6-1-2006

Investigating the Impact of Remotely Sensed Precipitation and Hydrologic Model Uncertainties on the Ensemble Streamflow Forecasting

Hamid Moradkhani
Portland State University

K. Hsu
University of California, Irvine

Y. Hong
NASA Goddard Space Flight Center

S. Sorooshian
University of California, Irvine

Citation Details

Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian (2006), Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, Geophys. Res. Lett., 33, L12401.

GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L12401, doi:10.1029/2006GL026855, 2006

Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting

Hamid Moradkhani,[1] K. Hsu,[1] Y. Hong,[2] and S. Sorooshian[1]

Received 9 May 2006; accepted 23 May 2006; published 29 June 2006.

[1] In the past few years sequential data assimilation (SDA) methods have emerged as the best possible method at hand to properly treat all sources of error in hydrological modeling. However, very few studies have actually implemented SDA methods using realistic input error models for precipitation. In this study we use particle filtering as a SDA method to propagate input errors through a conceptual hydrologic model and quantify the state, parameter and streamflow uncertainties. Recent progress in satellite-based precipitation observation techniques offers an attractive option for considering spatiotemporal variation of precipitation. Therefore, we use the PERSIANN-CCS precipitation product to propagate input errors through our hydrologic model. Some uncertainty scenarios are set up to incorporate and investigate the impact of the individual uncertainty sources from precipitation, parameters and also combined error sources on the hydrologic response. Also probabilistic measures are used to quantify the quality of ensemble prediction. **Citation:** Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian (2006), Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, *Geophys. Res. Lett.*, *33*, L12401, doi:10.1029/2006GL026855.

1. Introduction

[2] The key to potential improvement of hydrologic prediction is associated with the input, parameter, and initial condition uncertainty interdependencies. Scenario analysis of hydrologic model by statistical characterization of streamflow uncertainty through ensemble forecasting-updating goes one step further for better understanding of these interactions. Precipitation is the key forcing variables and to a large degree responsible for model dynamics through its spatio-temporal variability. As shown by *Clark and Slater* [2006], uncertainty in model simulation is strongly influenced by reliability on forcing variable and adequate characterization of their associated uncertainties. Traditionally, the uncertainties of the rainfall runoff process and model response are captured through the calibration process. The automatic calibration methods using the Maximum Likelihood framework try to filter the effect of uncertainties through an additive error term [*Sorooshian and Dracup*, 1980].

[3] In this study, we seek to characterize various sources of uncertainties and their impacts on hydrologic model response. This is accomplished by incorporating the individual and combined error sources into the hydrologic model consistent with the limited data that allows quantitative assessment of prediction uncertainty. We employ the particle filter as a sequential ensemble forecasting/updating technique to characterize the uncertainties of model components. Our analysis is built upon a parsimonious conceptual rainfall-runoff model applied to the Leaf River basin in Mississippi. The PERSIANN-CCS (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network–Cloud Classification System) product [*Hong et al.*, 2004] is used as the forcing data into the system while its uncertainty in terms of variance is propagated into hydrologic model. Finally model predictive uncertainty is evaluated in conjunction with other uncertainty sources.

2. Precipitation Error Model

[4] A recent development by *Clark and Slater* [2006] provides conditional ensemble grid of precipitation in a sparse rain gage network for mountainous areas with the intention of model forcing ensemble generation for a distributed hydrologic model. However, for ungauged regions and also those regions where rain gage data are missing, one has to rely on remotely sensed precipitation products. In this study, we use the PERSIANN-CCS satellite precipitation product and estimate its error conditioned on the radar data which is assumed as true field. The error is quantified according to the variance of the point by point difference between the satellite estimate and radar truth. As demonstrated by *Hong et al.* [2006], this conditional error is related to precipitation magnitude at different spatio-temporal resolutions.

$$\sigma_e = f\left(\frac{1}{L}, \frac{\Delta t}{T}, P\right) = a \cdot \left(\frac{1}{L}\right)^b \cdot \left(\frac{\Delta t}{T}\right)^b (P)^d \qquad (1)$$

Where, σ_e is the error in precipitation (standard deviation between the satellite and radar data) which is a function of spatial coverage A (here substituted by L as spatial scale, the side length of A), temporal scale (T), satellite sampling frequency (Δt), and the space-time average of precipitation rate (P). Also a, b, c, and d are the parameters of error model need to be calibrated (for more detail see *Hong et al.* [2006]). The temporal scale used for calibrating precipitation error model ranges from hourly to daily and special scale considered ranges from 0.04° to 0.96°. As it appears from equation (1), the error for no precipitation is equal to zero which may not be a valid assumption. Considering the

[1]Department of Civil and Environmental Engineering, Center for Hydrometeorology and Remote Sensing, University of California, Irvine, California, USA.
[2]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

0094-8276/06/2006GL026855$05.00

spatio-temporal scale of our interest in this study, the assumption of zero precipitation is not crucial, noting that dealing with this problem is not the thrust of this study.

[5] Finally to generate the precipitation ensemble the following equation is used:

$$P_e = P + \sigma_e \bullet \varepsilon \tag{2}$$

Where, P is the precipitation in the desired spatio-temporal resolution, ε is a random number normally distributed with mean 0 and standard deviation of 1 and P_e is the perturbed precipitation. Propagating each member of the precipitation ensemble into the deterministic rainfall-runoff model will result in an ensemble of model states and associated streamflow predictions.

3. Hydrologic Model, Study Region and Data

[6] To demonstrate the influence of various uncertainty sources in hydrologic response, we used the conceptual Hydrologic MODel (HyMOD) that has been used previously [*Boyle et al.*, 2000; *Moradkhani et al.*, 2005a, 2005b]. HyMOD is an extension of some of the lumped storage models developed in 1960s, and later to the case of multiple storages representing a spatial distribution of different storage capacities in a watershed. The Leaf River basin in Mississippi ($\sim$1945 km^2) was considered as the study region. The satellite precipitation product was taken from PERSIANN-CCS [*Hong et al.*, 2004]. The data were extracted for two water years of 2002–2003. To estimate the error associated with satellite product, the National Weather Service WSR-88D Stage IV radar rainfall data was extracted from NCEP (National Center for Environmental Prediction) and used as ground reference rainfall data.

4. State-Parameter Uncertainty Estimation Using SIR-Particle Filter

[7] *Moradkhani et al.* [2005b] presented a Bayesian updating procedure for uncertainty assessment of conceptual watershed model components. The procedure was developed based on the Sequential Importance Resampling Particle Filter (SIR-PF) which is a class of Bayesian filtering algorithms derived from a discrete description of Bayes rule [*Arulampalam et al.*, 2002]. In filtering, two sequential estimations are discerned, (1) forecasting of state variables that is the evolution (propagation) of the states from one measurement time to the next and (2) updating (correction or analysis) of the forecasted state variables with the availability of new incoming observation. Because of its stochastic nature, state x_k is a random variable; hence from Bayesian inference the pertinent information about x_k given observations up to time k can be extracted from the filtering posterior distribution $p(x_k|y_{1:k})$.

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \tag{3}$$

Where $p(y_k|x_k)$ is the likelihood, $p(x_k|y_{1:k-1})$ is the forecast density of states and $p(y_k|y_{1:k-1})$ is the normalizing factor, known as predictive distribution or evidence given by $p(y_k|y_{1:k-1}) = \int_{x_k} p(y_k|x_k)p(x_k|y_{1:k-1})dx_k$. For hydrologic application a practical solution to exact Bayesian scheme is to rely on discrete approximations of the above densities as follows:

$$p(x_k|y_{1:k-1}) \approx \sum_{i=1}^{N} w_{k-1}^i \delta\left(x_k - x_{k-1}^i\right) \tag{4}$$

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta\left(x_k - x_k^i\right) \tag{5}$$

Where x^i and w^i denote the value of ith particle and its weight and δ is the Dirac delta function. The weights w_k^i (filtering posterior) are determined through the recursive Bayesian scheme as $w_k^i = \frac{w_{k-1}^i \cdot p(y_k|x_k^i)}{\sum_{i=1}^{N} w_{k-1}^i \cdot p(y_k|x_k^i)}$.

[8] In general, filtering is used to recursively estimate the posterior distribution of the model state variables; however, the successful use of sequential data assimilation relies on unbiased model state prediction, which is closely linked with identifiability of parameters [*Vrugt et al.*, 2005; *Moradkhani et al.*, 2005a, 2005b]. *Moradkhani et al.* [2005b] extended the application of the SIR-PF Bayesian recursive technique to the adaptive inference of the joint posterior distribution of parameters and state variables. The use of this methodology relaxes the need for restrictive assumptions regarding the variables' probability density function; i.e., it can readily handle the propagation of non-Gaussian distribution through a nonlinear model.

5. Uncertainty Analysis Scenarios and Results

[9] To investigate the impact of different uncertainty sources on hydrologic response, three scenarios are considered as depicted in Figure 1. The synthetic true states and streamflow (Figure 1a) are considered as observed quantities followed by the uncertainty scenario analysis displayed in Figures 1b–1d. The targeted filtering for the scenario 1 in Figure 1b is just the state updating, while for scenarios 2 and 3 (Figures 1c–1d) the combined state-parameter updating is implemented. For each of the uncertainty scenarios, we aim to evaluate the performance of ensemble streamflow prediction considering the relative impact of uncertainty. It is noted that for this study we considered the river basin as a lumped system with the daily time scale. For demonstration purpose we display the uncertainty limit associated with the ensemble streamflow forecast for the combined uncertainty scenario for a period of 240 days in Figure 2. However, the assimilation/calibration was made for the period of 2 years and the probabilistic performance indices are calculated according to this analysis period. A simplistic conclusion might be driven that calculated ensemble range (95 percentile) envelopes the observation. However, closer probabilistic interpretation of generated ensemble is required. Therefore, following verification measures are chosen and the comparative results are provided.

5.1. Normalized Root Mean Square Error Ratio (NRR)

[10] It is a normalized measure of ensemble dispersion indicating how confidently the ensemble mean can be extracted from the ensemble spread. In fact NRR evaluates

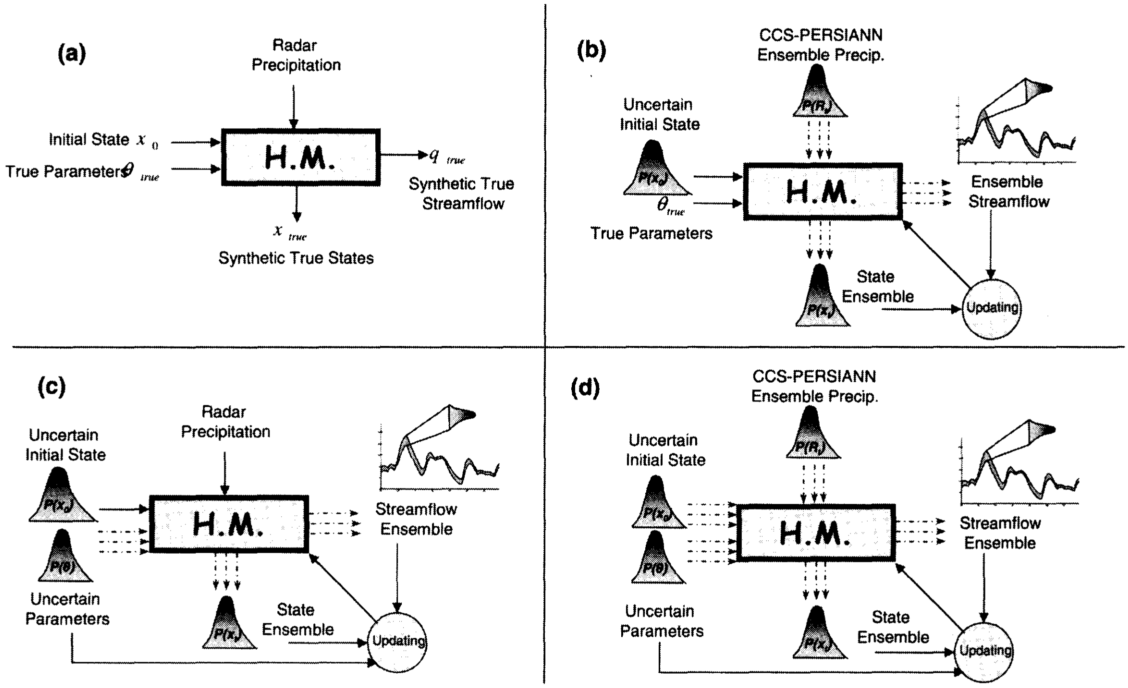

Figure 1. Scenario analysis for investigating the influence of various sources of uncertainties on ensemble streamflow forecasting: (a) Synthetic true, (b) forcing data error, (c) parameter uncertainty, and (d) combined uncertainties.

the sensitivity of the filtering scheme to the ensemble forecasting. According to this method, the ratio of the time-averaged *RMSE* of the ensemble mean to the mean *RMSE* of the ensemble members is calculated and the result is divided by $\sqrt{\frac{(n+1)}{2n}}$, where n is the ensemble size [*Moradkhani et al.*, 2005a]. The desirable ensemble is expected to have $NRR = 1$; while $NRR > 1$ indicates that the ensemble has too little spread, and $NRR < 1$ is an indication of an ensemble with too much spread. The *NRR* for three uncertainty scenarios is shown in Figure 3. The *NRR* of the forecast while including just forcing data uncertainty in modeling is $NRR = 0.76 < 1$ meaning that the ensemble has too much spread which is the result of fairly high value of uncertainty in satellite precipitation estimation for leaf river basin. The *NRR* while considering the parameter uncertainty is $NRR = 1.16 > 1$ meaning that ensemble has little spread although more precise with a little overconfidence. Finally, by accounting for all the

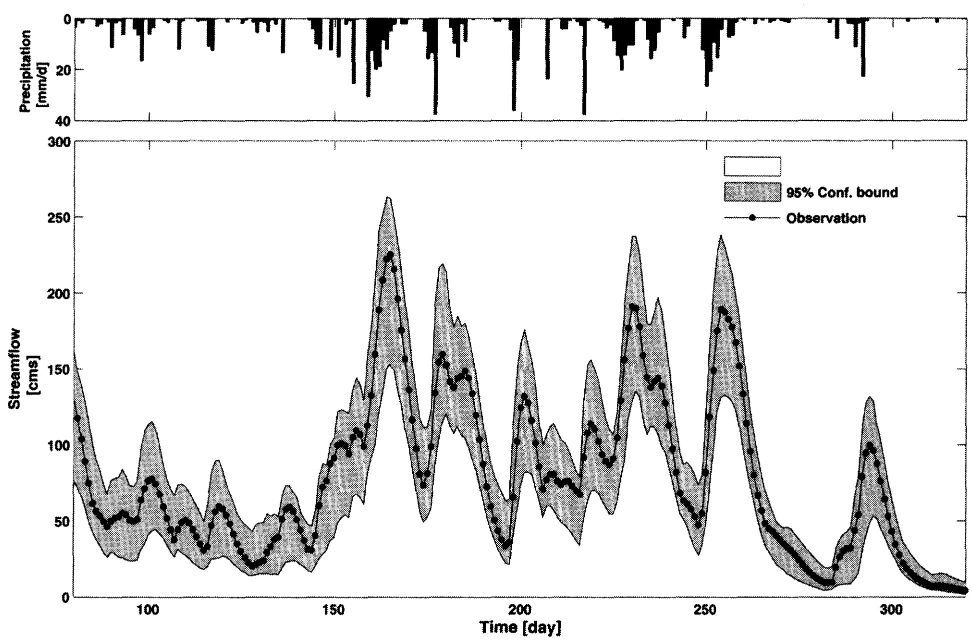

Figure 2. Time series of streamflow observation and 95% confidence bound associated with prediction while accounting for all uncertainty sources including initial condition, satellite forcing data and model parameter uncertainties.

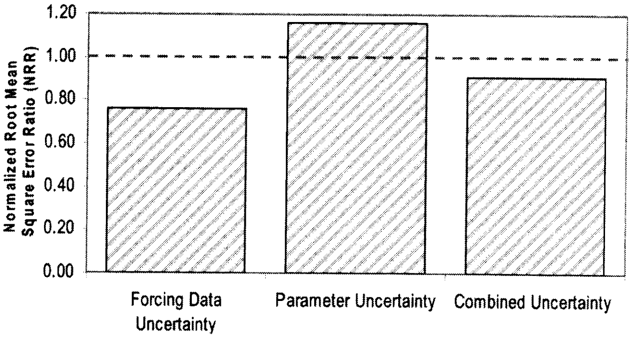

Figure 3. Normalized Root Mean Square Error Ratio (NRR) of ensemble streamflow forecasting by inclusion of different uncertainty sources in modeling.

uncertainty sources, the NRR = 0.91 is between the two $NRRs$ in individual uncertainty cases. This could be explained as the interaction of forcing data ensemble with parameter ensemble. In fact, assuming the time variation of model parameters through sequential Monte Carlo filtering will give the flexibility in interaction of model parameters and state variables which correspondingly alleviates the influence of forcing data error in model dynamics. This results to reconfiguration of model state and parameter ensemble through resampling of posterior distribution at each time step.

5.2. Exceedence Ratio (ER) and Uncertainty Ratio (UR)

[11] These two measures examine the spread (wideness) of prediction quantiles [*Borga*, 2002; *Hossain and Anagnostou*, 2005]. If the uncertainty bounds derived from ensembles are too wide, then the model is said to have high predictive capability with low precision and differently if the uncertainty bounds are derived too narrow then the overconfidence is put on prediction accuracy albeit precise, that is, simulation might be highly biased.

$$ER_n = \frac{N^n_{exceedence}}{T} * 100\% \tag{6}$$

$$UR_n = \frac{\sum_{t=1}^{T}\left(Q_t^{50+n/2} - Q_t^{50-n/2}\right)}{\sum_{t=1}^{T} Q_t^{obs}} * 100\% \tag{7}$$

Where, ER_n and UR_n are exceedence and uncertainty ratios respectively at nth percentile. $N^n_{exceedence}$ is the number of times during the total number of analysis period, T, that the observation Q_t^{obs} falls outside the ensemble bound at nth percentile. UR signifies the aggregate variability of prediction uncertainty ranges. Figure 4 represents the variation of exceedence and uncertainty ratios for a range of quantiles while accounting for each uncertainty source in simulation process. As seen in Figure 4, these two measures behave conversely where decreasing the ER for higher percentiles results in increasing UR. Closer examination of these figures reveal that for forcing data (precipitation) uncer-

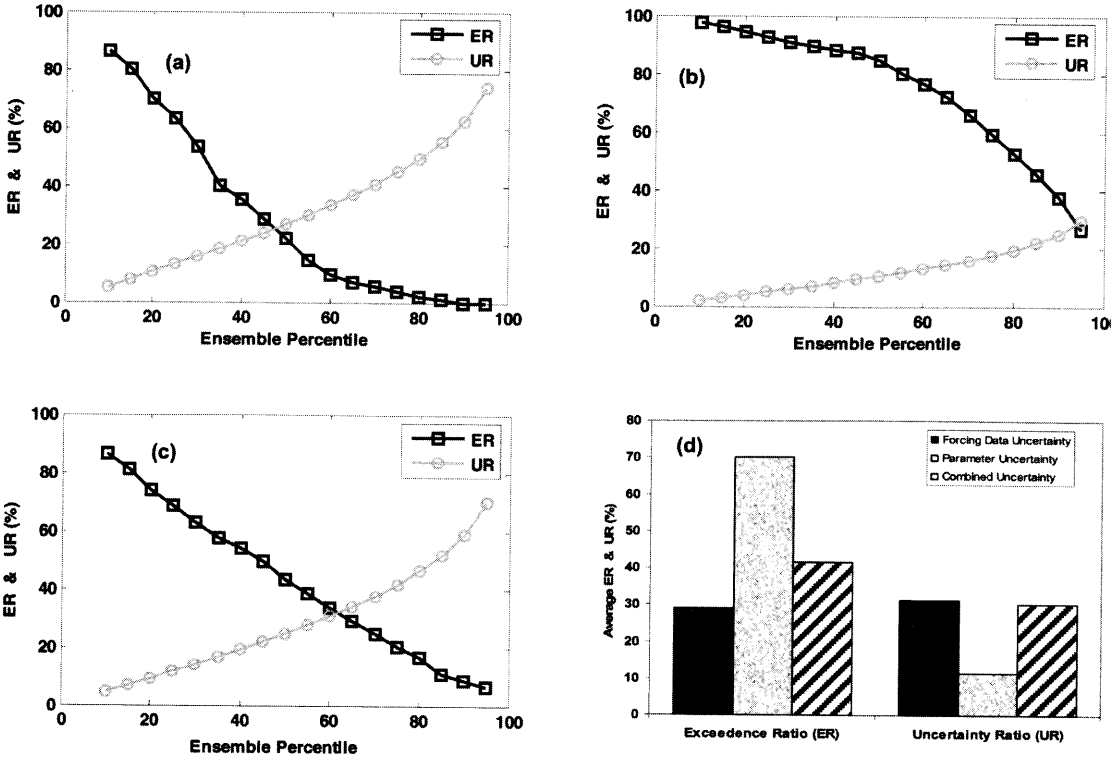

Figure 4. Exceedence and Uncertainty Ratios in ensemble streamflow forecasting: (a) Forcing data (precipitation) error, (b) parameter uncertainty, (c) combined uncertainty, and (d) comparison of average ratios for various uncertainty sources.

tainty, *ER* decreases rapidly as compared to parameter uncertainty case. This signifies lower precision in prediction due to the individual effect of precipitation uncertainty. On the other hand, *UR* for the precipitation case is generally greater than that of parameter uncertainty case meaning the wider uncertainty bound. One might argue that this is an indication of higher predictive capability than considering parameter uncertainty alone. However, this statistical measure needs to be interpreted cautiously in conjunction with *NRR* index discussed earlier. Figure 4d also shows the comparison of the average *ER* and *UR* over all quantiles for three scenarios. The result is consistent with the findings in *NRR* index as the combined uncertainty scenario ratios are between the ratios where individual uncertainty sources are considered.

6. Summary and Conclusion

[12] We aimed at building a framework for investigating the effect of various sources of uncertainties in hydrologic response. We used the PERSSIAN-CCS satellite precipitation product as forcing data to the hydrologic model and characterized its error trough a power law function. Assimilation of synthetic streamflow and state-parameter estimation was done within a sequential Bayesian framework. Three statistical measures suitable for verification of probabilistic (ensemble) forecasts were used and their competing characteristics illustrated while applying on a parsimonious hydrologic model. It was shown that satellite precipitation error reflects a wide uncertainty range in streamflow forecasting as opposed to the narrow range resulting from parameter uncertainty. It was also discussed that the ensemble filtering through combined state-parameter updating is capable of reducing the total uncertainty. In fact the resampling nested in such a filtering scheme allows the redistribution of model state and parameter samples. In summary, such scheme, in addition to introducing the sequential Bayesian method to estimate the model prediction uncertainty, offers three notable features that are lacking in conventional model calibration-assimilation schemes as (1) being sequential, therefore better able to take advantage of the temporal organization of information well suited for the ungauged basins such that by availability of new observation the model behavior is updated, (2) it is capable of incorporating all uncertainty sources into the modeling, (3) it employs the resampling scheme giving the chance to parameter and state particles to be relocated, thereby providing the flexibility in interaction of model components with forcing data through ensembles. This relaxes the impact of uncertainty of forcing data sequentially and result in reduced combined uncertainty.

References

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, *50*(2), 174–188.

Borga, M. (2002), Accuracy of radar rainfall estimates for streamflow simulation, *J. Hydrol.*, *267*, 26–39.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water. Resour. Res.*, *36*(12), 3663–3674.

Clark, M. P., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, *J. Hydrometeorol.*, *7*(1), 3–22.

Hong, Y., K. Hsu, X. Gao, and S. Sorooshian (2004), Precipitation estimation from remotely sensed imagery using Artificial Neural Network-Cloud Classification System, *J. Appl. Meteorol.*, *43*(12), 1834–1853.

Hong, Y., K. Hsu, H. Moradkhani, and S. Sorooshian (2006), Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response, *Water Resources Research*, doi:10.1029/2005WR004398, in press.

Hossain, F., and E. N. Anagnostou (2005), Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, *Adv. Water Resour.*, *28*, 1336–1350.

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

Moradkhani, H., K. Hsu, H. Gupta, and S. Sorooshian (2005b), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.

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

Y. Hong, NASA Goddard Space Flight Center, Mail Stop 613.1, Greenbelt, MD 20771, USA.

K. Hsu, H. Moradkhani, and S. Sorooshian, Department of Civil and Environmental Engineering, Center for Hydrometeorology and Remote Sensing, University of California, Irvine, CA 92697, USA. (moradkha@uci.edu)