Improving Flood Forecasting Skill with the Ensemble Kalman Filter

Mejoramiento de la capacidad de predicción de crecientes de caudales mediante el paquete de Kalman Filter

Humberto Vergara*, Yang Hong**, Jonathan J. Gourley***.

Abstract

The purpose of this particular work was to explore the benefits and drawbacks of sequential state updating for flood forecasting and identify factors or mechanisms affecting the updating process and thus controlling its performance. The Ensemble Kalman filter was employed to assimilate hourly streamflow observations into a simple but widely used conceptual rainfall-runoff model for flood prediction purposes. Ensembles were constructed by perturbing model forcing and parameters. Parametric perturbations were obtained from multiple model calibrations with an optimization algorithm. Errors in streamflow observations were characterized through an innovative yet simple empirical model. A sensitivity analysis was performed to evaluate the improvement of the first guess forecast. Additionally, the forecast skill was assessed as a function of lead-time. It was found that the improvement is mainly reflected in runoff volume, while the peak time

can be deteriorated as a trade-off of the assimilation process. Overall, ensemble-based models with sequential data assimilation outperformed the best-calibrated deterministic models for lead times of at least 1.5 days.

Keywords: Ensemble flood forecasting; Sequential data assimilation.

Resumen

El propósito de este trabajo en particular fue explorar las ventajas y los inconvenientes de la actualización del estado secuencial para la predicción de crecidas e identificar los factores o mecanismos que afectan el proceso de actualización y, por tanto controlan su funcionamiento. Se empleó el filtro de Kalman Ensemble de asimilar las observaciones de caudal por hora en un modelo de precipitación-escorrentía conceptual simple pero muy utilizado para fines de predicción de inundaciones. Conjuntos se construyeron perturbando modelo forzando y parámetros. Perturbaciones paramétricas se obtuvieron de múltiples calibraciones modelo con un algoritmo de optimización. Los errores en las observaciones de caudal se caracterizaron a través de un modelo empírico innovadora pero simple. Se realizó un análisis de sensibilidad para evaluar la mejora de la primera previsión conjetura. Además, la habilidad de pronóstico se evaluó como una función de tiempo de entrega. Se encontró que la mejora se refleja principalmente en volumen de escurrimiento, mientras que el tiempo de pico se puede deterioró como una compensación del proceso de asimilación. En general, los modelos basados en conjunto con la asimilación de datos secuencial superaron a los modelos deterministas mejor calibrados para los plazos de entrega de por lo menos 1,5 días.

Palabras Clave: Ensamble de predicción de crecidas; Asimilación de datos secuencial.

Recibido / Received: Mayo 25 de 2014 **Aprobado / Aproved:** Mayo 30 de 2014

Tipo de artículo / Type of paper: Artículo de investigacion científica y tecnólogica terminada

Afiliación Institucional de los autores / Institutional Affiliation of authors: *Grupo de investigación Choc Izone, Universidad El Bosque y Department of Civil Engineering and Environmental Science, **University of Oklahoma, Norman; Advanced Radar Research Center, ***University of Oklahoma, Norman; NOAA/National Severe Storms Laboratory, Norman.

Autor para comunicaciones / Author communications: Humberto Vergara, humber@ou.edu.com

Los autores declaran que no tienen conflicto de interés.

Subject Headings: Floods, Forecasting, Hydrologic models, Kalman filters, Stochastic models, Streamflow.

Introduction

Among all geophysical hazards, flooding events are considered to be the ones that produce the most devastating effects on lives and infrastructure on a global scale. On average, floods cause more than 20,000 deaths and adversely affect about 140 million people annually over the globe (Adhikari et al. 2010). During 2010, many major floods occurred worldwide (e.g., Australia, China, Colombia, and Pakistan) causing immeasurable damage and claiming the lives of many people. Many of these were reported by the media to be caused by very unusual heavy rainfall seasons. It is not surprising, thus, that intensive research efforts have been devoted to the prediction of this natural hazard.

Due to their success in meteorological forecast applications, data assimilation techniques are becoming popular in different areas of hydrologic modeling, including flood forecasting (e.g. Seo et al. 2003; Komma et al. 2008; Smith et al. 2008; Salamon and Feyen 2009). Data assimilation has been extensively studied in meteorology and oceanography for the past few decades. Indeed, the first attempts of data assimilation may be associated to early numerical weather prediction experiments in the 1920's where manual interpolation of point observations to a regular grid were performed, which was a tedious and time consuming procedure, and quickly motivated the development of various objective analysis, or data assimilation algorithms (Kalnay 2003). Applications in hydrology have been studied more recently. One of the first studies involving data assimilation was the work of Kitanidis and Bras (1980), where filtering techniques were employed to improve streamflow forecast lead times in a real-time set-up. These early efforts focused on the implementation of updating procedures for operational purposes, but research on the details related to how data assimilation works in hydrologic modeling has become more active only in recent years.

Some of the focus in hydrological data assimilation has been on the use of soil moisture data to improve prediction skill of land-surface and hydrologic models (e.g. Houser et al. 1998; 2001; Reichle et al. 2002; Aubert et al. 2003; Huang et al. 2008; Moradkhani 2008). Other endeavors have concentrated on the assimilation of streamflow observations. Seo et al. (2003) applied variational data assimilation to assimilate hourly streamflow, precipitation and potential evaporation data for operational purposes. Vrugt et al. (2005) developed the Simultaneous Optimization and Data Assimilation (SODA) algorithm where the EnKF is used in conjunction with a parameter optimization routine to get improved estimates of parameter and output uncertainties. In a similar work by Moradkhani (2005), the EnKF is used to update both model states and parameters at every assimilation cycle. More recently, emphasis has been placed on flood forecasting applications. Vrugt et al. (2006) discussed the applicability of SODA for real-time flood prediction using the Sacramento Soil Moisture Accounting model (SAC-SMA; Burnash et al. 1973) used operationally by the US National Weather Service (NWS). Neal et al. (2007) and Komma et al. (2008) used the EnKF for flood prediction purposes in different regions. Clark et al. (2008) attempted to propagate information to neighboring basins using an implementation of EnKF with a distributed hydrologic model. However, most of these studies have mainly evaluated the skill of the model prediction from one assimilation cycle to the next one (i.e., the first guess or one-step-ahead forecast) using overall measures of error such as the mean absolute error (MAE) or the broadly used root mean squared error (RMSE). The present study aims to examine the improvement of the hydrologic predictions, achieved through sequential state updating, in a more comprehensive manner. Specifically, three forecast components are analyzed from the first guess and from predictions of different lengths (i.e. lead times): a) Runoff volume error, which is a measure of the capability of the modeling system to predict the total amount of water in the flooding event; b) Peak magnitude error, which is an indicator of model's ability to predict how rare or extreme an event is; and c) Peak time error, which measures the skill of the model to make timely predictions. Consequently, an event-based approach is utilized. The overarching goal of this research is to develop a better understanding on the benefits and drawbacks of hydrological sequential state updating through the EnKF, as well as on the mechanisms that control its performance. Results from this work are expected to reveal information that can be used to optimize the implementation of EnKF for hydrologic applications.

The remainder of the document is organized as follows: section 2 presents the implementation of the modeling and state updating system, as well as the study area and datasets used in the experiments. Section 3 presents the experiments, results and discussion. Finally, section 4 summarizes the study and presents the main conclusions.

Hydrological data assimilation implementation.

Study area and hydrologic model

The Tar-Pamlico River basin in coastal North Carolina, which is part of the National Oceanographic and Atmospheric Administration (NOAA)'s Coastal and Inland Flooding Observation and Warning (CI-FLOW) project (Van Cooten et al. 2011) was selected as the area of study for this work. The basin is periodically affected by heavy rainfall from tropical storms and hurricanes, at which time major flood events occur. This study focuses on the sub-catchment of the US Geological Survey (USGS) station located at Tarboro (Streamgage No 02083500; Fig. 1), which includes the upper Tar River and Fishing Creek sub-basins. The catchment has a drainage area of 5,653 km^2 and is located on the coastal plain. The Tar is a perennial river with a mean daily flow value of about 62 m^3/s at Tarboro. The minimum recorded daily flow is 0.79 m^3/s, while the maximum is 1,996 m3/s, which occurred after the landfall of hurricanes Floyd and Dennis in September of 1999. Flooding and subsequent catastrophic events caused by these two storms were the trigger of the genesis of the CI-FLOW project. Extreme streamflow events are, therefore, a major concern for this basin. Table 1 presents streamflow levels for different recurrence intervals.

Figure 1. *Study area highlighting the upper Tar River and Fishing Creek basins. The hydrography of the basin is also presented.*

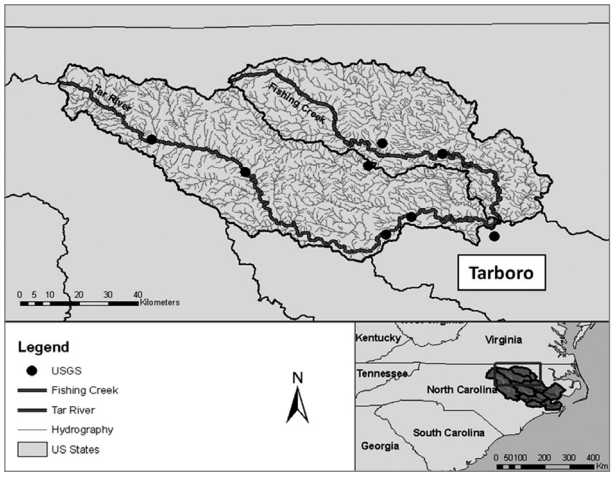


Table 1. *Frequency threshold flow levels for Tarboro. Thresholds for the 1, 2, 5, 10, 50, 100 and 500 year return period are presented. Levels are calculated using 24 years of hourly streamflow (1986-2010) and the Log-Pearson Type III distribution function.*

Frequency Threshold (years)	1	2	5	10	50	100	500
Streamflow (m3/s)	106	346	608	826	1,436	1,756	3,193

The Hydrologic MODel (HyMOD) was utilized to model streamflow at Tarboro. HyMOD is a lumped Conceptual Rainfall-Runoff (CRR) model, a subject of several hydrologic model calibration studies (e.g. Wagener et al. 2001; Vrugt and Bouten 2003; Vrugt et al. 2008) and more recently of research on hydrological data assimilation (e.g. Moradkhani et al. 2005; Vrugt et al. 2005; Vrugt and Robinson 2007; Smith et al. 2008). HyMOD is a rainfall excess model represented by a nonlinear tank, based on the Probability Distributed Model (PDM) developed by Moore (1985), whose output is routed through two series of linear tanks denoting quick and slow flow components of the streamflow.

The Multisensor Precipitation Estimate (MPE) rainfall product was selected as the rainfall forcing for HyMOD. MPE data combines information from satellite, radar, and rain gauges, and employs a blend of automated and interactive procedures for its derivation (Briedenbach and Bradberry 2001; Fulton 2002). Hourly MPE data are available for the Southeast River Forecast Center (SERFC) region at 4-km resolution from 2002 to present. Eight years of data (2002 – 2009) were collected for this study. Monthly mean potential evapotranspiration data were obtained for the same period from the Famine Early Warning Systems Network (FEWS NET; USGS and USAID 2011). Both datasets were spatially averaged for Tarboro's catchment.

Sub-hourly streamflow observations for Tarboro are available from October 1985 to present at the USGS Instantaneous Data Archive (IDA; USGS 2011). Hourly streamflow observations were derived from this dataset for the period of study. Given that the overall goal of this study is to improve flood forecasting, the performance of the hydrologic modeling system is assessed based on the simulation and prediction of significant rainfall-driven events. The criterion here used for defining an event is the frequency in terms of return period. Although the recurrence intervals associated to bankfull flow range from 1.0 to 2.5 years, a widely accepted average value for flood level is the 2-years return period (e.g. Carpenter et al. 1999; Gourley et al. 2011). In this work, flows exceeding the 1.0 year return period are considered an event while those exceeding 2.0 years level are assumed to be flooding events.

Twelve storm events, which produced flows exceeding the criterion described above, were selected from the past events inventory of the National Weather Service (NWS) forecast office at Raleigh in North Carolina (RAH-Web-Team 2011) and by examining the observed streamflow time series. Table 2 lists these events and some details related to them. Five of the eight flow events were driven by major storms (either tropical storm or hurricane), which are of particular concern for the region.

Table 2. *Selected events for the study. A short description indicating the type of storm, the exceeded frequency threshold, and the peak flow are presented. Peak flow is based on hourly streamflow at Tarboro.*

Date	Short Description	Exceeded Recurrence Interval (years)	Peak Flow (m^3/s)
March 1, 2003	No storm associated	2.0	374
March 20, 2003	Heavy rain event	2.0	402
April 9, 2003	No storm associated	4.0	595
May 23, 2003	No storm associated	1.5	337
September 17, 2003	Hurricane Isabel	1.5	273
August 9, 2004	Hurricane Charley	1.5	323
January 14, 2005	Severe weather event	1.0	209
June 14, 2006	Tropical storm Alberto	6.0	694
August 31, 2006	Tropical storm Ernesto	1.5	271
November 12, 2006	Severe weather event	2.0	433
April 3, 2008	Heavy rain event	1.0	179
November 10, 2009	Tropical storm Ida	1.0	203

Model Calibration

The hydrologic model was automatically calibrated in order to produce deterministic predictions to benchmark the performance of the assimilating system. The parameters of HyMOD were estimated using the Diffe-Rential Evolution Adaptive Metropolis (DREAM; Vrugt et al. 2009). DREAM is an adaptation of SCEM-UA (Vrugt et al. 2003) that uses a more sophisticated method for the estimation of the posterior probability density function of parameters in complex, high-dimensional sampling problems. The algorithm employs Markov Chain Monte Carlo sampling and uses a formal likelihood function. DREAM separates behavioral from non-behavioral solutions using a cutoff threshold that is based on the sampled probability mass (Vrugt et al. 2008). Time series of observed streamflow are used to objectively assess model's performance based on the proposed parameter values. The objective function used for the calibration was the Sum of Square Residuals (SSR), given by:

$$SSR = \sum_{i=1}^{N} \left(Q_i^{obs} - Q_i^{sim} \right)^2 \quad (1)$$

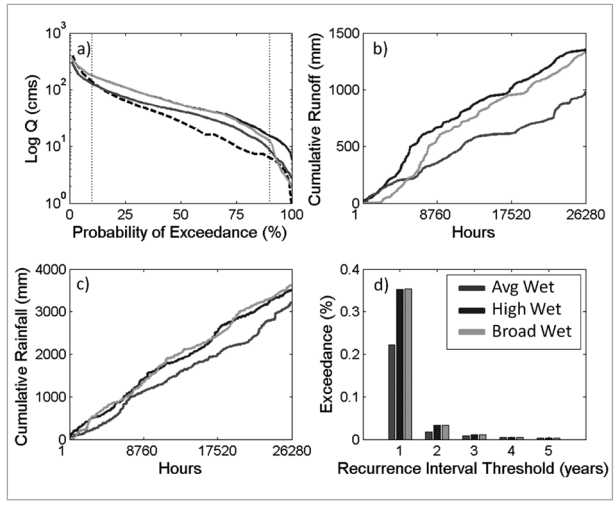
Where Q^{obs} and Q^{sim} are the observed and simulated streamflow, respectively, at time i and N is the number of data points. Parameter bounds used for the calibration of HyMOD are presented in Table 3. These ranges were derived from prior experience gained during calibration tests (not shown here).

Table 3. *Ranges for HyMOD parameters values specified for DREAM.*

Model Parameter	Lower Bound	Upper Bound
C_{max}	0	100,000
b_{exp}	0	2
α	0	1
N_q	1	10
R_q	0	1
R_s	0	1

A subset of data used for the calibration (i.e., calibration period) was selected objectively using a MATLAB® script that evaluates the data, for a prescribed length, based on information from total amounts and frequency of both rainfall and runoff. A length of three years was defined for the calibration process. The algorithm collected data for three-year windows starting from the beginning of the period of study and then advancing the window by three months until the last three years of available data are analyzed. Since the interest of the present study is flood prediction, wet periods with extreme events were the target. Although it is common to select one period for calibration in deriving deterministic models, three calibration periods were identified for this experiment. Figure 2 presents rainfall and runoff data for the selected periods. A common characteristic of the three periods is that they can all be considered wet periods (see panel a). The first period has the highest runoff as shown in panel b), and so is named "High

Figure 2. *Quantitative analysis of the selected three-year calibration periods; a) Flow Duration Curves (FDC) compared to the Period Of Data (POD) FDC. Vertical dashed lines indicate 10 and 90% probability of exceedance; b) Accumulated Runoff; c) Accumulated rainfall; d) Percent of time the 1,2,3,4 and 5 recurrence intervals are exceeded.*



Wet". The second period contains flows that most closely approximate the average or typical flow regime of the basin as indicated by the mid-segment of the FDCs in panel a), therefore referred to as "Avg Wet". The third and last period covers flow values as high as those in the "High Wet" period but also has the minimum flow values,

thus named "Broad Wet". The total amount of runoff and rainfall of the three periods are comparable, with the exception of the runoff of period "Avg Wet". Additionally, the three periods contain some extreme events as shown by panel d), which is a desirable characteristic given the modeling objective of this study.

In order to identify the best deterministic model, performance of the resulting models parameterized with the different parameter sets was evaluated on an independent time period. Table 4 presents the results of the three calibration runs during the calibration period and Table 5 during independent validation periods. The measures presented in these two tables are widely used in hydrologic modeling studies for assessing the overall skill of the model for simulating the long-term basin response to rainfall through a time series of streamflow. The percent bias is a measure of total volume difference (Smith et al. 2004) and is given by:

$$Bias(\%) = \frac{\sum_{i=1}^{N}\left(Q_i^{sim} - Q_i^{obs}\right)}{\sum_{i=1}^{N} Q_i^{obs}} \times 100 \qquad (2)$$

Table 4: Calibration results during calibration period. *The calibration period for every run is presented along with measures of closeness. NSCE for flows exceeding the 1-yr return period (H1YRP) and the 2-yr return period (H2YRP) thresholds are also included.*

Calibration	Bias (%)	RMSE (%)	NSCE	NSCE (H1YRP)	NSCE (H2YRP)
High Wet 09/01/02 – 08/31/05	0.63	74.53	0.39	-0.47	-8.61
Avg Wet 11/23/03 – 11/22/06	-10.39	68.99	0.63	0.17	-2.10
Broad Wet 06/01/02 – 05/31/05	5.70	77.64	0.37	-0.45	-9.52

Table 5. *Calibration results during validation period. This period refers to intervals other than the calibration period. Measures of closeness are presented. NSCE for flows exceeding the 1-yr return period (H1YRP) and the 2-yr return period (H2YRP) thresholds are also included.*

Calibration	Bias (%)	RMSE (%)	NSCE	NSCE (H1YRP)	NSCE (H2YRP)
High Wet	70.35	131.81	0.24	0.20	-1.73
Avg Wet	15.54	110.98	0.31	-0.90	-9.84
Broad Wet	67.19	122.66	0.32	0.18	-2.96

The range of Bias (%) is between -∞ to ∞, with a perfect value of zero. A Bias less than zero indicates underestimation, while a positive value implies the opposite. The percent Root Mean Square Error (RMSE) refers to the random error of the model performance (Willmott and Matsuura 2005) and is computed as:

The range of Bias (%) is between -∞ to ∞, with a perfect value of zero. A Bias less than zero indicates underestimation, while a positive value implies the opposite. The percent Root Mean Square Error (RMSE) refers to the random error of the model performance (Willmott and Matsuura 2005) and is computed as:

$$RMSE(\%) = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(Q_i^{sim} - Q_i^{obs}\right)^2}{N}}}{\bar{Q}^{obs}} \times 100 \qquad (3)$$

Where $\bar{Q}^{obs}$ is the mean of the observations. RMSE (%) ranges from zero to ∞, with zero as its perfect value. Lastly, the Nash-Sutcliff Coefficient of Efficiency (NSCE) which is a measure that indicates how well the model fits the observed data (Krause et al. 2005), is given by:

$$NSCE = 1 - \frac{\sum_{i=1}^{N}\left(Q_i^{sim} - Q_i^{obs}\right)^2}{\sum_{i=1}^{N}\left(\bar{Q}^{obs} - Q_i^{obs}\right)^2} \qquad (4)$$

The range of the NSCE score has an upper bound value of 1.0 (perfect) and a lower bound of -∞. A negative value of NSCE indicates that the mean value of the observations is a better predictor than the model being evaluated. Additionally to the results in Table 4 and Table 5, aggregated

event-based error statistics developed by Smith et al. (2004; equations 5 - 7) were computed for the specific events considered in this study and are presented in Table 6

$$E_R(\%) = \frac{\sum_{i=1}^{N} |B_i|}{NY_{avg}} \times 100 \quad (5)$$

$$E_P(\%) = \frac{\sum_{i=1}^{N} |Q_{P,i} - Q_{SP,i}|}{NQ_{Pavg}} \times 100 \quad (6)$$

$$E_T(hours) = \frac{\sum_{i=1}^{N} |T_{P,i} - T_{SP,i}|}{N} \quad (7)$$

Table 6. *Calibration results from the event-based evaluation. The percent absolute event runoff error (ER), percent absolute peak error (ER), and percent absolute peak time error (ET) are presented.*

Calibration	E_R (%)	E_P (%)	E_T (*hours*)
High Wet	42	44	24
Avg Wet	41	48	27
Broad Wet	37	43	24

Where ER is the aggregated event-based runoff error, EP is the aggregated event-based peak error, and ET is the averaged event-based peak time error. Bi is the bias of the i-th event, Yavg is the average observed flood event runoff volume of N events, QP is the observed peak discharge, QSP is the simulated peak discharge, QPavg is the average observed peak discharge of the N events, TP is the observed peak time, and TSP is the simulated peak time.

Even though the calibration from the "Avg Wet" period had the best long-term performance for the calibration interval, all three deterministic models performed very similarly during the independent validation period. As for the event-based analysis, the "Broad Wet"

Model seems to be the best. However, the three models' skill measures are comparable. Therefore, all deterministic models are considered as reference benchmarks in the following data assimilation experiments.

Assimilation approach

The technique selected in this study for updating model states was the Ensemble Kalman Filter (EnKF; Evensen 2003). The EnKF is a Monte Carlo simplification of the Extended Kalman Filter (EKF; Jazwinski 1970). The most important advantage of EnKF is that background error statistical information is retrieved from the ensembles, thus the linearized model and observation operator of the EKF are not necessary. There are two approaches in EnKF: Stochastic and Deterministic (Hamill 2006). The difference between the two is that in the former observations are perturbed by adding noise from a normal distribution with zero mean and standard deviation R, while the latter does not perturb observations. Perturbations of the observations are necessary because otherwise the error covariance of the analysis is systematically underestimated (Hamill 2006; Wang 2009). However, the noise added to the observations can have a detrimental effect (Clark et al. 2008). Whitaker and Hamill (2002) developed a deterministic EnKF version entitled the Ensemble Square Root Filter (EnSRF). The EnSRF uses a reduced Kalman gain (i.e. the weight of the innovations) to update the perturbations. The equations for this implementation are as follows:

$$X_{i,k}^{b} = M_{i-1,k}(X_{i-1,k}^{a}), \mathrm{K} = 1,2,3,...,\mathrm{L} \quad (8)$$

$$\overline{X^b} = \frac{1}{L}\sum_{k=1}^{L} X_k^b \quad (9)$$

$$X_k'^b = X_k^b - \overline{X^b} \quad (10)$$

Where L is the ensemble size, $x_{(i,k)}^{b}$ is the k[th] member of the background ensemble, $\overline{x}^b$ is the mean of the background ensemble, and x'^{b}_{k} is the perturbation. As for the background error covariance, its value is calculated from the ensemble:

$$P^b H^T \approx \frac{1}{L-1}\sum_{K=1}^{L}(X_k^b - \overline{X}^b)(\mathrm{H}(\mathrm{X}_k^b) - \overline{H(X^b)})^T \quad (11)$$

$$HP^b H^T \approx \frac{1}{L-1}\sum_{k=1}^{L}(H(X_k^b) - \overline{H(x^b)})(H(x_k^b) - \overline{H(x^b)})^T \quad (12)$$

The data assimilation step is divided into mean update and the perturbation update. The mean update:

$$\overline{x}^a = \overline{x}^b + K\left(y - \overline{H\left(x^b\right)}\right) \quad (13)$$

$$K = P^b H^T (HP^b H^T + R)^{-1} \quad (14)$$

Where K is the traditional Kalman gain. Now, the perturbation update:

$$X'^a_k = X'^b_k - \tilde{K}H(X^b_k)' \quad (15)$$

$$\tilde{K} = \left(1 + \sqrt{\frac{R}{HP^b H^T + R}}\right)^{-1} K \quad (16)$$

Where $\tilde{K}$ is the reduced Kalman gain. It can be seen that the perturbations are reduced less with $\tilde{K}$ than with K, yielding the same effect as with the EnKF with perturbed observations. The final analyses are then computed by:

$$x^a_k = \overline{x}^a + x'^a_k \quad (17)$$

The input data for the filter are ensembles of the states to be updated, an ensemble of the observation priors, and the actual observation along with its error information. The states in HyMOD are the initial soil moisture C, the content of water in the slow tank *x4*, and the content of water in each of the quick flow tanks denoted by x1, *x2* and *x3*. The implementation of HyMOD used in this study allows for different number of quick flow tanks N_q, and so the number of states will vary as the value of N_q changes.

The ensembles of model states and streamflow are produced perturbing different components of the modeling system. In this experiment, forcing and parameters are the only model components being perturbed prior to the assimilation of observed streamflow (model states are also perturbed through EnKF).

It is assumed that uncertainty from other sources (e.g. model structure) is implicitly accounted for to a certain degree by the ensembles and, thus, what is not explicitly addressed can be neglected. A more detailed explanation of the perturbations is presented in the following section. Error in observations is explicitly treated through a simple approach as explained in section 2.3.3.

Characterizing forcing, model and observation errors Rainfall

To account for uncertainty in model forcing, rain rates were perturbed using uniform random multipliers. The multipliers were generated by a simple model described in the work of Clark et al. (2008), and has the following form:

$$p' = p \times \varphi \quad (18)$$

Where p is a time series of the observed precipitation, p' is the perturbed time series of precipitation, and φ is the uniform random multiplier defined by:

$$\varphi = (1 - \varepsilon) + 2 \times u \times \varepsilon \quad (19)$$

Where u is a uniform random number such that φ is a realization from a uniform distribution ranging from $(1 - \varepsilon)$ to $(1 + \varepsilon)$, and ε is the expected error in the precipitation.

This model is mainly designed to account for volume error in the precipitation, an important source of uncertainty in hydrologic predictions. Since rainfall used in this study is the product used operationally in the SERFC, its error is assumed to be relatively low and of random nature. However, values of ε were arbitrarily inflated to account for the effect of basin aggregation.

Model parameters

Besides finding the best model parameters values, given a set of observations and the objective function used during calibration, DREAM has the ability to store information related to parametric uncertainty. Vrugt et al. (2008) discuss how the probability distribution functions of the hydrologic model parameters produced by DREAM contain all required information to summarize the predictive uncertainty. Data from the deterministic calibrations performed with DREAM were thus used to obtain parameter perturbations.

The multiple calibration runs described in section 2.1.1 were selected with the purpose of incorporating information from different sets (i.e. calibration periods). Even though these periods were selected to have wet conditions and extreme high flow events, the objective function in DREAM is intended to represent the mean error of the simulations, and therefore, parameter sets derived from these calibrations might not be skillful for reproducing extremes. For this reason, four additional calibrations with modified objective functions were run.

Instead of evaluating the SSR of the entire streamflow time series, a threshold is defined to separate high flows from low or average flows. The four threshold-based calibrations are referred to as H1YRP, L1YRP, H2YRP, and L2YRP. The first two used the 1-yr return period level as a threshold, while the latter two applied the 2-yr level. The "H" refers to flows higher than the threshold, while the "L" indicates the opposite. This information was expected to be translated into a more complete estimation of the parametric uncertainty than if doing it from a single calibration exercise.

The perturbations were extracted from DREAM by sampling from the parameter sets' pdf found during the calibration process. Since DREAM starts with no information about the prior distribution of the parameter sets (i.e. it assumes the distribution is uniform), the algorithm spends several iterations until reasonable parameter sets are proposed. Therefore, it was necessary to define a subspace of the parameter sets population (i.e. all parameters sets tested in DREAM) to assure stability of the perturbations and avoid misspecification of error statistics during the assimilation process, which can lead to ensemble simulations departing from the solution (i.e. filter divergence). Subspaces were then defined for each calibration run and different samples were obtained from them.

Streamflow observations

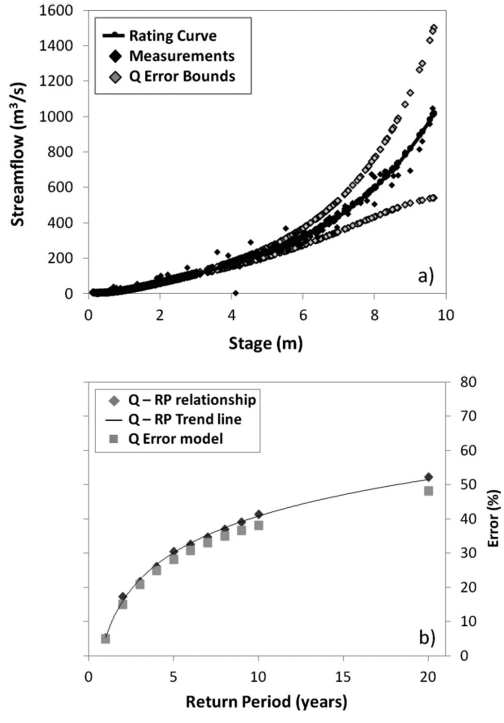
A simple model for estimating uncertainty in streamflow observations was developed for this study. The error model is based on standard error values reported in the literature (e.g. Sauer and Meyer 1992; Di Baldassarre and Montanari 2009), and frequency analysis of streamflow to address error heteroscedasticity in rating curve estimates. While this simple model can serve the purpose of this particular study, it is acknowledged that a strict treatment of the rating curve uncertainty deserves more elaborated and comprehensive approaches such as the ones proposed by Petersenoverleir (2004), Di Baldassarre and Montanari (2009), and McMillan et al. (2010), although these are not always easy to implement, especially when the required data is not available.

To tackle the issue of heteroscedasticity, it was assumed that streamflow observation errors behave as a function of the rareness of the flow in terms of recurrence interval. This assumption agrees with Sorooshian and Dracup (1980) who claim that high flows are more likely to have larger errors with higher deviations compared to low flows. It is well known that the rating curve uncertainty increases when the flow exceeds river bank-full conditions (i.e. during flooding), which is expected to be represented by this error model. The relationship between streamflow and recurrence interval for the study basin can be described by the following simple equation:

$$Q = 308.56 \ln RP + 106.04 \qquad (20)$$

Where Q is the streamflow and RP is the recurrence interval in years. Equation (14) was derived by a log regression fit for flows with return periods between 1 and 20 years (Fig. 3 b). Vrugt et al. (2005) fitted a spline function to observation data for the Leaf river watershed and found that the size of streamflow error almost increases log-linearly, which agrees with the assumption made here. Equation (20) is thus used as a proxy to model streamflow observation errors.

Figure 3. *a) Rating curve for Tarboro (USGS gage station 02083500) and uncertainty bounds estimated by the frequency-based streamflow error model; b) Streamflow – Return Period relationship for recurrence intervals between 1.0 and 20.0 years. Modeled observation error is also plotted on the secondary y-axis.*



For the case of flows that are below the 1-year threshold, a constant error value was assigned. Sauer and Meyer (1992) report that the error associated to good measurements is 3 to 6%. In their study of 17 different cross-sections and a wide range of streamflow values in the Po river in Italy, Di Baldassarre and Montanari (2009) found that the total uncertainty in the streamflow observations ranged from about 6 to 42%. In this study, a value of 5% was set for the error of low or average flows, and was used as the intercept in the log-linear component in the following equation:

$$Q_{error}(\%) = \begin{cases} 5, RP < 1.0\,yr \\ 14.4\,lnRP + 5, RP \geq 1.0\,yr \end{cases} \quad (21)$$

RP can be computed for any given flow using the reverse form of (20). The model was tested on historical field measurement data collected by USGS (Fig. 3 a).

Flood forecasting with EnKF

Two experiments involving the datasets and tools previously described were designed to analyze how data assimilation helps to improve flood forecasting. The first one was a sensitivity analysis to assess the overall improvement of the forecasts. The second one was a traditional evaluation of flood forecast skill of the modeling-assimilating system and a comparison with the benchmarks (i.e. deterministic models).

Sensitivity analysis

The primary objective of the sensitivity analysis was to examine how the data assimilation implementation improved the forecasts of the following event-based objectives: runoff volume, peak magnitude, and peak timing. A total of 484 ensembles were derived varying different characteristics including the nature of the perturbations, the size, and the spread of the ensemble. Based on the nature of the perturbations, the ensembles could be rainfall-based, parametric-based, or combined. Rainfall perturbations were generated by using values of ε (see section 2.3.1) in the range between 0.1 and 2.0. Model perturbed ensembles were combined with rainfall perturbations to create the "Combined" ensemble type. Different ensembles for these three categories were generated with sizes ranging from ten to two hundred members. The spread was given in terms of the mean standard deviation of the ensemble. Five of the selected events (within the "Avg Wet" period) were simulated with and without data assimilation using the first guess (i.e. the one-step-ahead forecast) to compute error and improvement statistics. Improvement scores were defined based on the aggregated event-based error measures defined in section 2.1.1:

$$I_R(\%) = \frac{E_{R,noDA} - E_{R,DA}}{E_{r,noDA}} \times 100 \quad (22)$$

$$I_P(\%) = \frac{E_{P,\mathrm{noDA}} - E_{P,DA}}{E_{\mathrm{P},noDA}} \times 100 \quad (23)$$

$$I_T(\%) = \frac{E_{T,\mathrm{noDA}} - E_{T,DA}}{E_{T,noDA}} \times 100 \quad (24)$$

Where I_R, I_P, and I_T are the percentage relative improvement in forecasting runoff error, peak error and peak time error, respectively, following data assimilation. The subscript *noDA* refers to the error associated to the ensemble mean simulation without data assimilation, while the subscript *DA* is for the error associated to the ensemble mean simulation after assimilating streamflow observations.

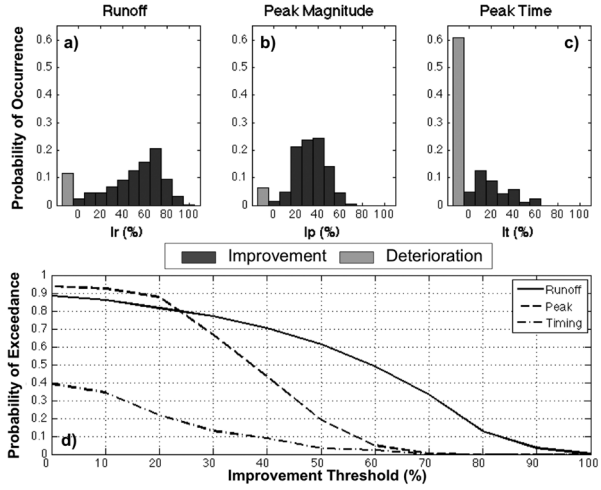
Where I_R, I_p, and I_T are the percentage relative improvement in forecasting runoff error, peak error and peak time error, respectively, following data assimilation. The subscript *noDA* refers to the error associated to the ensemble mean simulation without data assimilation, while the subscript *DA* is for the error associated to the ensemble mean simulation after assimilating streamflow observations.

Results from the sensitivity analysis are summarized in histograms for measure of improvement (see Fig. 4 a-c). Of the total number of ensembles generated for this experiment, only about 38% improved in all three objectives following data assimilation. The other portion is composed of cases where the forecast was deteriorated for at least one of the three objectives (light gray bars in Fig. 4). In general, the timing was the measure with the highest degree of deterioration (i.e. improvement with lowest probability of occurrence), while the peak

magnitude had the greatest improvement (i.e. the sum of the dark gray bars), although closely followed by runoff volume. In terms of levels of effectiveness, the runoff volume was superior for improvement rates above 30% (see Fig. 4 d) and also had the highest rates of improvement with a maximum of about 96%. These results suggest that the main forecast component improved by sequential data assimilation is the runoff volume. The improvement in predicted runoff volume, however, is contrasted by the deterioration in peak timing. Both phenomena might be explained by the fact that the error statistics are evaluated for each assimilation cycle, and thus are solely based on bias (i.e. the difference in magnitude of the predicted and observed variable). Additional information from previous cycles could potentially be included in the updating process in order to address the timing issue (e.g. the slope between the streamflow values of the prior and current cycles can be used to determine whether the predicted flow will be in the rising or in the recession limb of the hydrograph). However, the effect that this approach would have on the total skill of the forecast and/or the stability of the updating process is unknown and beyond the scope of the present study.

In addition to the overall assessment of event-based statistical improvement following data assimilation discussed above, information from the sensitivity analysis was used to select an ensemble for the evaluation of forecasts for single events. Association between the performance of the filter and the size and spread of the ensembles was first evaluated. Figure 5 presents scatterplots for the three directions of improvements versus ensemble size (upper panels) and ensemble spread (lower panels). Although it was observed that increasing the size of the ensemble could help in cases where the filter diverged (not shown here), it can be seen from the figure that the size of the ensemble does not have a clear impact on the improvement. The spread, on the other hand, has a well-defined direct relationship with all directions of improvement. However, it is important to understand the meaning of the spread in terms of total and specific uncertainty in the modeling system. Considerations in the next step of the selection process helps to elaborate on this issue.

Figure 4. *Distribution of the relative improvement in event-based error statistics for the first guess: a) Runoff error; b) Peak magnitude error; c) Peak time error; and d) Probability of exceedance for a), b) and c). The light gray bars (far left) indicate cases where the assimilation process deteriorated the forecast. The magnitude of the deterioration was not considered.*

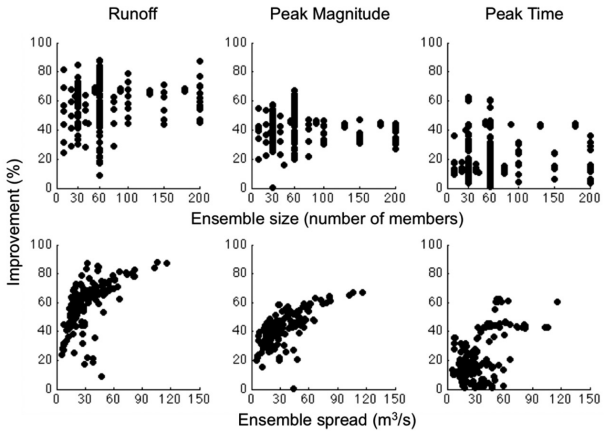


The best ensemble in each of the directions of improvement was identified from the portion of ensembles that reduced all three types of error (i.e. *Ep*, *Er* and *Et*). The size was fixed to 60 while the rainfall error was set to a maximum of 50% and parametric perturbations were limited to single calibration runs (i.e. no combination among calibration runs). In this selection approach, the modeling error was characterized by individually constraining each major source of uncertainty (i.e. model parameters and forcing) to ranges that are believed to be reasonable. The first two rows of Table 7 present the selected 60-member ensembles. Although the improvement rates are high, Fig. 5 shows that better performance of the filter is achieved with higher values of spread than the spread represented by the selected ensembles. The relatively low spread is most probably caused by deficiencies in the methodology for generating parametric and rainfall perturbations. Ensembles with higher spread, and thus higher improvement rates, were those generated by combining parametric perturbations from different calibration runs (e.g. "Avg Wet + H2YRP") and by allowing the rainfall error to be higher than 50% in the uniform random multiplier model. In this way, deficiencies in one component are compensated in other parts, and so the uncertainties from the modeling system are not treated as separate sources but as a whole, which results in better characterization of the total uncertainty. The best from this group of ensembles is presented in the last row of Table 7 and is used for the forecast evaluation in the next section.

Table 7. *Best 60-member ensembles from sensitivity analysis. The selection criteria were the highest rates of improvement.*

Criteria	Ensemble	Spread (m^3/s)	Er (%)	Ep (%)	Et (*hours*)	Ir (%)	Ip (%)	It (%)
Ir	Avg Wet 50% Rain Error	32	2	26	13	84	39	35
Ip/It	L2YRP 50% Rain Error	57	5	26	17	70	48	61
Ir/Ip	L1YRP + H1YRP 160% Rain Error	116	5	12	10	87	67	60

Figure 5. *Relationship between ensemble size and forecast improvement (upper panels) and between ensemble spread and forecast improvement (lower panels).*



Event-based forecast evaluation

To assess the forecast skill of the hydrologic model with data assimilation, four of the selected events outside the period used in the sensitivity analysis were simulated to emulate a real-time system. At time $t = i$, observations are assimilated while at times $t > i$, no assimilation of streamflow was performed while the model runs the forecast. However, quantitative precipitation forecast (QPF) products were not utilized. Instead, the same Quantitative precipitation estimate (QPE) product (i.e. MPE) was used in the forecast portion of the simulation. In this way, the system would be run under "ideal" conditions, using the best available forecast information. This also prevents the introduction of uncertainty from model inputs that has not been properly accounted for (i.e. the input error model used herein is assumed to only apply for the case of QPE). More importantly, the present study focuses on the effect of the assimilation of streamflow observations over the forecast.

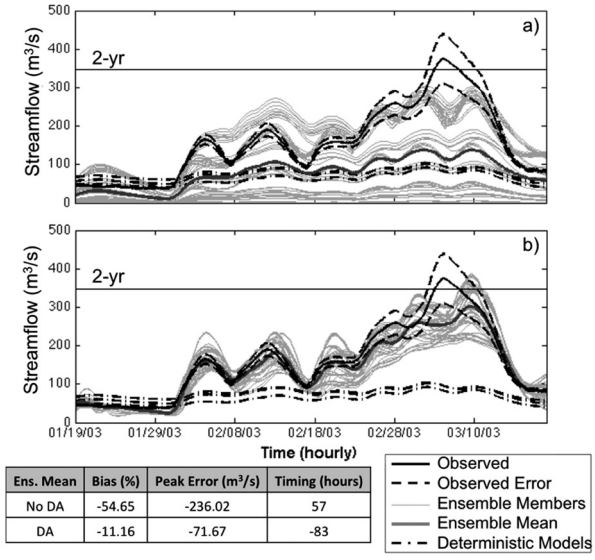
The best ensemble from the sensitivity study was employed for the assimilation and forecast runs, and the deterministic models were used as benchmarks. Figures 6, 8, 10 and 12 present the hydrographs while Figs. 7, 9, 11 and 13 contain plots of error statistics (i.e. bias, peak error, and peak time error) as a function of forecast lead time. A summary of the one-step-ahead forecast skill for all of the independent events are presented in Table 8. Overall, it is evident that the assimilating system outperformed the deterministic models on the one-step-ahead forecast of all the events. The forecast skill of the updated ensemble mean was superior for lead times of at least 3 days in terms of runoff volume and peak magnitude, and for general lead times of 1.5 days in terms of timing of peak flow. Table 8 shows that the forecast of only two of the seven events was deteriorated in regards to timing.

Table 8. *Summary of error and improvement measures for the one-step-ahead forecast of events used for validation.*

Event	Bias (%)	Peak Error (%)	Time Error (*hours*)	Ir (%)	Ip (%)	It(%)
March 1, 2003	-11.2	-19.2	-83.0	79.6	69.6	-45.6
March 20, 2003	28.1	-22.5	-6.0	60.8	66.2	89.5
April 9, 2003	-10.4	-15.4	10.0	77.9	61.2	80.0
May 23, 2003	-11.1	-10.1	-3.0	57.8	62.8	88.5

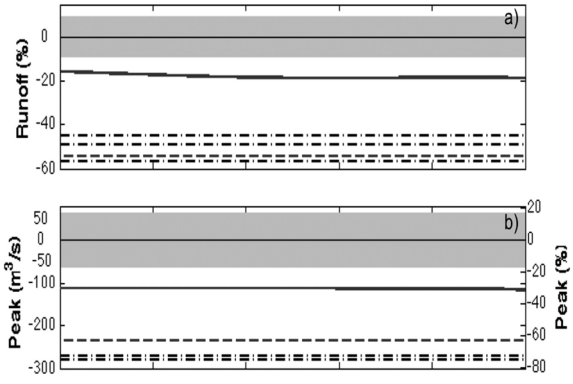
Event	Bias (%)	Peak Error (%)	Time Error (*hours*)	Ir (%)	Ip (%)	It(%)
September 17, 2003	2.0	8.7	82.0	96.5	84.1	-228.0
April 3, 2008	-4.1	-2.3	-6.0	82.6	89.8	33.3
November 10, 2009	-1.3	-11.0	0.0	99.4	92.7	100.0

Figure 6. *One-step-ahead forecast of the March 01, 2003 event. Ensemble prediction is compared to deterministic models and observed streamflow. a) Simulation without data assimilation; b) simulation with data assimilation. Error measures are presented for the ensemble mean.*



Ens. Mean	Bias (%)	Peak Error (m³/s)	Timing (hours)
No DA	-54.65	-236.02	57
DA	-11.16	-71.67	-83

Figure 7. *Forecast error measures as functions of lead-times for March 01, 2003 event. a) Runoff error in terms of percent bias; b) Peak magnitude error in cubic meters per second (m3/s); c) Peak time error in hours. Lead-times range from 0 to 120 hours (5 days).*



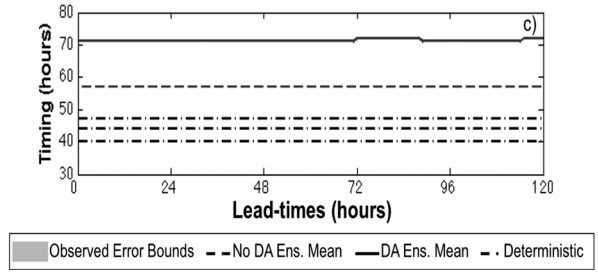
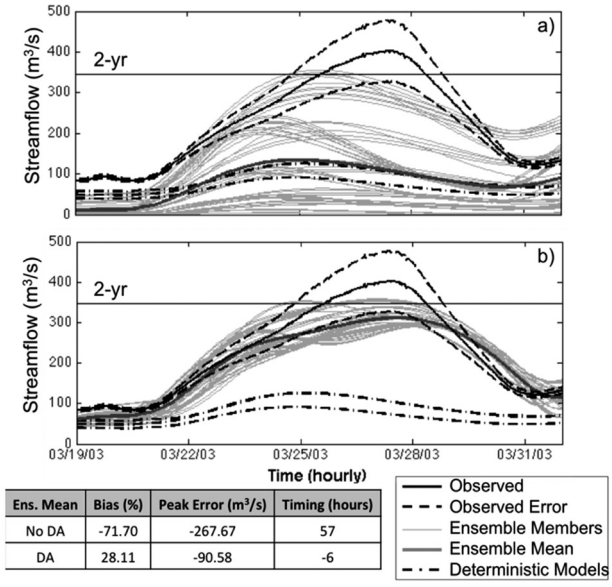


Figure 8. *Same as Figure 6 but for the March 20, 2003 event.*



Ens. Mean	Bias (%)	Peak Error (m³/s)	Timing (hours)
No DA	-71.70	-267.67	57
DA	28.11	-90.58	-6

Figure 9. *Same as Figure 7 but for the March 20, 2003 event.*

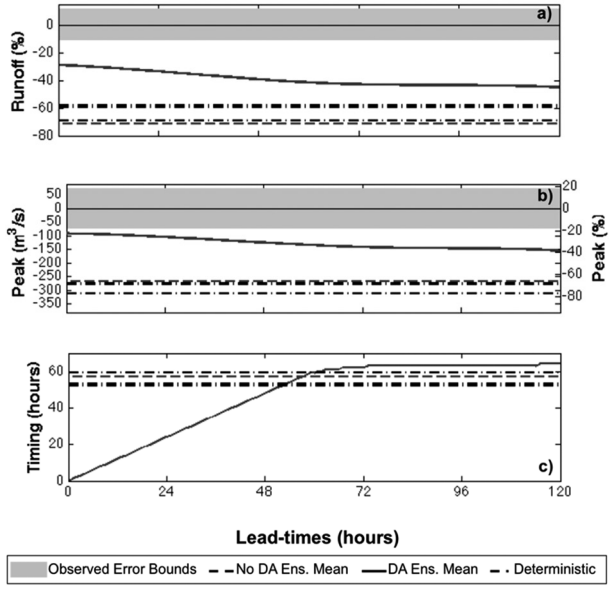
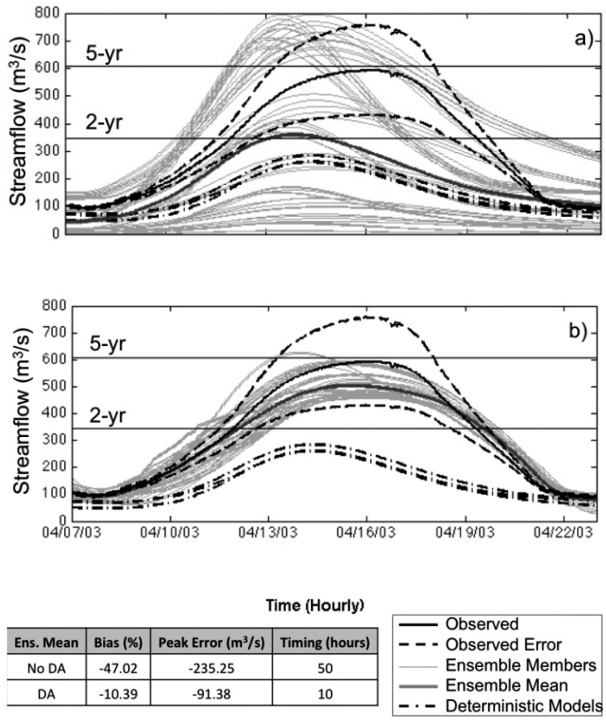


Figure 10. *Same as Figure 6 but for the April 09, 2003 event.*



Ens. Mean	Bias (%)	Peak Error (m^3/s)	Timing (hours)
No DA	-47.02	-235.25	50
DA	-10.39	-91.38	10

Figure 11. *Same as Figure 7 but for the April 09, 2003 event.*

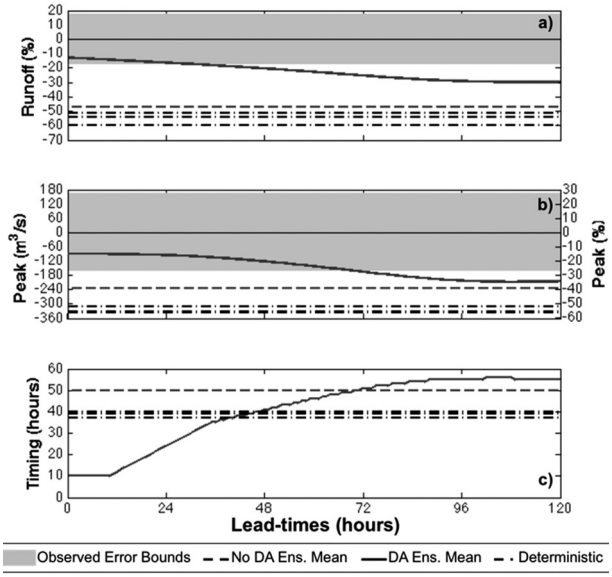
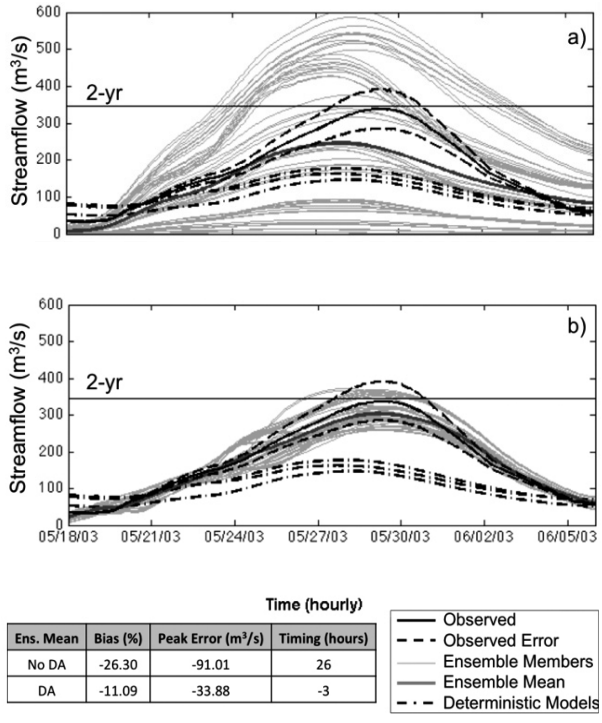
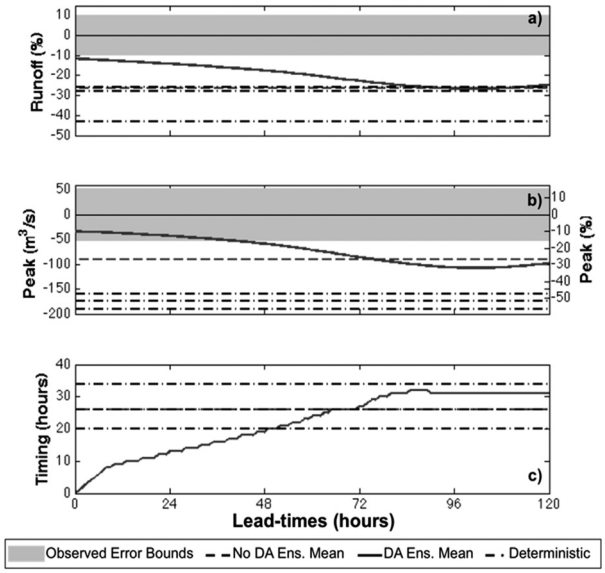


Figure 12. *Same as Figure 6 but for the May 23, 2003 event.*



Ens. Mean	Bias (%)	Peak Error (m^3/s)	Timing (hours)
No DA	-26.30	-91.01	26
DA	-11.09	-33.88	-3

Figure 13. *Same as Figure 7 but for the May 23, 2003 event.*



The performance of the assimilating system still indicates some room for improvement. In general, the runoff volume and peak magnitude of the measured streamflow were underestimated. However, the prediction lied within the uncertainty bounds of the observations for most of the events. Moreover, it is possible that in some cases, like the events in March 01 and March 20 of 2003, the errors in either the rainfall estimates or the observed streamflow are much higher than expected, as the predictions from the deterministic models suggest. In these two specific cases, the updating process actually did a reasonable job accounting for high errors and dramatically improving the forecast, most likely due to the ensemble characteristics discussed in the previous section (i.e. high spread obtained from the combination of two different parametric perturbations plus the high values of rainfall multipliers). Even though timing was identified as the major shortcoming as previously discussed, the prediction of most of the independent events were greatly improved with relatively low errors. It is important to highlight that although an analysis on ensemble characteristics that favor the performance of the filter was presented, the objective of this work was not to optimize the assimilating system, but rather to understand how sequential data assimilation improves the forecast and to identify factors affecting the updating process.

Summary and conclusions

This study explored the application and implementation of hydrological data assimilation for flood prediction using a Monte Carlo technique based on the Kalman filter. The study region was the Tar River basin in North Carolina in the United States, a basin of NOAA's CI-FLOW project where major flooding events have occurred due to heavy rainfall from tropical systems. Nine storms were selected from the past significant events inventory of the NWS forecast office at Raleigh in North Carolina, plus three additional events with peak flow values exceeding the 2-yr recurrence interval. A simple but widely used lumped, conceptual rainfall-runoff model was employed to simulate streamflow at the basin outlet. The model was calibrated with a novel global optimization algorithm called DREAM, which was also used to generate parametric perturbations for ensemble prediction. Perturbations on rainfall were produced using random uniform scalars to account for volume errors. Errors in streamflow observations were characterized through an innovative yet simple empirical model.

A sensitivity analysis was conducted on 484 different ensembles varying in size, spread, and the perturbations used to produce them (i.e. parameter perturbations, rainfall perturbation, and combination). Event-based aggregated statistics were computed for each ensemble from simulations of five of the selected events, to determine the overall direction of improvement and relationships between ensemble attributes (i.e. size and spread), and the performance of the assimilation technique. Additionally, specific events were simulated emulating a real-time forecasting system to evaluate the skill of the filter as a function of different lead times.

The main objective of this particular work was to analyze some of the benefits and drawbacks of sequential state updating through the EnKF for flood forecasting, as well as some of the mechanisms that control its performance. Specifically, the improvement on three forecast components were analyzed from the first guess and from forecasts of different lengths (i.e. lead times): a) prediction of the total amount of water in the flooding event, represented by the runoff volume error; b) prediction of how rare or extreme an event is, represented by the peak magnitude error; and c) prediction's timing, represented by the peak time error. The most important conclusions and remarks derived from this study can be summarized as follows:

Data assimilation improved the forecast skill of the hydrologic model for the events considered. Overall, the reduction of volume errors was more effective than that of peak magnitude and peak time errors.

The timing of the prediction was found to be more prone to deterioration than improvement. This might be a consequence of the fact that the difference between simulated and observed streamflow is the only information (i.e. innovation) considered in the assimilation process at every cycle.

Strong association was found between the spread of the ensembles and the performance of the filter in terms of the rates of improvement. The opposite was found regarding the size of the ensemble.

Combination of different parametric perturbations and exceedingly high error bounds for the rainfall perturbations resulted in higher spread of the ensembles, and

thus better performance of the updating system, most probably due to compensation effects on error characterization deficiencies.

Ensemble-based models with data assimilation outperformed the best deterministic models found in this study for lead times of at least 3 days for runoff volume and peak magnitude, and 1.5 days for timing.

Although skillful predictions of streamflow were obtained with the ensembles tested in this study, it is clear that further improvement can still be achieved. More research towards optimization of ensemble prediction and implementation of data assimilation is necessary to improve hydrologic forecasts. Future efforts will include an in-depth analysis of the error statistics during the assimilation process and an assessment of the value of using more sophisticated and complex error models for observations and forcing data, especially for the case of QPF.

References

[1] Adhikari, P., Y. Hong, K. Douglas, D. Kirschbaum, J. Gourley, R. Adler, and G. Robert Brakenridge (2010). "A digitized global flood inventory (1998–2008): compilation and preliminary results." *Natural Hazards 55* (2):405-422.

[2] Aubert, D., C. Loumagne, and L. Oudin (2003). "Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model." *Journal of Hydrology* 280 (1-4):145-161.

[3] Briedenbach, J. P., and J. S. Bradberry (2001). "Multisensor precipitation estimates produced by the National Weather Service River Forecast Centers for hydrologic applications." Paper read at 2001 Georgia Water Resources Conference, at Institute of Ecology, University of Georgia, Athens, Ga.

[4] Burnash, R. J. C., R. L. Ferral, and R. A. McGuire. 1973. A general streamflow simulation system - Conceptual modeling for digital computers. *In Report by the Joint Federal State River Forecasts Center, Sacramento, California.*

[5] Carpenter, T. M., J. A. Sperfslage, K. P. Georgakakos, T. Sweeney, and D. Fread (1999). "National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems." *Journal of Hydrology* 224:21-44.

[6] Clark, M., D. Rupp, R. Woods, X. Zheng, R. Ibbitt, A. Slater, J. Schmidt, and M. Uddstrom (2008). "Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model." *Advances in Water Resources* 31 (10):1309-1324.

[7] Di Baldassarre, G., and A. Montanari (2009). "Uncertainty in river discharge observations: a quantitative analysis." *Hydrol. Earth Syst. Sci.* 13 (6):913-921.

[8] Evensen, G. (2003). "The Ensemble Kalman Filter: theoretical formulation and practical implementation." *Ocean Dynamics* 53 (4):343-367.

[9] Fulton, R. (2002). "Activities to improve WSR-88D Radar Rainfall Estimation in the National Weather Service." Paper read at 2nd Federal Interagency Hydrologic Modeling Conference, July, 2002, at Las Vegas, Nevada.

[10] Gourley, J. J., Z. L. Flamig, Y. Hong, and K. W. Howard (2011). "Evaluation of past, present, and future tools for radar-based flash flood prediction." *Hydro. Sci. J.* (in review).

[11] Hamill, T. M. 2006. Ensemble-based atmospheric data assimilation. In *Predictability of weather and climate,* edited by T. Palmer and R. Hagedorn. Cambridge ; New York: Cambridge University Press.

[12] Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich (1998). "Integration of soil moisture remote sensing and hydrologic modeling using data assimilation." *Water Resour. Res.* 34 (12):3405-3420.

[13] Huang, C., X. Lin., L. Lu, and J. Gu (2008). "Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter." *Remote Sensing of Environment* 112 (3):888-900.

[14] Jazwinski, A. (1970). "Stochastic Processes and Filtering." Dover Publications.

[15] Kalnay, E. (2003). "Atmospheric modeling, data assimilation, and predictability." Cambridge Univ Pr.

[16] Kitanidis, P. K., and R. L. Bras (1980). "Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty." *Water Resour. Res.* 16 (6):1025-1033.

[17] Komma, J., G. Bloschl, and C. Reszler (2008). "Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting." *Journal of Hydrology* 357 (3-4):228-242.

[18] Krause, P., D. P. Boyle, and F. Base (2005). "Comparison of different efficiency criteria for hydrological model assessment." *Advances in Geosciences* 5:89-97.

[19] McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010). "Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions." H*ydrological Processes* 24:1270–1284.

[20] Moore, R. J. (1985). "The probability-distributed principle and runoff production at point and basin scales." *Hydrological Sciences* 30 (2):273-297.

[21] Moradkhani, H. (2008). "Hydrologic Remote Sensing and Land Surface Data Assimilation." *Sensors* 8 (5):2986-3004.

[22] Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser (2005). "Dual state?parameter estimation of hydrological models using ensemble Kalman filter." *Advances in Water Resources* 28 (2):135-147.

[23] Neal, J., P. Atkinson, and C. Hutton (2007). "Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements." *Journal of Hydrology* 336 (3-4):401-415.

[24] Petersen-Overleir, A. (2004). "Accounting for heteroscedasticity in rating curve estimates." *Journal of Hydrology* 292 (1-4):173-181.

[25] RAH-Web-Team. (2011). "Past Events." Event inventory of the National Weather Service (NWS) forecast office at Raleigh, <http://www.erh.noaa.gov/rah/events/> May 15, 2011

[26] Reichle, R. H., D. Entekhabi, and D. B. McLaughlin (2001). "Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach." *Water Resour. Res.* 37 (9):2353-2364.

[27] Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002). "Hydrologic data assimilation with the ensemble Kalman filter." *Monthly Weather Review* 130 (1):103-114.

[28] Salamon, P., and L. Feyen (2009). "Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter." *Journal of Hydrology* 376 (3-4):428-442.

[29] Sauer, V. B., and R. W. Meyer. 1992. Determination of Error in Individual Discharge Measurments. Norcross, Georgia: U.S. Geological Survey.

[30] Seo, D.-J., V. Koren, and N. Cajina (2003). "Real-Time Variational Assimilation of Hydrologic and Hydrometeorological Data into Operational Hydrologic Forecasting." *Journal of Hydrometeorology* 4 (3):627-641.

[31] Smith, M., D. Seo, V. Koren, S. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004). "The distributed model intercomparison project (DMIP): motivation and experiment design." *Journal of Hydrology* 298 (1-4):4-26.

[32] Smith, P. J., K. J. Beven, and J. A. Tawn (2008). "Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach." *Water Resources Research* 44 (1).

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

[34] USGS. (2011). "IDA." Instantaneous Data Archive - IDA, <http://ida.water.usgs.gov/ida/> June, 15, 2010

[35] USGS, and USAID. (2011). "USGS FEWS NET Data Portal." USGS FEWS NET Data Portal, <http://earlywarning.usgs.gov/fews/> July, 01, 2010

[36] Van Cooten, S., K. E. Kelleher, K. Howard, J. Zhang, J. J. Gourley, J. S. Kain, K. Nemunaitis-Monroe, Z. Flamig, H. Moser, A. Arthur, C. Langston, R. Kolar, Y. Hong, K. Dresback, E. Tromble, H. Vergara, R. A. Luettich, B. Blanton, H. Lander, K. Galluppi, J. P. Losego, C. A. Blain, J. Thigpen, K. Mosher, D. Figurskey, M. Moneypenny, J. Blaes, J. Orrock, R. Bandy, C. Goodall, J. G. W. Kelley, J. Greenlaw, M. Wengren, D. Eslinger, J. Payne, G. Olmi, J. Feldt, J. Schmidt, T. Hamill, R. Bacon, R. Stickney, and L. Spence (2011). "The CI-FLOW Project: A System

for Total Water Level Prediction from the Summit to the Sea." *Bulletin of the American Meteorological Society* 92 (11):1427-1442.

[37] Vrugt, J. A., and W. Bouten (2003). "Toward improved identifiability of hydrologic model parameters: The information content of experimental data." *Water Resources Research* 38 (12).

[38] Vrugt, J. A., C. J. F. Braak, H. V. Gupta, and B. A. Robinson (2008). "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" *Stochastic Environmental Research and Risk Assessment* 23 (7):1011-1026.

[39] Vrugt, J. A., C. G. H. Diks, H. V. 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 Resources Research* 41 (1).

[40] Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian. 2003. A Shuffled Complex Evolution Metropolis Algorithm for Estimating Posterior Distribution of Watershed Model Parameters. In *Calibration of Watershed Models*, edited by Q. Duan, S. Sorooshian, H. V. Gupta, A. N. Rousseau and R. Turcotte. Washington, DC: American Geophysical Union.

[41] Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten (2006). "Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model." *Journal of Hydrology* 325 (1-4):288-307.

[42] Vrugt, J. A., and B. A. Robinson (2007). "Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging." *Water Resources Research* 43 (1)

[43] Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009). "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling." *International Journal of Nonlinear Sciences and Numerical Simulation* 10 (3):273-290.

[44] Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian. 2001. A framework for development and application of hydrological models. Copernicus GmbH.

[45] Wang, X. 2009. *Advance topics in data assimilation: Ensemble Kalman filter techniques*. Norman, OK: Graduate level class METR 6803, The University of Oklahoma.

[46] Whitaker, J. S., and T. M. Hamill (2002). "Ensemble data assimilation without perturbed observations." *Monthly Weather Review* 130 (7):1913-1925.

[47] Willmott, C. J., and K. Matsuura (2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." *Climate Research* 30 (1):79-82.

Los Autores

Humberto J. Vergara

Received the B.Sc. degree in environmental engineering from El Bosque University, Colombia, and the M.Sc. degree in water resources engineering from the University of Oklahoma, Norman, OK, USA. He is currently working toward the Ph.D. degree in the Department of Civil Engineering and Environmental Science at the University of Oklahoma, Norman. He is currently a Graduate Research Assistant at the Hydrometeorology and Remote Sensing (HyDROS; hydro.ou.edu) Laboratory and the National Severe Storms Laboratory (NSSL; http://www.nssl.noaa.gov) in Norman, Oklahoma. His primary field of study is hydrological modeling for flood forecasting. He focuses on model development, ensemble forecasting and data assimilation.

Yang Hong

Received the B.S. and M.S. degrees in geosciences and environmental sciences from Peking University, Beijing, China, and the Ph.D. degree, major in hydrology and water resources and minor in remote sensing and spatial analysis, from the University of Arizona, Tucson, AZ, USA. Following a postdoctoral appointment in the Center for Hydrometeorology and Remote Sensing, University of California, Irvine, CA, USA, he joined the National Aeronautics and Space Administration Goddard Space Flight Center, Greenbelt, MD, USA, in 2005. He is currently an Associate Professor with the School of Civil Engineering and Environmental Sciences and the School of Meteorology, University of Oklahoma, Norman, OK, USA, where he is also directing the Remote Sensing Hydrology research group (http://hydro.ou.edu). He also serves as the Co-director of the Water Technologies for Emerging Regions Center (http://water.ou.edu) and an affiliated Faculty Member with the Atmospheric Radar Research Center (http://arrc.ou.edu). He has served in the editorial boards of the International Journal of Remote Sensing, the Natural Hazards journal, and the Landslides journal. His primary research interests are in remote-sensing retrieval and validation, hydrology and water resources, natural hazard prediction, land surface modeling, and data assimilation systems for water resource planning under changing climate. Dr. Hong is currently the American Geophysical Union Precipitation Committee Chair.

Jonathan J. Gourley

Received the B.S. and M.S. degrees in meteorology with a minor in hydrology and the Ph.D. degree in civil engineering and environmental science from the University of Oklahoma, Norman, OK, USA. He is currently a Research Hydrometeorologist with NOAA's National Severe Storms Laboratory, is an affliate Associate Professor with the School of Meteorology, University of Oklahoma, and Director of the National Weather Center's seminar series. His research focuses on rainfall observations from remote sensing platforms with an emphasis on ground-based radars and implementing these high-resolution observations into hydrologic models. He completed a postdoctoral study with researchers in Paris, France, to demonstrate the capabilities of dual-polarimetric radar in improving data quality, microphysical retrievals, and precipitation estimation. MeteoFrance has subsequently upgraded several of their operational radars with polarimetric technology. Dr. Gourley received the Department of Commerce Silver Medal Award in 1999 "For developing an important prototype Warning Decision Support System for weather forecasters and making significant enhancements to the NEXRAD system, resulting in more timely and reliable warnings." He also received an Honorable Mention in 2004 from the Universities Council on Water Resources Dissertation Awards Committee.