to address the uncertainty in state estimation. To extend the applicability of the EnKF to simultaneous state-parameter estimation, we need to treat the parameters similar to state variables. Combined estimation can be provided by joint estimation where state and parameter vectors are concatenated into a single joint state vector (state augmentation). An alternative approach to joint estimation is dual estimation; designed as two interactive filters motivated either by the need to estimate state from the model (parameters) or by the need to estimate the model from state (*Moradkhani et al.*, 2004). In combined estimation, parameter evolution needs to be set up artificially, i.e., it is assumed that the parameters follow a random walk. The drawback of such parameter sampling is the loss of information between time points resulting in posterior distribution of parameters that are too diffuse comparing to the posteriors of fixed parameters (*Liu*, 2000). One remedy to this problem is to use the Kernel smoothing of parameter samples introduced by *West* (1993) wherein the conditional evolution density of parameters is written as follows:

$$P\left(\theta_{t+1}|\theta_t\right) \sim N\left(\theta_{t+1}^{i-}|a\theta_t^{i+} + (1-a)\overline{\theta}^{i+}, h^2 V_t\right) \tag{5.1}$$

Where, θ_t^{i+} and θ_{t+1}^{i-} are respectively the updated and forecasted parameter vectors of i^{th} kernel at time t and $t+1$, a is a factor ranging in $0.95 \sim 0.99$, V_t is the variance of normal kernels, and h is the smoothing parameter.

The generic discrete-time nonlinear stochastic dynamic system and predictions in the EnKF framework can be respectively expressed in the form of

$$x_{t+1}^{i-} = f\left(x_t^{i+}, u_t^i, \theta_{t+1}^{i-}\right) \tag{5.2}$$

$$\hat{y}_{t+1}^{t} = h\left(x_{t+1}^{i-}, \theta_{t+1}^{i-}\right) \tag{5.3}$$

Where x_t^{i+} and x_{t+1}^{i-} are updated and forecasted state ensemble members at time t and $t+1$ respectively. u_t^i is the perturbed forcing data according to $u_t^i = u_t + \zeta_t^i$, $\zeta_t^i \sim N\left(0, \sum_t^u\right)$.

In the updating step, observation y_{t+1} needs to be perturbed in the amount of η_{t+1}^i, therefore parameters are updated as follows:

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^{\theta}\left(y_{t+1} + \eta_{t+1}^i - \hat{y}_{t+1}^i\right), \qquad \eta_{t+1}^i \sim N\left(0, \sigma_{t+1}^y\right) \tag{5.4}$$

Where, K_{t+1}^{θ} is the Kalman gain associated with the parameters (*Moradkhani et al.*, 2004). Now using the updated parameters, we regenerate the model state and prediction trajectories as follows:

$$x_{t+1}^{i-} = f\left(x_t^{i+}, u_t^i, \theta_{t+1}^{i+}\right) \tag{5.5}$$

$$\hat{y}_{t+1}^{i} = h\left(x_{t+1}^{i-}, \theta_{t+1}^{i+}\right) \tag{5.6}$$

Model states ensemble is similarly updated as follows:

$$x_{t+1}^{i+} = x_{t+1}^{i-} + \Sigma_{t+1}^{xy}\left[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^{y}\right]^{-1}\left(y_{t+1}^i + \eta_{t+1}^i - \hat{y}_{t+1}^i\right) \tag{5.7}$$

Where K_{t+1}^x is the Kalman gain associated with the state variables. The flowchart of dual state-parameter estimation using EnKF with kernel smoothing of parameters is demonstrated in Figure 5.3.

Results and Discussion The applicability and usefulness of the dual EnKF on state-parameter estimation of the conceptual Hydrologic MODel (HyMOD) described by *Boyle* (2000) was investigated (Figure 5.4). State variables in this system are S: storage in the nonlinear tank representing the watershed soil moisture content, x_1, x_2 and x_3: the quick-flow tank storages representing the temporary (short-time) detentions, e.g., depression storages, and x_4 as the slow-flow tank storage (subsurface storage). Correspondingly parameters of this model are C_{max}, as the maximum storage capacity within the watershed, b_{exp} the degree of spatial variability of the soil moisture

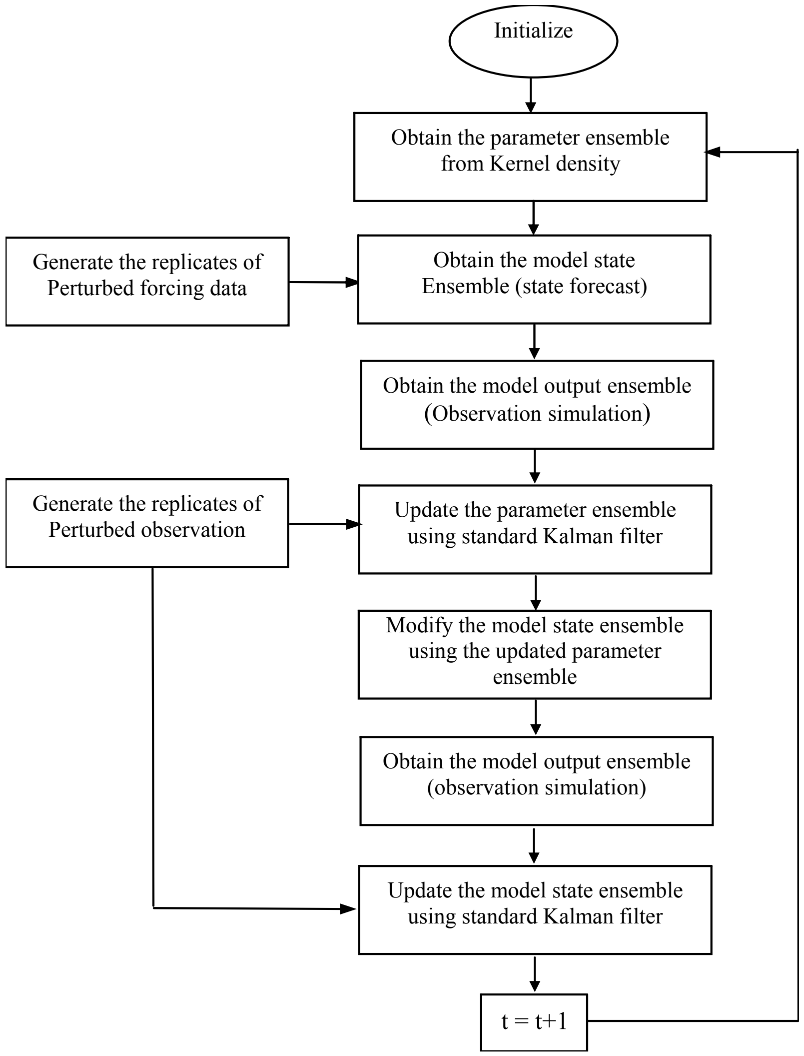

Figure 5.3: Dual state-parameter estimation flowchart using ensemble Kalman filter and kernel smoothing of parameters.

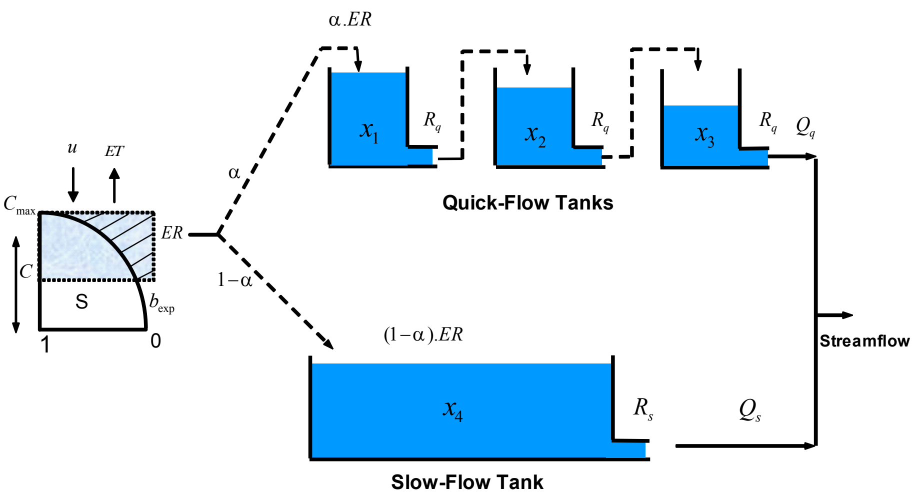

Figure 5.4: Hydrologic MODel (HyMOD) conceptualization.

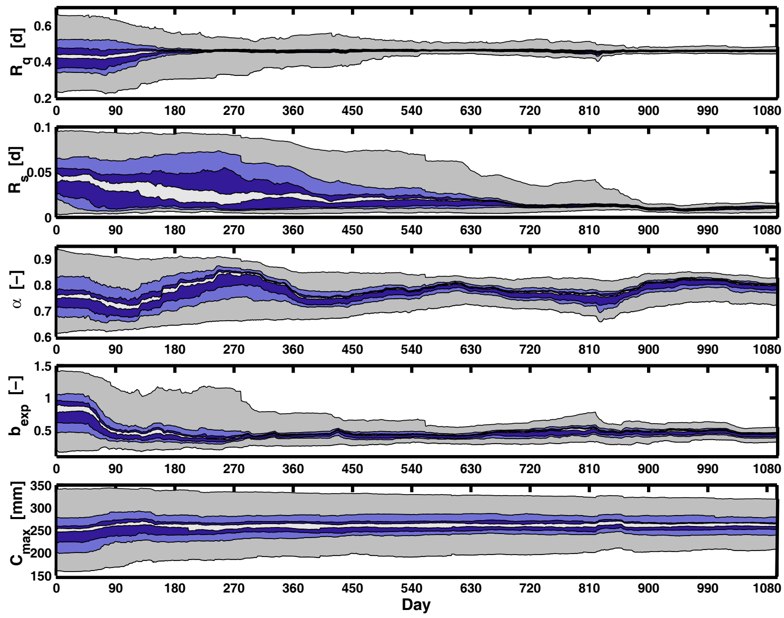

Figure 5.5: Time evolution of HyMOD model parameters for 3 years of dual ensemble filtering in Leaf River Watershed. Shaded areas correspond to 95, 75, 66 and 10 percentile confidence intervals.

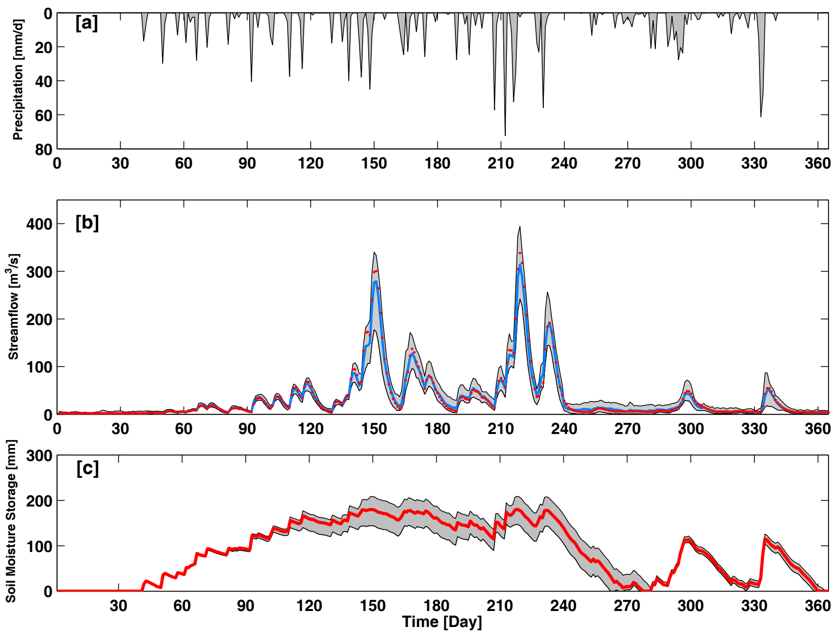

Figure 5.6: Results of the dual EnKF by application to the HyMOD, [a] Precipitation (forcing data); [b] streamflow forecasting with 95% uncertainty range; [c] Soil moisture storage variation (storage in the nonlinear tank of the HyMOD model). The solid line denotes the mean ensemble prediction.

capacity within the watershed, α, a factor for partitioning the flow between two series of tanks, R_q and R_s as the residence time parameters of quick-flow and slow-flow tanks respectively. The system is initialized by defining the prior uncertainty range associated with the parameters and state variables. Figure 5.5 displays the time evolution of HyMOD model parameters after dual filtering for the water years of 1950–1953. As seen, quick flow tank parameter R_q is the most identifiable parameter by showing the fastest convergence with minimum degree of uncertainty comparing to the others. In contrast, the maximum storage capacity of the watershed displayed by C_{max}, is less identifiable than the others and shows the slowest convergence. Ensemble time variation of the key state variable, S, representing the watershed soil moisture content, along with streamflow forecasting as predictive variable in the system are demonstrated in Figure 5.6. The streamflow forecasting result is very consistent with the observation; as a result state estimation as non-observable quantity shown in Figure 5.6 could be a reliable estimate.

In summary, the current algorithm introduces a number of novel features against the traditional calibration schemes: (1) both model states and parameters can be estimated simultaneously, (2) the algorithm is recursive and therefore does not require storage of all past information, as is the case in the batch calibration procedures, (3) the various sources of uncertainties can be properly addressed, including input, output and parameter uncertainties.

Bibliography

Boyle, D., Multicriteria calibration of hydrological models, Ph.D. thesis, Univ. of Arizona, Tucson, 2000.

Bras, R., and I. Rodriguez-Iturbe, *Random functions in hydrology*, Addison Wesley, Reading, Massachusetts, USA, 1985.

Burgers, G., P. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724, 1998.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, 1994.

Kitanidis, P., and R. Bras, Adaptive filtering through detection of isolated transient errors in rainfall-runoff models, *Water Resour. Res.*, *16*(4), 740–748, 1980.

Liu, F., Bayesian time series: analysis methods using simulation based computations, Ph.D. thesis, Duke University, 2000.

Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser, Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, in review, 2004.

Todini, E., A. Szollosi-Nagy, and E. Wood, Adaptive state-parameter estimation algorithm for real time hydrologic forecasting; a case study, in *IISA/WMO workshop on the recent developments in real time forecasting/control of water resources systems; Laxemburg (Austria)*, 1976.

West, M., Mixture models, Monte Carlo, Bayesian updating and dynamic models, *Comp. Sci & Stat.*, *24*, 325–333, 1993.

Young, P., Advances in real-time flood forecasting, *Philos. Tr. Roy. Soc.*, *360*, 1433–1450, 2002.