A Novel Modeling Framework for Computationally Efficient and Accurate Real-Time Ensemble Flood Forecasting With Uncertainty Quantification

Vinh Ngoc Tran[1] , M. Chase Dwelle[2], Khachik Sargsyan[3] , Valeriy Y. Ivanov[2] , and Jongho Kim[1]

[1]School of Civil and Environmental Engineering, University of Ulsan, South Korea, [2]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA, [3]Sandia National Laboratories, Livermore, CA, USA

Abstract A novel modeling framework that simultaneously improves accuracy, predictability, and computational efficiency is presented. It embraces the benefits of three modeling techniques integrated together for the first time: surrogate modeling, parameter inference, and data assimilation. The use of polynomial chaos expansion (PCE) surrogates significantly decreases computational time. Parameter inference allows for model faster convergence, reduced uncertainty, and superior accuracy of simulated results. Ensemble Kalman filters assimilate errors that occur during forecasting. To examine the applicability and effectiveness of the integrated framework, we developed 18 approaches according to how surrogate models are constructed, what type of parameter distributions are used as model inputs, and whether model parameters are updated during the data assimilation procedure. We conclude that (1) PCE must be built over various forcing and flow conditions, and in contrast to previous studies, it does not need to be rebuilt at each time step; (2) model parameter specification that relies on constrained, posterior information of parameters (so-called *Selected* specification) can significantly improve forecasting performance and reduce uncertainty bounds compared to *Random* specification using prior information of parameters; and (3) no substantial differences in results exist between single and dual ensemble Kalman filters, but the latter better simulates flood peaks. The use of PCE effectively compensates for the computational load added by the parameter inference and data assimilation (up to ~80 times faster). Therefore, the presented approach contributes to a shift in modeling paradigm arguing that complex, high-fidelity hydrologic and hydraulic models should be increasingly adopted for real-time and ensemble flood forecasting.

1. Introduction

Real-time forecasting is an important component of flood risk management and mitigation but is subject to multiple uncertainties caused by meteorological inputs, initial states, model structures, and model parameters (Ajami et al., 2007; Beven, 1989; Mockler et al., 2016; Moradkhani & Sorooshian, 2008). Due to the complexities of natural phenomena represented by equifinality (Beven, 2006; Beven & Freer, 2001), hysteresis (Fatichi et al., 2015; Ivanov et al., 2010; Wei & Dewoolkar, 2006), nonuniqueness (Beven, 2000; McKenna et al., 2003; Kim & Ivanov, 2014; Kim, Dwelle, et al., 2016), nonlinearity (Kim & Ivanov, 2015; Kitanidis & Bras, 1980; Xie & Zhang, 2010), and internal variability (Kim et al., 2016a, 2016b; Kim et al., 2018; Lafaysse et al., 2014; Mondal & Mujumdar, 2012; Nikiema & Laprise, 2011), perfect predictions using numerical models are infeasible. The problem exacerbates, if one attempts to simulate constitutive models derived from empirical or phenomenological observations rather than basic conservation laws of physics that would also require embracing a large number of parameters. Forecasting systems must therefore rely on approaches with intrinsic tools to quantify and reduce associated uncertainties and allow end users to make informed decisions (Todini, 1999, 2004).

Forecasts made with sufficient lead time can mitigate flood damages considerably. Results should therefore be provided within a predetermined time horizon and accurate enough to promote community confidence in actions taken for emergency preparedness (Associated Programme on Flood Management (APFM), 2013; Todini, 2004). Extensive efforts have been devoted to enhance forecast accuracy, predictability, and efficiency in real time with uncertainty quantification (Table 1). However, simultaneous improvement of predictive

Table 1
Literature Review of Applications Involving Real-Time, Ensemble Streamflow Forecasting

Study	Deterministic model	Surrogate model	Parameter specification	State initialization	DA	A
Zhang et al. (2018)	Xinanjiang	-	Optimization	Warm-up	Dual	A12
Abbaszadeh et al. (2018)	SAC-SMA	-	Random	Arbitrary	Dual	A12
Wang et al. (2018)	HyMOD	PCE	NA	Warm-up	Dual	A6
Davison et al. (2017)	MESH	-	Random	NA	Dual	A3
Thiboult et al. [2016]	Multimodels	-	NA	Warm-up	Single	A2
Fan et al. [2016]	HyMOD	PCE	Random	NA	Dual	A6
Zahmatkesh et al. (2015)	HyMOD, HBV, SWMM	-	Bayesian inference	Warm-up	None	A10
Li et al. (2014)	GR4H	-	Optimization	NA	Dual	A12
DeChant and Moradkhani (2014)	VIC	-	NA	Warm-up	Dual	A3
Xie and Zhang (2013)	SWAT	-	Random	Warm-up	Dual	A3
Chen et al. (2013)	HyMOD	-	Bayesian inference	NA	Single	A11
Moradkhani et al. (2012)	HyMOD	-	Random	Warm-up	Dual	A3
He et al. (2012)	SNOW17+ SAC-SMA	-	Bayesian inference	Warm-up	Single	A11
Mendoza et al. (2012)	TopNet	-	Manual calibration	Warm-up	Single	A11
Clark et al. (2008)	TopNet	-	Bayesian inference	Warm-up	Single	A11
Ajami et al. (2007)	HyMOD, SWB	-	Bayesian inference	Warm-up	None	A10
Weerts and El Serafy (2006)	HBV-96	-	NA	NA	Single	A2
Vrugt et al. (2005)	HyMOD	-	Random	Arbitrary	Dual	A3
Moradkhani et al. [2005]	HyMOD	-	Random	Arbitrary	Dual	A3
Madsen and Skotner [2005]	Mike 11	-	Optimization	Warm-up	Single	A11
Beven and Freer (2001)	TOPMODEL	-	Bayesian inference	Warm-up	Dual	A12

Note. The last column corresponds to the taxonomy of the predictive approach (A) that we define in Table 3. "Warm-up" methods have a warm-up period, while "Arbitrary" methods do not. NA = Not Available.

accuracy and efficiency, while evaluating effectiveness, remains a major challenge (Cintra & Velho, 2018; Liu et al., 2012).

For the purpose of enhancing model accuracy in real-time flood forecasting where no information of model states and parameters is available, data assimilation (DA) has been proven useful. Due to the nature of forecasting, the effect of future unknowns (model parameters and states) on flood prediction will change over time. In addition, uncertainty can be amplified not only by the features of the model itself but also by errors in forcing data and observations. Therefore, model adjustment for the forecasting period may be necessary (Moradkhani, Sorooshian, et al., 2005; Young, 2002). Several assimilation methods have been developed using Kalman or particle filters and optimization or inference techniques such as the back-fitting algorithm (Zhang et al., 2018), shuffled complex evolution algorithm (Li et al., 2014), shuffled complex evolution metropolis (Vrugt et al., 2005), generalized likelihood uncertainty estimation (GLUE) (Beven & Freer, 2001), and sequential Bayesian combination (DeChant & Moradkhani, 2014). Due to the higher computational requirements of the latter techniques, filter-type approaches have attracted attention as assimilation tools (Gharamti et al., 2013; Moradkhani & Sorooshian, 2008).

Currently, the ensemble Kalman filter (EnKF) (Evensen, 1994) and its modifications (e.g., ensemble Kalman smoothers, ensemble square root filters, and gain function) are the most commonly used techniques in the hydrology community (Table 1), despite the issue of slow convergence caused by intrinsic assumptions, especially for domains with complexities (Moradkhani et al., 2012; Moradkhani, Hsu, et al., 2005; Wang et al., 2017; Weerts & El Serafy, 2006). Recent studies have suggested that particle filtering (Arulampalam et al., 2002) is an alternative method to resolve the inclusion of unrealistic Gaussian assumptions in the EnKF. The particle filtering method has more advantages than EnKF in reducing numerical instability by providing particle weights and using non-Gaussian state-space models (Liu et al., 2012). However, this method is computationally more expensive as it generally requires more ensemble members (Liu et al., 2012; Moradkhani, Hsu, et al., 2005).

When assimilating data, model parameter specification and state initialization may play a crucial role, especially for short-range forecasting (Houtekamer & Zhang, 2016). Generally, ensemble initialization of model states and parameters for the forecasting period can be generated approximately, for example, using a random selection from uniform distributions for parameters and setting up the initial state values as an

arbitrary number (e.g., zero) at the beginning of the forecasting period (Abbaszadeh et al., 2018; Davison et al., 2017; DeChant & Moradkhani, 2014; Moradkhani, Hsu, et al., 2005; Moradkhani et al., 2012; Vrugt et al., 2005; Xie & Zhang, 2013). Alternatively, the ensemble can be generated more carefully, for example, specifying parameters from relevant distributions (Ajami et al., 2007; Beven & Freer, 2001; Chen et al., 2013; Clark et al., 2008; He et al., 2012; Madsen & Skotner, 2005; Mendoza et al., 2012; Zahmatkesh et al., 2015) and using a warm-up technique for states (Ajami et al., 2007; DeChant & Moradkhani, 2014; He et al., 2012; Mendoza et al., 2012; Wang et al., 2018), as summarized in Table 1.

The assimilation techniques described above generally require a large number of model evaluations to update parameter and state values and present predictive uncertainties, leading to computational challenges (Vrugt, 2016; Vrugt et al., 2008; Zhang et al., 2017), even with the benefit of parallel computation with multiple processors (Cintra & Velho, 2018). Because keeping calculation time to a minimum is a key element for timely flood warnings and responding to emergency situations (Ballio & Guadagnini, 2004; Sene, 2008), it is necessary to find alternatives that significantly increase forecast lead time. Surrogate modeling can address this challenge by substituting computationally intensive models with computationally efficient metamodels, such as the polynomial chaos expansion (PCE). Through the expansion of orthogonal polynomials, approximate functions can be constructed and applied to hydrologic models. Recent studies have used PCE to perform robust uncertainty assessment of diverse hydrologic problems (Dwelle et al., 2019; Fan et al., 2014, 2016; Sochala & Le Maître, 2013; Wang et al., 2015, 2017, 2018; Wu et al., 2014) rather than running deterministic models. However, few studies have tested its effectiveness in a setting of real-time flood forecasting (Fan et al., 2016; Wang et al., 2015, 2017, 2018).

To fill the above gaps, we propose a novel integrated modeling framework that improves accuracy, predictability, and efficiency of real-time flood forecasting. Eighteen approaches to the framework are presented, combining ways of constructing the surrogate models, specifying model parameters and states, and assimilating newly observed data. This study investigates (i) the effects of building methods of the PCE model and its capacity for real-time flood forecasting, (ii) the effects of specifying methods on predictive performance, (iii) the effects of single- and dual-assimilation techniques, and (iv) the computational time of the proposed approaches.

2. Methodologies and Frameworks

2.1. Methodologies

2.1.1. Deterministic Rainfall-Runoff Model: NAM

To construct a surrogate model, simulate streamflow, quantify uncertainty, and assimilate observed data, a Nedbør-Afstrømnings model (NAM) (Nielsen & Hansen, 1973) is employed. As one of the widely used deterministic, lumped models, it is considered useful and flexible and has been applied to many catchments (Butts et al., 2004; Madsen, 2000; Mockler et al., 2016; O'Brien et al., 2013; Thompson et al., 2004). Specifically, its design assumes three different and mutually integrated storages representing a surface zone, lower zone, and routing components that simulate overland flow, interflow, and base flow, respectively. The model requires two input forcing variables (M_I) of spatially averaged precipitation and evapotranspiration, five model states ($M_S = 5$), and nine model parameter values ($M_P = 9$) listed in Table 2 (Danish Hydraulic Institute, 2014). The latter states and parameters control the amount of water content and the rates of release from the conceptualized storage compartments of the model. Because evapotranspiration is assumed to be negligible during the rainy season with flooding events, the number of inputs used in this study is 1 ($M_I = 1$).

2.1.2. Surrogate Model: Polynomial Chaos Expansion

PCE (Ghanem & Spanos, 1991; Wiener, 1938) can build a surrogate model ($\mathscr{M}^{PCE}$) for any (deterministic rainfall-runoff) model ($\mathscr{M}$) through the expansions of orthogonal polynomials. This enables a polynomial approximation of the model through its deterministic input/output relationship. The form of a PCE model approximating a model output (e.g., streamflow y_t) as a function of model parameters θ_t is given as follows:

$$y_t = \mathscr{M}(\theta_t) \approx \mathscr{M}^{PCE_t}(\theta_t) \tag{1}$$

Note that the surrogate model ($\mathscr{M}^{PCE}$) in equation (1) has the subscript of t, indicating that the surrogate model is a collection of PCEs constructed at each time step of interest. Also, only the parameter θ_t (this

Table 2
Description of the NAM Model States and Parameters

		Unit	Description	Lower Bound	Upper Bound
States	U	mm	Water content in surface storage	0	35
	L	mm	Water content in lower zone/root storage	0	400
	OF	m³/s	Overland flow	0	+∞
	IF	m³/s	Inter flow	0	+∞
	BF	m³/s	Base flow	0	+∞
Parameters	Um	mm	Maximum water content in surface storage	5	35
	Lm	mm	Maximum water content in lower zone/root storage	50	400
	CQOF	[-]	Overland flow coefficient	0	1
	CKIF	hr	Interflow drainage constant	200	2,000
	TOF	[-]	Overland flow threshold	0	0.9
	TIF	[-]	Interflow threshold	0	0.9
	TG	[-]	Groundwater recharge threshold	0	0.9
	CK12	hr	Time constant for routing interflow and overland flow	3	72
	CKBF	hr	Time constant for base flow	500	5,000

includes a subscript of t as well) is chosen as an input variable during PCE construction, and other forcing or state inputs required to simulate hydrologic models are held constant (Dwelle et al., 2019; Fan et al., 2016; Meng & Li, 2018; Sochala & Le Maître, 2013; Tran & Kim, 2019; Wang et al., 2018). This mathematical formulation conveys that PCE should be built separately for each time step at which a meteorological condition or model state is updated.

Unlike previous studies based on equation (1), this study constructs the surrogate PCE model with equation (2), which has three characteristics: (i) the model input consists of meteorological data, model states, and model parameters; (ii) model parameters do not change over time, which is different from equation (1); and (iii) there is no need to constantly create the PCE model over time (which is the most important practical feature). The single PCE model represents streamflow phenomena over the entire calibration period during which the PCE model was generated. Specifically, ensemble model output (Y_t) at each time step, including streamflow (y_t) and states (x_t), can be written as a function of model inputs (X_t), including states (x_{t-1}), climate data (u_t), and time-invariant parameters (θ):

$$Y_t = \mathcal{M}(X_t) \approx \mathcal{M}^{PCE}(X_t) = \sum_{\alpha \in A} \varepsilon_\alpha \Psi_\alpha(X_t) \tag{2}$$

$$Y_t = [y_t \; x_t], \quad X_t = [x_{t-1} \; u_t \; \theta] \tag{3}$$

where ε_α represents the PCE coefficients to be determined for all multiindices, $\alpha = \{\alpha_1, ..., \alpha_{M_X}\}$ belonging to a set of candidate polynomials $A, \{\alpha \in A\}$. $\Psi_\alpha(X_t)$ represents the multivariate polynomials corresponding to the given input X_t. The polynomials are constructed as the product of univariate orthonormal polynomials:

$$\Psi_\alpha(X_t) = \prod_{j=1}^{M_X} \Psi_{\alpha_j}^{(j)}\left(X_t^j\right) \tag{4}$$

where $\Psi_{\alpha_j}^{(j)}$ is the univariate orthonormal polynomials of the jth variables of the degree α_j. In general, the size of X_t, M_X, is equal to the summation of the number of parameters, states, and forcing inputs of the deterministic NAM model (i.e., $M_X = M_P + M_S + M_I$). Set A is determined by M_X and the polynomial degree, p of the PCE model as follows:

$$A = A^{M_X, p} = \left\{\alpha \in \mathbb{N}^{M_X} : |\alpha| \le p\right\}, card \, A^{M_X, p} = \binom{M_X + p}{p} \tag{5}$$

Various polynomial bases (e.g., uniform, Gaussian, beta, and gamma) can be chosen for $\Psi_\alpha(X_t)$ from the Weiner-Askey scheme, depending on the probabilistic characteristics of model input variables X_t (Xiu & Karniadakis, 2002).

Given the set of multivariate orthonormal polynomials ($\Psi_\alpha(\boldsymbol{X}_t)$), the next step is to compute the PCE coefficients (ε_α), which are influenced by the number of model evaluations (called the experimental design, N) and the polynomial degree, p (Blatman & Sudret, 2010, 2011). Increasing these numbers requires significant computational resources and the requirement is higher. When using the projection method (Ghiocel & Ghanem, 2002; Le Maître et al., 2002), one of the methods employed to compute PCE coefficients, N is determined based on p and the size of $\boldsymbol{X}_t$, M_X as $N = (p + 1)^{M_X}$. This number is large enough that it takes a considerable time to construct the surrogate model (Blatman & Sudret, 2011; Tran & Kim, 2019). Reducing N is desirable as it lessens the computational cost. For the least squares regression method adopted in this study, N is not defined a priori and is provided by the researcher, which can enable a significant decrease for the value of N (Berveiller et al., 2006; Blatman & Sudret, 2010; Sudret, 2008). Also, p can be determined by the complexity of model outputs and the subjectivity of researcher, with many studies choosing values of 2 or 3 (Fan et al., 2014; Sochala & Le Maître, 2013; Wang et al., 2015, 2017). Investigating the effects of varying values of N and p on the PCE model allows for determination of the optimal values of the both parameters. According to the approach by Blatman and Sudret (2010), a metric of the leave-one-out (LOO) cross-validation error in equation (A5) can illustrate the performance of the PCE model. A brief overview of the construction of the PCE surrogate model is detailed in Appendix A.

2.1.3. Parameter Inference: GLUE

GLUE (Beven & Binley, 1992) refers to a series of procedures for inferring parameter posterior distributions and quantifying the associated uncertainties. The objective of GLUE is to select "behavioral" model runs based on the threshold values of likelihood functions with observations, among a large number of runs simulated with random combinations of parameter values. The latter parameter's values can be sampled randomly from the prior distributions of each parameter (constrained in this study with upper and lower bounds of Table 2) using Monte Carlo or Latin hypercube sampling (LHS). For more efficient performance, LHS was used (Helton & Davis, 2003). The likelihood functions proposed in this study are three metrics of Nash-Sutcliffe efficiency (NSE), peak error (PE), and volume error (VE) defined in Appendix B, representing the model performance with respect to the shape, peak, and volume of hydrograph, respectively. Acceptance threshold values are determined according to an approach (Tran & Kim, 2019) in which relationships between accuracy and efficiency indices are identified for their determinations. Specifically, cutoff threshold values for the likelihood functions of NSE, PE, and VE are suggested as 0.8, 5%, and 5%, respectively (Tran & Kim, 2019). The model runs (or parameters) that satisfy the modelling error within the above thresholds for all the likelihood functions are defined here as "behavioral" runs (or parameters).

2.1.4. Ensemble Data Assimilation: Single and Dual EnKFs

Among many reported techniques, the single EnKF and the dual-ensemble Karman filter (dual EnKF) are often chosen to optimally update the ensemble of model states (and parameters) of forecasting systems with real-time observations, which can be coupled with any models (Burgers et al., 1998; Evensen, 1994; Moradkhani, Sorooshian, et al., 2005; Whitaker, 2012). Specific details are provided in Appendix C.

The EnKF allows for the perturbation of observations to generate replicates of $\boldsymbol{x}_{t-1}$ and θ_t and the correction of the ensemble forecast members through an update step (equations (C10) and (C18)) (Moradkhani, Sorooshian, et al., 2005). This prevents the EnKF from a collapse in which all ensemble forecast members are likely to have similar values (Burgers et al., 1998). As shown in equation (C11), observations can be perturbed by adding stochastic noise to the observed value. This observed error in measurements is assumed to be independent and is set to be proportional to the observed values, following a Gaussian distribution with predetermined variance. In this study, we assume that the standard deviation of the observational error is 5% of streamflow observations (i.e., noise) at each time step, similar to prior studies (Bauser et al., 2018; Clark et al., 2008; DeChant & Moradkhani, 2012; Fan et al., 2016). Sensitivity analysis on the observation error is illustrated in Text S1 in the supporting information. Furthermore, overshooting or filter divergence problem in data assimilation happens when the ensemble size is small or the initial values of ensemble members are quite different from the true. To resolve this issue, we used a sufficiently large ensemble size and the posterior information of parameters to initialize the ensemble of EnKF.

2.2. New Modeling Framework

2.2.1. Obtaining Prior and Posterior Parameter Distributions of a Deterministic Model

The first preparation step of the modeling framework is to obtain the prior and posterior parameter distributions for a deterministic model. There could be various ways to handle this, but in this study the
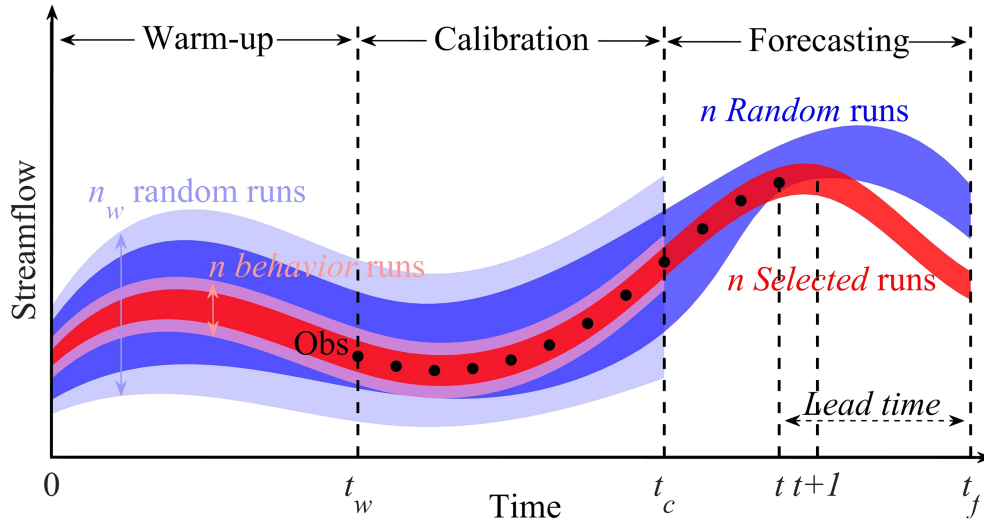
Figure 1. A schematic illustration for real-time ensemble flood forecasting, which consists of three intervals: warm-up, calibration, and forecasting periods. The light red shaded region in the warm-up and calibration periods refers to the n behavior results of GLUE that are employed to estimate posterior parameter distributions, while the light blue region refers to the n_w random results obtained from parameter sets sampled from prior (uniform) distributions to attain the n behavior runs. The construction of PCE models is carried out over the calibration period: PCE-I model is built from the N_I training samples extracted from the light blue region, while PCE-II is from the N_{II} samples from the light red region. The (dense) blue and red shaded regions correspond to the approaches using the "*Random*" (A1 to A9) and "*Selected*" (A10 to A18) parameter specifications with the same n ensemble runs, respectively.

following assumptions and methodologies are specifically applied. We first assume that each of the parameters follows a uniform distribution within specified bounds—the prior parameter distributions are simply attained by utilizing prior-known information for the bounds in Table 2. In contrast, the posterior parameter distributions are fitted to the 500 behavior parameters of GLUE—the 500 NAM behavior samples are identified as an optimal number from our previous study which has confirmed that more than the 500 parameter sets does not change the shape of the posterior distributions (Tran & Kim, 2019). For consistency, this number will be also used for making the posterior distributions of PCE-I and PCE-II in section 2.2.2.

The mathematical expression of this step is as follows. For the warm-up and calibration periods, a model $\mathcal{M}$ (NAM) can be simulated to attain behavioral runs with GLUE, that is,

$$\left[y_t^{ii} \; x_t^{ii}\right] = \mathcal{M}\left(x_{t-1}^{ii}, u_t, \theta^{ii}\right), \quad ii = 1, ..., n_w; t = 1, ..., t_c \tag{6}$$

where n_w is the number of model runs to obtain the n number of the behavioral set based on the likelihood scores estimated with the GLUE method (Tran & Kim, 2019). Among the n_w random runs (referring to the light blue shaded region in Figure 1) that are simulated by using parameter sets (θ^{ii}) sampled randomly from the prior (uniform) distributions, the only n behavior runs (referring to the light red shaded region in Figure 1) are employed for making the posterior distributions.

Reducing the effects of uncertainty by initial conditions (x_0^{ii}) is necessary for modeling. In this framework, a "warm-up" technique was employed to calibrate the deterministic model. Generally, a sufficient period of time (called the warm-up period) can be set such that the influence of the initial condition is dissipated, and the warm-up is performed before entering the calibration period. This technique produces behavioral parameter sets much faster in GLUE, compared with cases that do not use the warm-up technique.

2.2.2. Building PCE With Two Types of Experimental Design

We propose two types of approaches for constructing the PCE model, depending on how the sample collections of the experimental design ($\mathcal{X}_t$) is composed. One approach is to build a PCE model (PCE-I) by collecting the training samples that are generated from the prior parameter distributions. The other approach is (PCE-II) uses samples that are formed by *the posterior* parameter distributions. The associated mathematical expression is

$$\left[y_t^{iii} \; x_t^{iii}\right] = \mathscr{M}\left(x_{t-1}^{iii}, \boldsymbol{u}_t, \theta^{iii}\right), \quad iii = 1, ..., N; t = 1, ..., t_c \tag{7}$$

where the N_I set of $\mathscr{X}_t$ (i.e., $N = N_I$ for PCE-I) consists of model $\mathscr{M}$ simulation results calculated from parameters sampled from the prior distributions (correspond to N_I set sampled randomly from the results in the light blue shaded region over the calibration period in Figure 1) (Blatman & Sudret, 2010; Blatman & Sudret, 2011; Sudret, 2008). In contrast, the experimental design of the latter approach assumes that the N_{II} set of $\mathscr{X}_t$ (i.e., $N = N_{II}$ for PCE-II) are drawn from the more constrained, posterior parameter distributions (correspond to the light red shaded region over the calibration period in Figure 1) (Tran & Kim, 2019). All the samples were taken through LHS sampling (McKay et al., 1979).

The former approach can be implemented easily and therefore has been used more commonly in the literature (Blatman & Sudret, 2010; Blatman & Sudret, 2011; Sudret, 2008). However, for past periods in which observations exist, the second approach using a well-calibrated set of parameters is beneficial in significantly reducing computational time (Tran & Kim, 2019). It takes less time to build PCE in the second approach because less training samples (N_I is generally larger than N_{II}) are required when estimating coefficients. On the other hand, in the context of real-time forecasting when no observations have been attained, the latter approach might cause a problem. Specifically, PCE models built with a set of "good" posterior parameters sets obtained only for a certain historic period of time would not necessarily demonstrate validity for unknown prediction periods. Evaluation of the applicability of the two approaches to real-time flood forecasting will be addressed in section 4.

Once the PCE models were constructed, the same GLUE procedure is made to obtain the posterior parameter distributions of both PCE models:

$$\left[y_t^{ii} \; x_t^{ii}\right] = \mathscr{M}^{PCE}\left(x_{t-1}^{ii}, \boldsymbol{u}_t, \theta^{ii}\right), \quad ii = 1, ..., n_w; t = 1, ..., t_c \tag{8}$$

Note that the number n_w is different depending on $\boldsymbol{Model} = \{\text{NAM}, \text{PCE} - \text{I}, \text{PCE} - \text{II}\}$.

2.2.3. Specifying Model Parameters for Data Assimilation

Determining initial conditions and parameter values before assimilating real-time observations over the forecasting period is a necessary step. The mathematical expression for preparing data assimilation (forecasting) is written as follows:

$$\left[y_t^{i} \; x_t^{i}\right] = \boldsymbol{Model}\left(x_{t-1}^{i}, \boldsymbol{u}_t, \theta^{i}\right), \quad i = 1, ..., n; t = 1, ..., t_c \tag{9}$$

where the initial ensemble of states (x_0^i) is set to an arbitrary number (e.g., zero) at the beginning of simulation (i.e., $t = 0$) (Figure 1). In terms of specifying the model parameters, two types of approach are proposed. First, similarly to most previous studies of data assimilation (Davison et al., 2017; DeChant & Moradkhani, 2014; Gharamti et al., 2013; Moradkhani, Sorooshian, et al., 2005; Vrugt et al., 2005; Wang et al., 2009; Xie & Zhang, 2013), the ensemble of parameters over the periods ($0 \leq t \leq t_c$) is assumed to follow a prior distribution. That is, the n number of parameter sets are sampled from uniform distributions with predefined bounded ranges (i.e., from the results in the light blue shaded region in Figure 1). The values of parameters remain unchanged, while those of state vectors are continuously updated until the beginning of the forecasting period (i.e., $t = t_c$). This is hereafter named "*Random*" set—referring to the use of random parameter sets for running **Model** of NAM, PCE-I, and PCE-II.

An alternative way to this *Random* specification method is enabled by taking the advantage of the ability to calibrate model parameters with observed data before the forecasting period. Specifically, this method uses the posterior results of GLUE behavioral runs (referring to the light red shaded region in Figure 1), that is, selected parameter sets for running **Model**—called "*Selected*" specification method. The selected parameter sets for **Model** remain unchanged over the warm-up and calibration periods as well. As with the former approach, the values of state vectors are initially set to be zero at $t = 0$ but are continuously updated until $t = t_c$. We expected to see the EnKF process converge much faster and the forecasting results improve.

2.2.4. Modeling Approaches for Forecasting

In total, 18 modeling approaches (see Figure 2) were developed by combining the modeling options with various techniques (NAM + PCE + GLUE + EnKF) in sections 2.2.2 and 2.2.3. The modeling techniques were
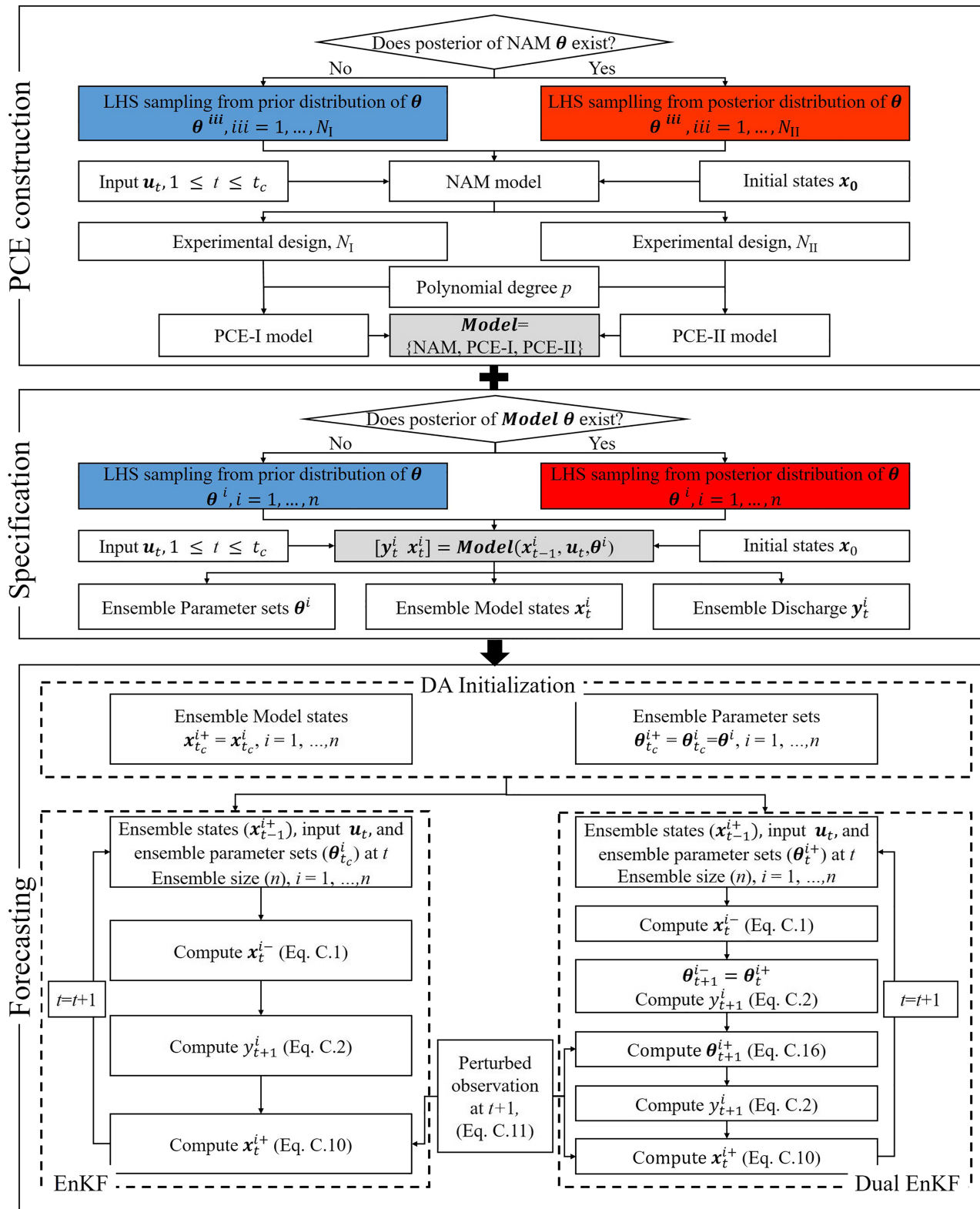
Figure 2. The overview of an ensemble flood forecasting framework. The top box, "PCE construction," is for the process of building two PCE models (blue box for PCE-I and red box for PCE-II). The middle box, "Specification," describes two distinct approaches of specifying model parameters before forecasting including *Random* (blue box) and *Selected* (red box). The bottom box, "Forecasting," corresponds to data assimilation for flood forecasting in real-time (single and dual EnKFs). The top and middle blue boxes correspond to sampling N_I and n, independently from the same prior uniform distributions, respectively, while the red boxes sampling N_{II} and n from the same posterior distributions.

Table 3
Forecasting Approaches Employed in This Study

Approach	Specification	*Model*	Data assimilation
A1	*Random*	NAM	None
A2			EnKF
A3			Dual EnKF
A4		PCE-I	None
A5			EnKF
A6			Dual EnKF
A7		PCE-II	None
A8			EnKF
A9			Dual EnKF
A10	*Selected*	NAM	None
A11			EnKF
A12			Dual EnKF
A13		PCE-I	None
A14			EnKF
A15			Dual EnKF
A16		PCE-II	None
A17			EnKF
A18			Dual EnKF

coupled to successfully perform ensemble flood forecasting and to meet the need for accurate and efficient flood forecasting. The 18 approaches represent permutations of the $3 \times 2 \times 3$ subcases (Table 3). First, they were divided into three subcases corresponding to **Model**, depending on whether a deterministic model or a PCE model was used over the calibration period (see section 2.2.2) and how the latter was developed. Second, these modeling sets were divided into two subcases corresponding to *Random* or *Selected* sets, depending on how the parameter sets before the forecasting period were specified (see section 2.2.3). Lastly, they were divided into three subcases depending on the methodology of data assimilation. The first of the three subcases did not use any data assimilation, and the other two used single- and dual-ensemble Kalman filters (see section 2.1.4). We evaluated the modeling performance of the coupling framework by assessing accuracy, efficiency, and predictability in section 4.2. The performance comparisons of the 18 approaches are expected to be a guide to which approach demonstrates better skill and most appropriate and which should be avoided.

2.3. Performance Metrics

To assess the modeling performance of the 18 approaches, metrics representing accuracy, predictability, and efficiency were chosen, beginning with the accuracy metrics of Nash-Sutcliffe efficiency (NSE), absolute error (AE), and relative entropy (RE) (Kleeman, 2002; Kullback, 1997; Kullback & Leibler, 1951). Second, Brier (1950) scores (BSs) and the range of uncertainty (UR) were used to assess the predictability of probabilistic forecasts. Lastly, a metric calculating total runtime (TRT) was evaluated to compare the computational efficiency of the tested approaches.

NSE, which is traditionally used to evaluate the accuracy power of deterministic models, is computed for each ensemble member (i) over the entire computation time. In this study, NSE is expressed as follows:

$$NSE^i = 1 - \frac{\sum\limits_{t=1}^{T}\left(y_t^{Obs} - y_t^i\right)^2}{\sum\limits_{t=1}^{T}\left(y_t^{Obs} - y_{mean}^i\right)^2}, \quad i = 1, ..., n \tag{10}$$

where y_t^{Obs} and y_t^i are the actual observation and ith predicted output at time t; y_{mean}^i is the mean of the ith predicted output over the entire forecasting period; T is the total number of time steps over the forecasting period from t_c to t_f.

AE is differences between actual observations and predictions of each ensemble members at each time t. Thus, it varies with time and can be written as follows:

$$AE_t^i = \left|y_t^{Obs} - y_t^i\right|, \quad t = 1, ..., T; i = 1, ..., n \tag{11}$$

RE is a measure of the statistical difference between probability distributions over the entire forecasting period of observations and model simulations (Giannakis & Majda, 2012; Kleeman, 2002; Shukla et al., 2006). Following Kleeman (2002) and Heo et al. (2014), RE can be defined as follows:

$$RE^i = \left[\log\frac{\sigma_{y^{Obs}}^2}{\sigma_{y^i}^2} + \frac{\sigma_{y^i}^2}{\sigma_{y^{Obs}}^2} - 1\right] + \left[\frac{\left(\mu_{y^i} - \mu_{y^{Obs}}\right)^2}{\sigma_{y^{Obs}}^2}\right], \quad i = 1, ..., n \tag{12}$$

where $\mu_{y^{Obs}}$ and μ_{y^i} are the mean, while $\sigma_{y^{Obs}}$ and σ_{y^i} are the variance of streamflow observation and the ith model prediction over the entire computation time from t_c to t_f. Small values of relative entropy indicate that distribution of a given model is close to that of the observation. This is also called Kullback-Leibler divergence between the two distributions, model and data, assuming Gaussianity of both.

The BS is one of the most commonly used verification measures for assessing the predictability of probabilistic forecasts. The score is defined as the mean squared error of the probabilistic forecasts over the verification sample, expressed as follows:

$$BS = \frac{1}{T} \sum_{t=1}^{T} \left(p_t^f - o_t \right)^2 \tag{13}$$

where p_t^f is the forecast probability for the tth time, which refers to the ratio among ensemble reaching a predefined flow threshold; o_t is the observed probability, which is 1 if observation at tth time, y_t^{Obs} is larger than the threshold, and 0 if it is not. In this study, this threshold value was chosen as the proportional rate of 90% of the true discharge peak.

The uncertainty range (UR) is the range between the 5th and 95th percentiles of the ensemble outcomes (q). It is computed over each computational time t in hydrographs, expressed in equation (14):

$$UR_t = q_t^{95} - q_t^5, \quad t = 1, ..., T \tag{14}$$

Lastly, the TRT for all of the approaches is defined as follows:

$$TRT = \left(RT_{w+c,\textbf{Model}} \times fac_{\textbf{Model}} + RT_{f,\textbf{Model},DA} \right) \times n + RT_{build,\textbf{Model}} \tag{15}$$

where $RT_{w+c,\textbf{Model}}$ is the run time to compute one simulation of **Model** (NAM, PCE-I, and PCE-II) over the warm-up and calibration periods, that is, from 0 to t_c; $RT_{f,\textbf{Model},DA}$ is the run time to compute one simulation of **Model** with different DA methods over the forecasting period, that is, from t_c to t_f; and $RT_{build,\textbf{Model}}$ is the run time needed for building **Model**. For example, because it is unnecessary for constructing the deterministic model, the time for NAM is zero. The building run times for PCE-I and PCE-II will be calculated in detail in section 4.1.2. The factor $fac_{\textbf{Model}}$ represents the number of **Model** runs to obtain a single behavior run in GLUE and remains 1 in A1 to A9, while it depends on **Model** for the rest of approaches.

Equation (15) is a linear function with respect to the number of ensembles run, in which $RT_{w+c,\textbf{Model}} \times fac_{\textbf{Model}} + RT_{f,\textbf{Model},DA}$ serves as the slope of the linear function and $RT_{build,\textbf{Model}}$ the intercept. The values of the slope and intercept and the executed times of the 18 approaches are addressed in section 4.2.

3. Study Area and Experimental Setups

In this study, the unified framework is applied to predict hourly streamflow in the Vu Gia watershed as shown in Figure 3. The watershed is one of the largest in central Vietnam, with a total area of 1,679.8 km^2 in the tropical region. It experiences a typical continental monsoon climate, with concentrated rainfall mainly from September to December. As the Vu Gia watershed is characterized by a large difference in elevation (slopes of approximately 30%), floods occur rapidly and frequently. The region has experienced intense severe flooding and significant damage (Nga et al., 2015; United Nations Development Programme (UNDP), 1999).

Streamflow data used for the outlet of the basin was collected hourly at Thanh My station—the only hydrometric station in the domain. Rainfall data was also observed hourly and obtained from two weather stations near the study area (Thanh My and Kham Duc station). The average rainfall over the basin (Figure 4) was calculated through the Thiessen polygon method. Observations, from 1 to 17 December 2016, are employed, in which the data from 1 to 13 December was used for the warm-up period (i.e., from 0 to t_w), the data from 13 to 15 December for the calibration period (i.e., from t_w to t_c), and the remaining data (assuming numerically that this data was newly provided at an hourly basis) corresponds to the forecasting period (i.e., from t_c to t_f) (Figure 4). The effects of the length of warm-up period are illustrated in the supporting information. Note that rainfall forecasts have not been considered in this experimental design, what is done is hindcasting but one refers to the period between t_c and t_f as the "forecasting period," allowing for replicating real-life operational flood-forecasting process. Also note that a source of uncertainty for rainfall forecasts has not
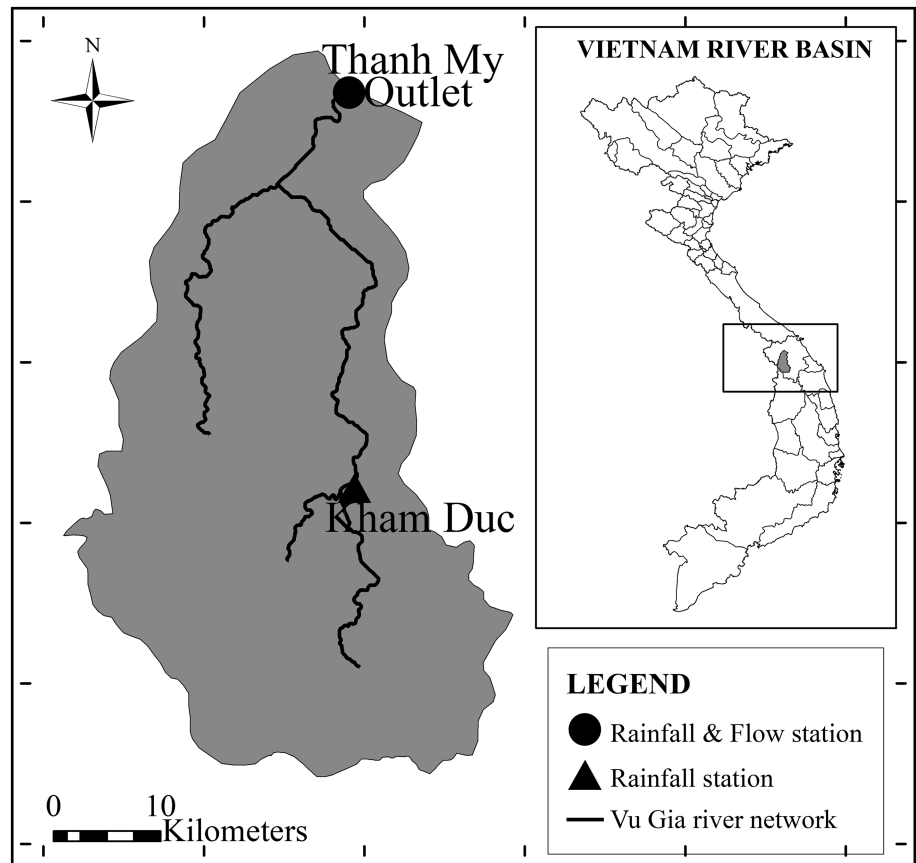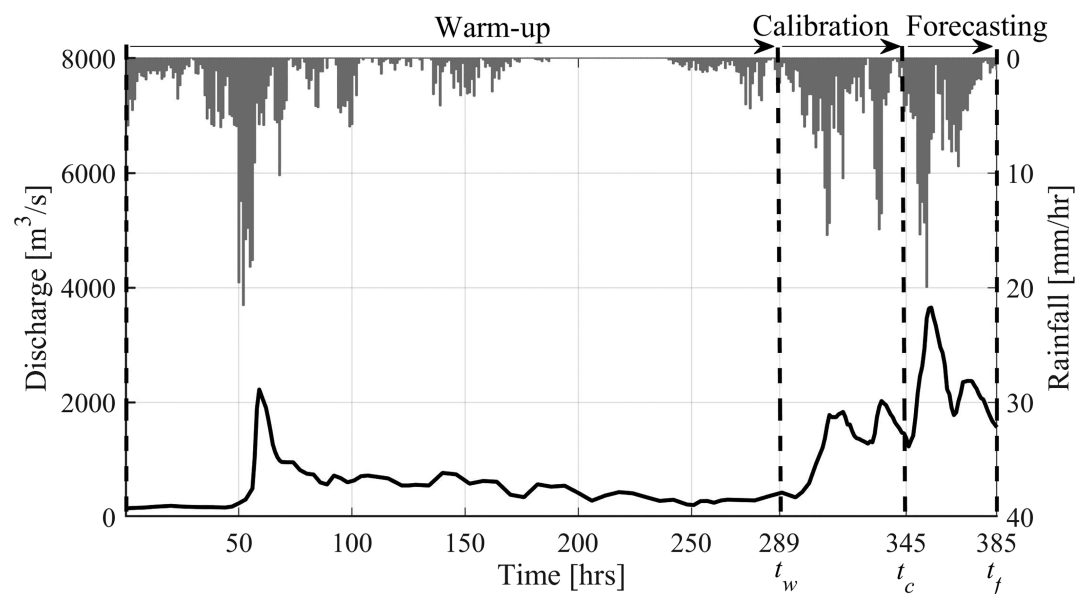
Figure 3. Study area: Vu Gia watershed.

Figure 4. Flood event used in the study. The black line is the discharge of outlet, and the gray hyetograph represents the average rainfall of Vu Gia watershed.

been presented, but it could have been addressed in equation (2) that has the flexibility to include *ensemble* precipitation inputs ($\boldsymbol{u}_t$).

Determining the size of ensemble for forecasting (n) is related to quantifying the uncertainty bounds and representing the EnKF. In previous studies, the ensemble size was selected randomly or large enough (at least 100 members) to fully identify the uncertainty confidence intervals (Beven & Freer, 2001; Blasone et al., 2008; Cameron et al., 2000; Choi & Beven, 2007; Hossain & Anagnostou, 2005; Jin et al., 2010; Shen et al., 2012). A sufficient number of ensemble parameter sets to achieve both goals of efficiency and uncertainty quantification should be determined. Following our previous study (Tran & Kim, 2019), we used an n of 500 as the optimal size of the ensemble.

4. Results

4.1. Preparation Steps Before Forecasting

4.1.1. Attaining Parameter Posterior Distributions

The posterior distributions of parameters can be generally attained by using Bayesian inference. As detailed in section 2.1.3, we employed a relatively simple and robust method, GLUE (Beven & Binley, 1992), that does not require reformulation of the deterministic model code. Details on why we choose the likelihood functions of NSE, PE, and VE (presented in Appendix B), how we determine the cutoff threshold values of each function, and which parameters are more sensitive, are described in Tran and Kim (2019). We confirmed the benefits of a warm-up technique that significantly speeds up the GLUE process of finding the behavioral sets: without warm-up, no behavioral set was obtained from GLUE even after a sufficiently large number of NAM model runs, while with warm-up, a behavioral set was obtained after approximately 118.0 model runs for NAM (A10 to A12), 26.9 for PCE-I (A13 to A15), and 3.6 for PCE-II (A16 to A18), respectively. Therefore, the factors, $fac_{\boldsymbol{Model}}$ are 118.0, 26.9, and 3.6 for NAM, PCE-I, and PCE-II, respectively, in A10 to A18.

4.1.2. Constructing the PCE Models

Determining the coefficients of the PCE-I and PCE-II models depends on the number of the experimental design (N) and the polynomial degree (p) (Blatman & Sudret, 2010, 2011). To discover appropriate values for N and p, the effect of experimental design N on PCE performance was first evaluated. Specifically, a number of simulations were repeated with the N value varied between 10 and 5,000 but the value of p was set as 3, and the performance results of LOO for streamflow (y) and the five model states (Table 2) computed. Figure 5 shows that the LOO values for streamflow and five state variables become smaller as the value of N increases and ceases to become smaller when N approaches a certain value. For N values larger than this threshold, the model performance was almost indistinguishable (the left column plots in Figure 5). From a visual inspection of Figure 5, the optimal N value for constructing the PCE-I and PCE-II models would be 1,000 and 100, respectively.

A selection of the polynomial degree p was made in a fashion similar to the aforementioned procedure. The value of p was varied from 1 to 6 and N was set as 1,000 (PCE-I) and 100 (PCE-II). From the results of Figure 5 (the right column), the gradients of the LOO metrics assessed changed considerably when p was set to 3 and the values remained stable for large magnitudes of p. In terms of reducing the computational time to construct a PCE model, a low polynomial degree would be preferred. Thus, a p of 3 would be an appropriate value to use when building both PCE models. With optimal values of N of 1,000 and 100 and a p of 3, PCE-I and PCE-II models can be built to quantify the uncertainty range for flow prediction and to compare the degree of accuracy and efficiency with the results of the deterministic NAM.

The total time to establish both PCE models is described (further details are in Tran & Kim, 2019). Obviously, the larger the number of the experimental design set, the more time is needed for computing N ensemble runs. The time required to perform the N_{I} and N_{II} ensemble runs of NAM was 121.9 and 12.6 s for PCE-I and PCE-II, respectively. It also takes much more time to estimate PCE-I coefficients if one uses an ensemble set (N_{I}) generated from the prior distribution of the parameters than to compute PCE-II coefficients from parameter sets informed by the likelihood function. The time required to estimate PCE coefficients was 419.3 and 11.3 s, respectively. The summation of these two times was considered to be the total time required to build the PCE models before forecasting: approximately 541.2
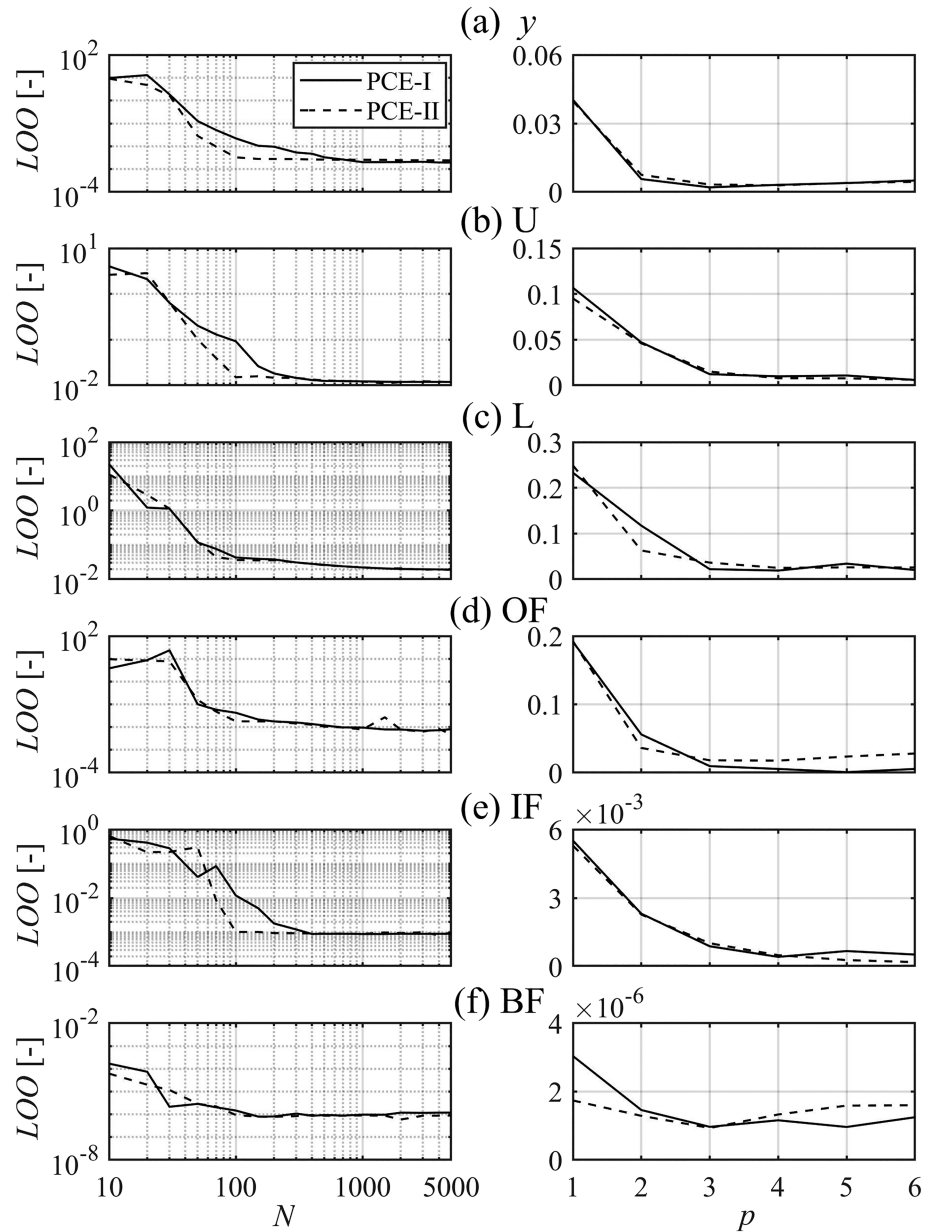
Figure 5. The effects of (left plots) the experimental design, N and (right plots) the polynomial degree, p on the leave-one-out cross-validation error (LOO) in constructing PCE-I and PCE-II models, for (a) streamflow and (b–f) 5 model states in Table 2.

and 23.9 s for PCE-I and PCE-II, respectively. The construction time of PCE-II is much (~22 times) faster than that of PCE-I.

4.1.3. Comparing the Ensemble Results of NAM and PCE Models

Over the calibration period, ensemble results composed of 500 *Random* and *Selected* runs were compared for three different models. To make the 500 *Selected* behavioral results, 58,977, 13,444, and 1,822 (n_w) random runs were required for NAM, PCE-I, and PCE-II, respectively. Compared with the NAM itself, using PCE models can reduce the amount of computational runs by a factor of about 4.4 for PCE-I and 32.4 times for PCE-II model. The composing behavioral set for PCE-II was even faster (~7.4 times) than for PCE-I.

Figure 6 shows hydrographs for the 500 *Random* (A1 to A9) and *Selected* (A10 to A18) simulations for the three models. Their uncertainties are illustrated with a 90% confidence interval, which corresponds to 5 and 95% quantiles of the 500 ensemble members. Because we controlled the conditions for the behavioral
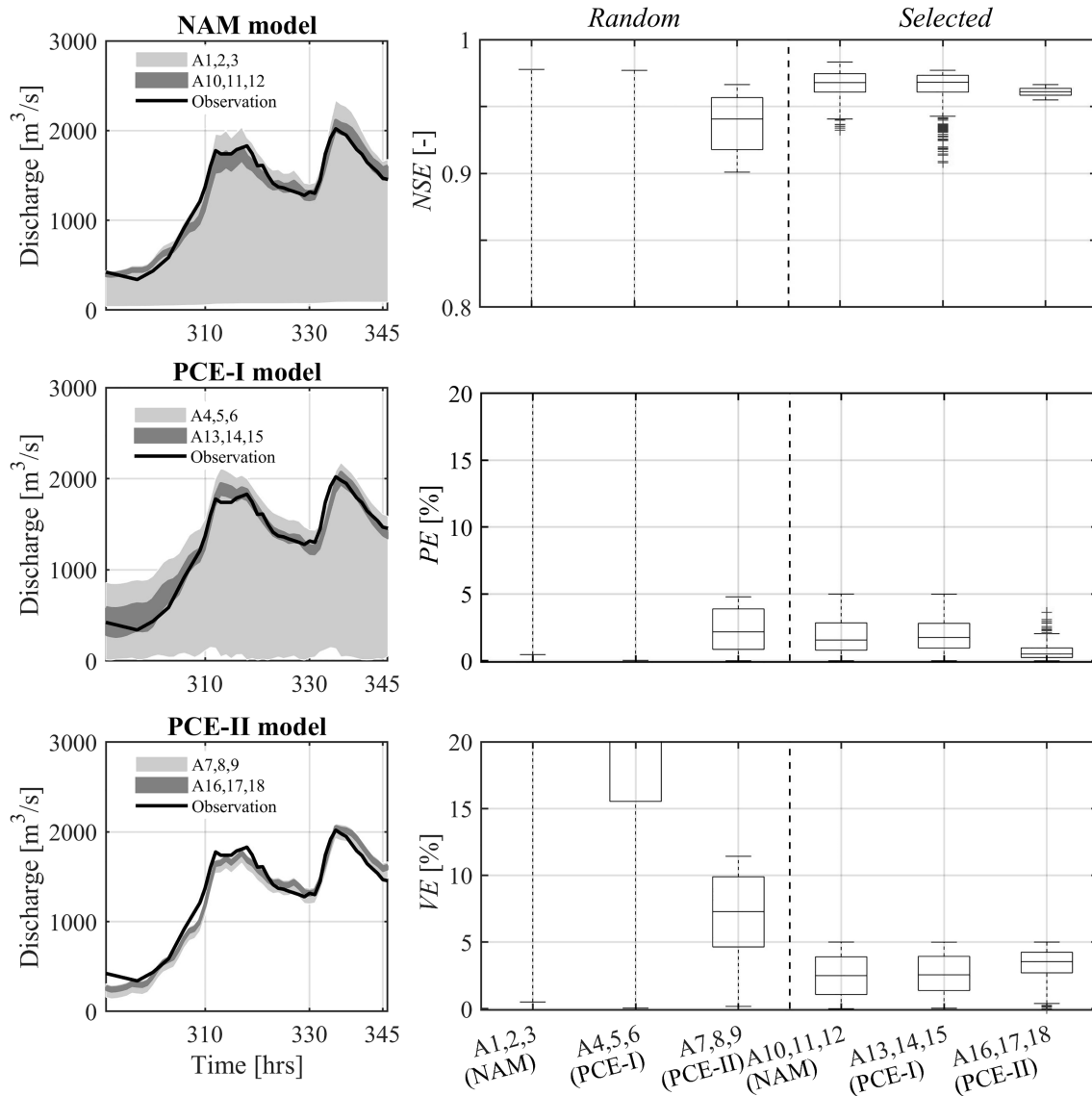
Figure 6. The left column plots show hydrographs for the calibration period. The shades in the plots correspond to 90% confidence interval for 500 *Random* model runs (A1 to A9, light gray shade) and 500 *Selected* model runs (A10 to 18, dark gray shade) for 18 approaches in Table 3. The boxplots in the right column demonstrate the verification metrics of *NSE*, *PE*, and *VE* for the 18 approaches used.

set of GLUE, the overall comparison with the observed values for the results of the *Selected* cases (A10 to A18) is very satisfactory. Specifically, the *NSE* value was always higher than 0.9 and both *PE* and *VE* values were less than 5% for all cases. However, streamflow curves for the *Random* simulations (A1 to A9) clearly show different patterns depending on the model. It can be anticipated that the results of these *Random* cases will not be encouraging and their uncertainties will be large. However, the results of some cases using PCE-II model were very satisfactory and their uncertainties small.

As mentioned above, when making using observations to constrain the parameter sets (A10 to A18), the results of both PCE models are similar to those of the NAM and no substantial differences were observed. This confirmed that both PCE models have an equivalent degree of accuracy as the NAM and can provide an excellent match to the deterministic model. In terms of efficiency, it is also advantageous to use the PCE model (discussed in sections 4.2.1 and 5.1), and there is no reason to hesitate adopting the PCE model for streamflow prediction.
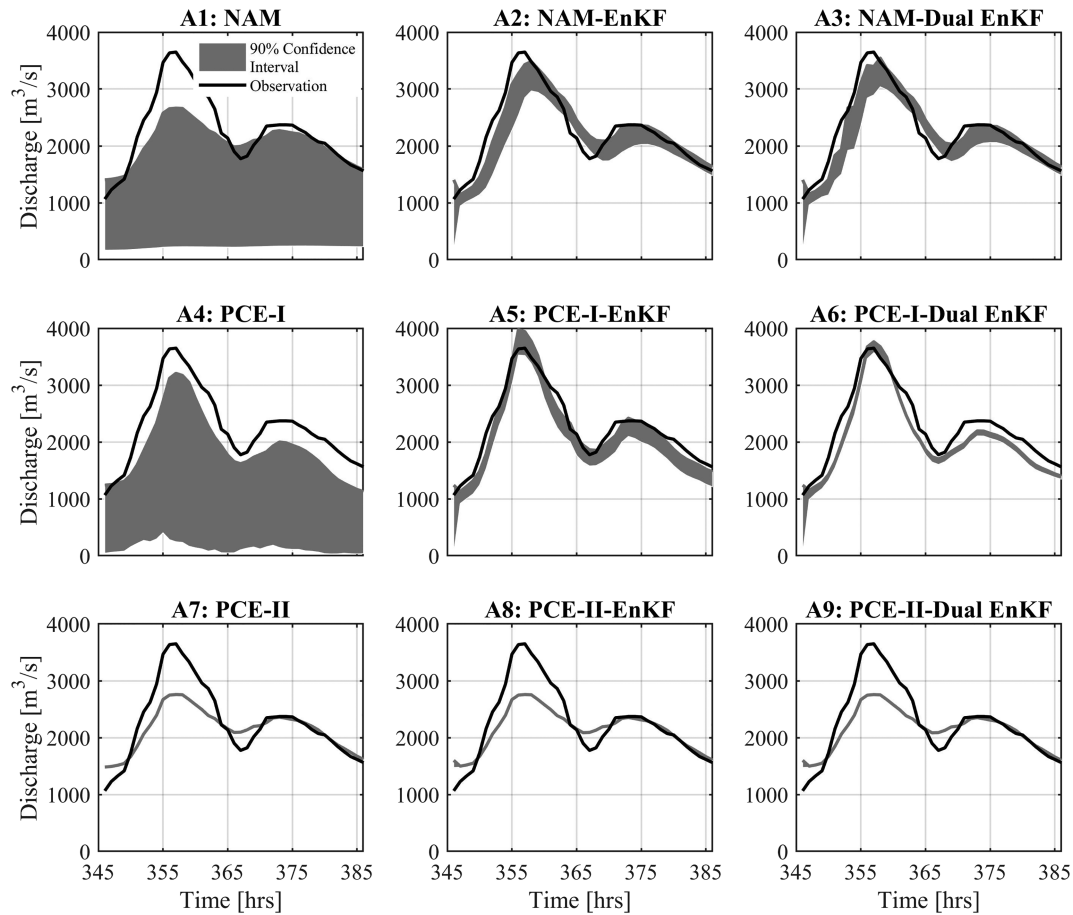
Figure 7. Hydrographs over the forecasting period, with a 90% confidence interval of 500 *Random* model runs (A1 to A9).

4.2. Flood Forecasting With 18 Approaches

For past periods when observations for calibration were collected, all models performed well. We then tested the forecasting performance of the three models using the permutation that resulted in 18 approaches in Table 3. Results of the real-time flood forecasting for these approaches are shown in Figures 7 and 8, in which the 500 ensemble results are illustrated with a 90% confidence interval at each time step. The verification metrics for the simulations, specifically NSE, AE, RE, BS, and UR, and the peak values of hydrographs are compared in Figures 9 and 11. Lastly, the total run time with respect to the ensemble size was computed for the 18 approaches in Figure 10. In this section, we analyze the results and draw conclusions from the following four perspectives: (i) the applicability of PCE-I and PCE-II models for real-time flood forecasting, (ii) the impact of estimating appropriate parameter conditions for forecasting, (iii) the effect of using EnKF and dual EnKF, and (iv) the degree of improving efficiency performance among the approaches.

4.2.1. PCE-I Versus PCE-II Model for Real-Time Flood Forecasting

Depending on the model used in forward simulations (NAM, PCE-I, and PCE-II), the results for the 18 approaches were divided into three groups. Almost all of the results of the six approaches using the PCE-II model were worse than those obtained with both NAM and PCE-I (Figures 7 and 8). The only exception is for the A1 and A4, which did not have assimilation and whose parameter sets used were based on prior uniform distributions. No verification metrics computed using the results of forecasting based on PCE-II were satisfactory, except for the metric of $\overline{UR}$. However, if the accuracy is not ensured, the better performance in terms of $\overline{UR}$ is not meaningful. Specifically, *NSE* values were low, approximately 0.7; AE values at flood peak time (AE_{peak}) were larger than 750 m³/s; RE was approximately 0.01; and *BS* was equal to 1 (Figure 9). No metric improvements were found for the approaches based on PCE-II, even if combinations of assimilation and calibration techniques were applied. We concluded that the PCE-II model can reproduce streamflow characteristics well for the past period but not for the future.
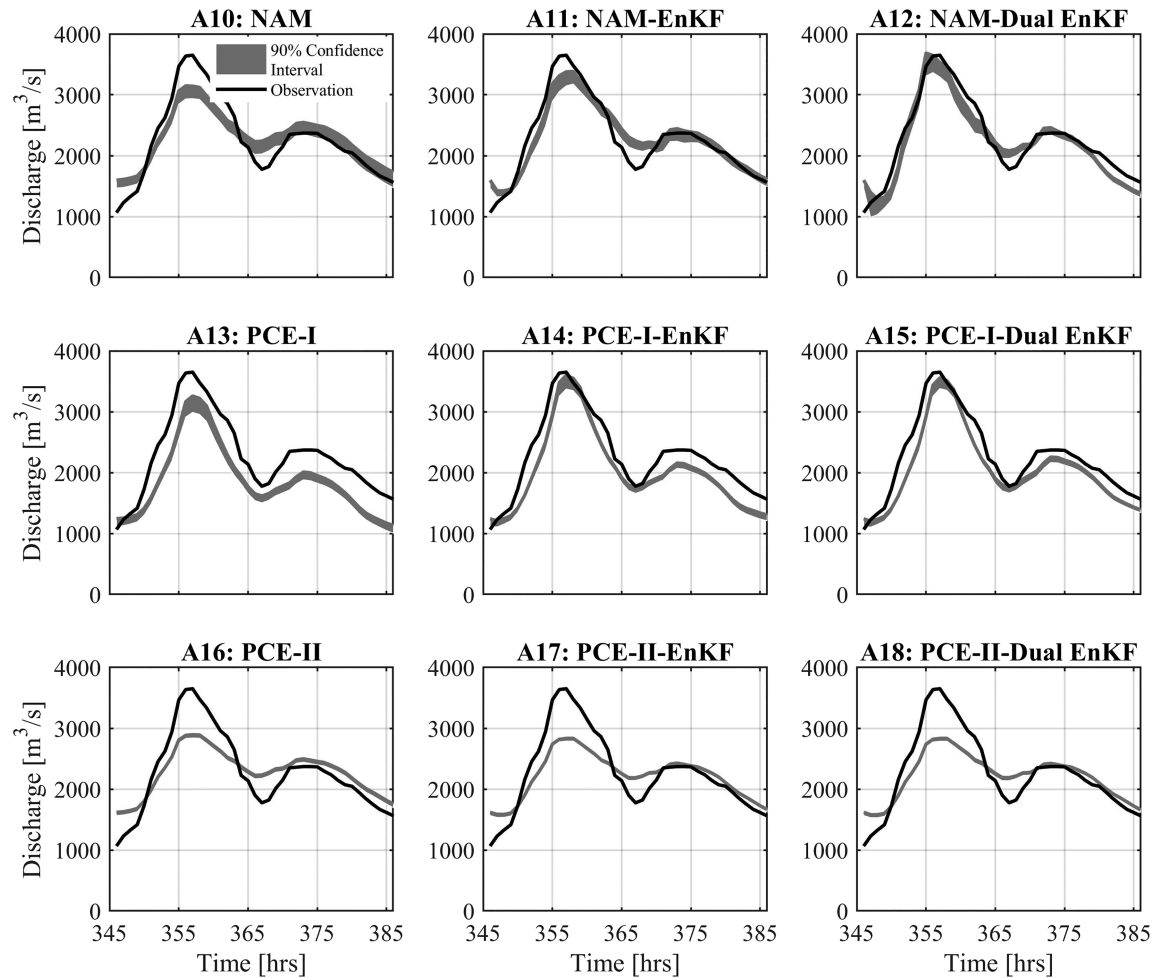
Figure 8. Hydrographs over the forecasting period, with a 90% confidence interval of 500 *Selected* model runs (A10 to A18).

Conversely, the forecasting results of the approaches based on the PCE-I model are almost similar to those obtained with NAM and in some cases even better. The latter can be seen in Figure 9; the verification metrics of *NSE*, AE_{peak}, *RE*, and $\overline{UR}$ show better performance for PCE-I than for NAM results (e.g., A5 vs. A2, A6 vs. A3, A14 vs. A11, and A15 vs. A12) (see Table 4). In particular, the RE results in Figure 9c illustrate that the PCE-I results are closer to the observed values than those obtained with NAM (A15 is the best result with the smallest value of RE). *BS* corresponding to PCE-I also has smaller values, close to zero, which indicates instances of when predictability of probabilistic forecasts matched predictability of observation (Figure 9d). Therefore, the PCE-I model can be adapted to substitute the NAM in performing real-time flood forecasting, as well as in capturing the uncertainty of calibration period.

Comparing the modeling results in terms of the computation speed, it is clear that simulating a surrogate model using the PCE theory is significantly faster than with a deterministic model such as NAM. The "slopes" of the runtime curves of Figure 10 indicate both PCE approaches are approximately 20 times faster (A4 to A9) and ~80 times faster (A13 to A18) than the corresponding approaches using the NAM. Similarly, if we compare efficiency between PCE model approaches, using PCE-II may or may not offer much improvement in efficiency over PCE-I. There is only 10% improvement when *Random* specification is applied (see the slope of A4, A5, and A6 vs. A7, A8, and A9 in Figure 10), while there is about 6 times improvement when simulating *Selected* approaches (see the slope of A13, A14, and A15 vs. A16, A17, and A18). The use of surrogate models therefore did not sacrifice accuracy. The flood prediction accuracy of PCE-1 model presented here is similar to that of the original NAM, and computational efficiency has been found to be highly superior.
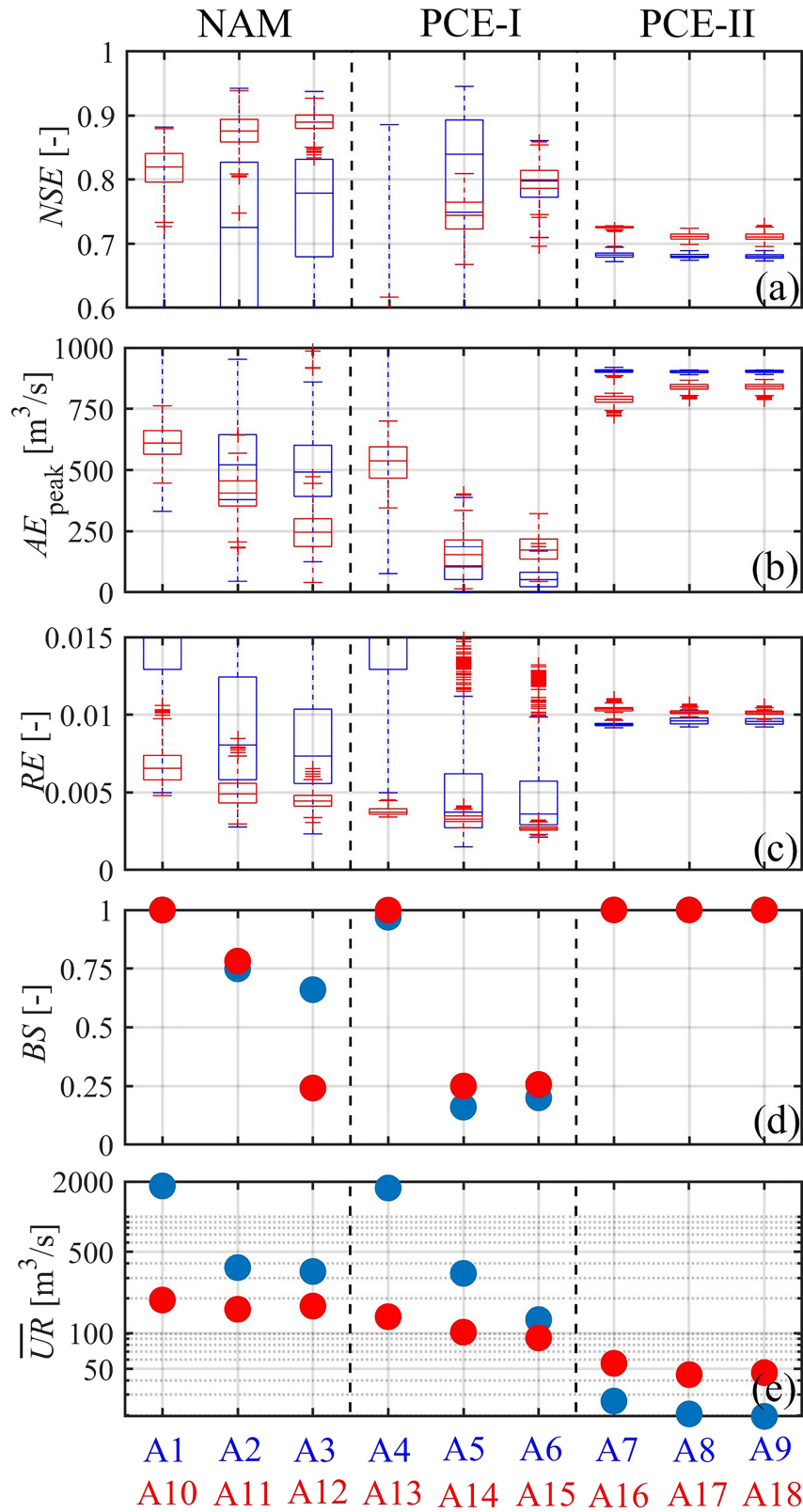
Figure 9. The performance metrics reflecting accuracy and predictability of the 18 approaches for the forecasting period. Boxplots of (a) NSE, (b) AE_{peak} (AE at flood peak time), and (c) RE show 500 ensemble values with the statistics of median (central mark), the 25th and 75th percentiles (edges of the box), and maximum and minimum except for outliers (whiskers). (e) $\overline{UR}$ is the mean of uncertainty range, UR_t over the entire forecasting period.
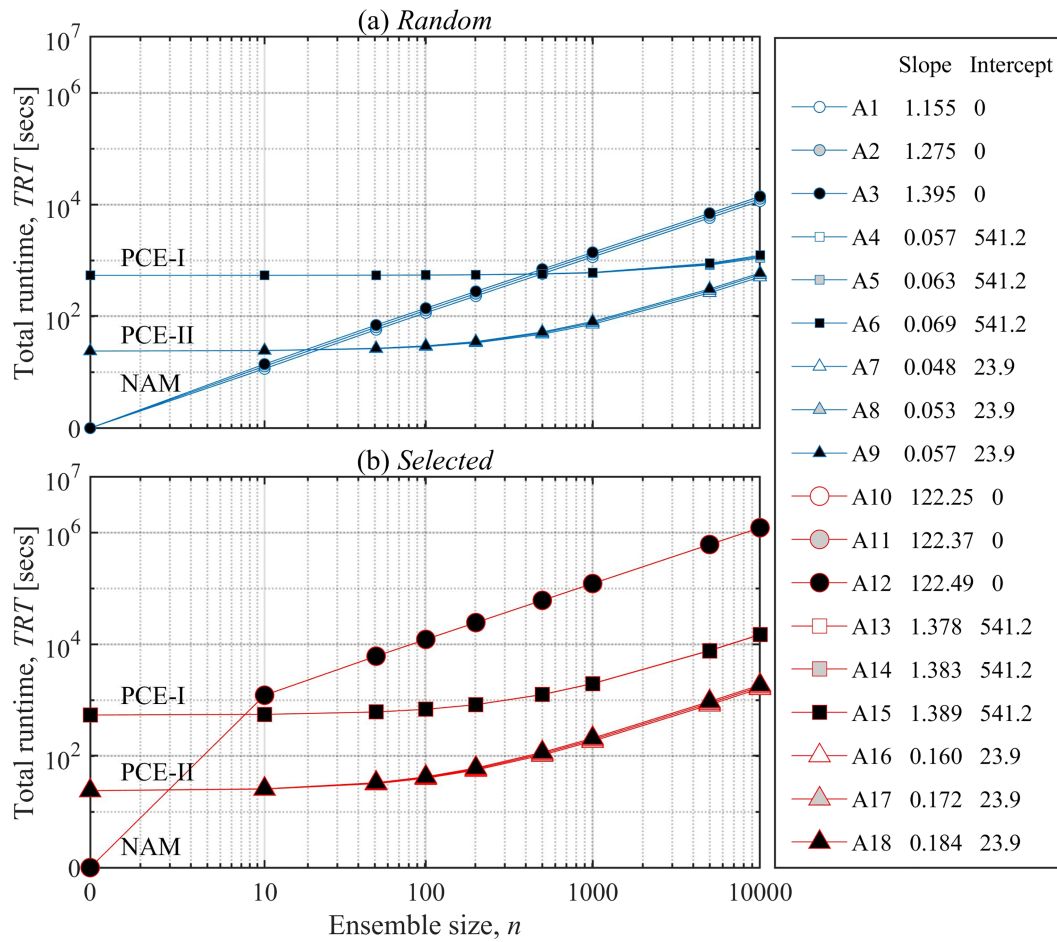
Figure 10. The total runtime (*TRT*) corresponding to 18 approaches in the forecasting period versus ensemble size (*n*). Note that although we plot on logarithmic axis, the actual total runtime has the form of a linear function with the ensemble size at linear scale; its slope and intercept values for all approaches are tabulated on the right.

Table 4
Performance Metric Values of 18 Approaches

Approach	Median of *NSE* [-]	Median of AE_{peak} [m^3/s]	Median of *RE* [−]	*BS* [−]	$\overline{UR}$ [m^3/s]
A1	−3.62	2292.19	0.019	1.00	1834.18
A2	0.70	515.64	0.009	0.75	367.34
A3	0.75	500.45	0.009	0.66	340.12
A4	−3.84	1928.86	0.015	0.97	1760.11
A5	0.82	126.48	0.005	0.16	327.86
A6	0.79	55.84	0.005	0.20	131.92
A7	0.68	904.21	0.009	1.00	26.52
A8	0.68	901.91	0.010	1.00	20.59
A9	0.68	902.79	0.010	1.00	19.63
A10	0.82	612.60	0.007	1.00	193.95
A11	0.88	401.26	0.005	0.78	161.69
A12	0.89	242.62	0.005	0.24	172.18
A13	0.44	532.35	0.004	1.00	139.54
A14	0.74	157.07	0.003	0.25	102.98
A15	0.80	176.00	0.003	0.26	91.98
A16	0.73	787.18	0.010	1.00	55.75
A17	0.71	839.31	0.010	1.00	44.64
A18	0.71	840.19	0.010	1.00	46.37

Note. The values in the first three columns are the medians of *NSE*, AE_{peak}, and *RE*, which match the values in Figure 9.
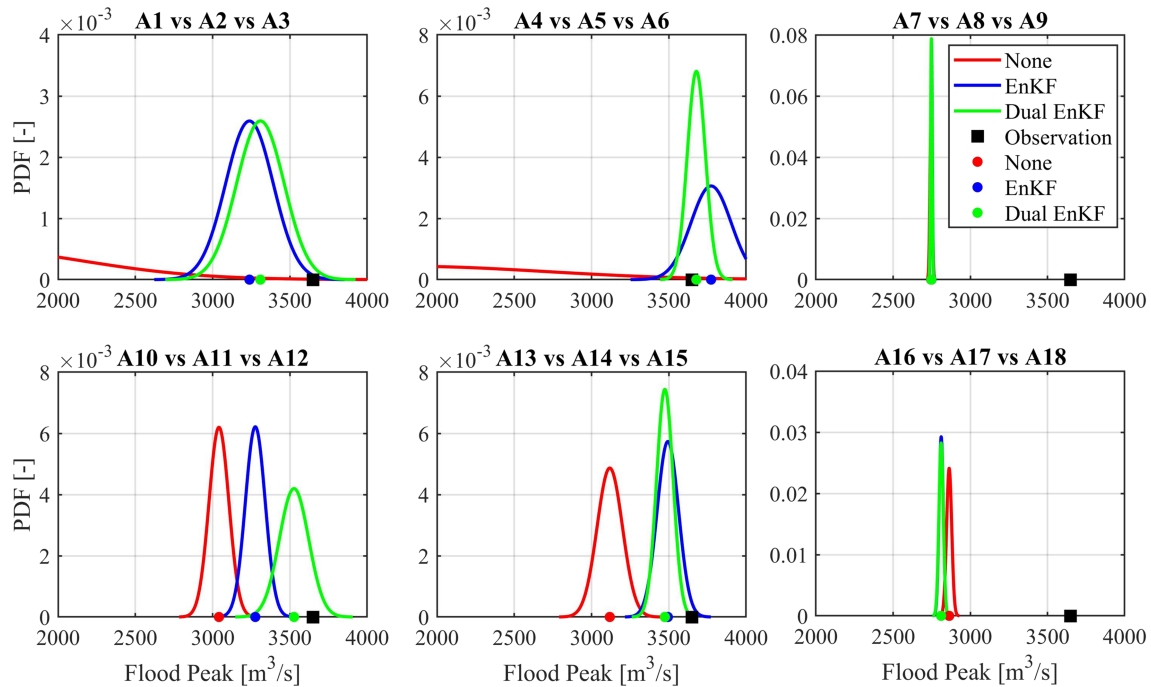
Figure 11. Comparisons of the three assimilation methods (none, EnKF, and Dual EnKF) for 500 ensemble flood peaks over the forecasting period. The left, middle, and right plots correspond to the approaches using NAM, PCE-I, and PCE-II, respectively. The first and second row plots correspond to the approaches using *Random* and *Selected* methods for parameter specification. The black square represents observed value at peak time; the circles are the expected values of the sample probability density functions.

4.2.2. Random Versus Selected Specification for Forecasting

The approaches using the *Selected* specification generally show a better performance than those using the *Random* specification. This is especially noticeable in the NAM and PCE-I approaches and rarely in PCE-II. First, in the approaches without data assimilation, their accuracy was significantly improved (compare A1 vs. A10 and A4 vs. A13). The performance of A10, represented by the NSE, AE_{peak}, RE, and $\overline{UR}$ metrics, was improved by about 95%, 73%, 61%, and 89% compared with A1, and the performance of A13 about 86%, 72%, 79%, and 92% over A5, respectively. Despite the noticeable improvement of A10 and A13, these results were still not ideal. The large AE error at the peak of A10 and A13 was approximately 450 m^3/s less than the observation, and the BS value was close to 1 (Figure 9 and Table 4). On the other hand, in the approaches in which data assimilation was used, the improvement effect for *Selected* specification was not greater than when it was not used. The increasing performance for the same metrics was about 55%, 22%, 36%, and 56% (A2 vs. A11) and about 56%, 52%, 44%, and 49% (A3 vs. A12). Here the parameter specification effect was smaller because DA improves the absolute error magnitude.

Determination of states and parameters that can increase accuracy and predictability requires more computation time because a large number of model runs are carried out to make an inference for posterior distributions. For approaches using NAM (A1 vs. A10, A2 vs. A11, and A3 vs. A12), it took 56, 41, and 30 times longer; while for PCE-I (A4 vs. A13, A5 vs. A14, and A6 vs. A15), it took 13, 10, and 8 times, respectively (Figure 10). Because of this computational burden, parameter inference can be a weakness for real-time flood forecasts where it is important to ensure sufficient time ahead. However, if the surrogate model is employed, the necessary repetition of estimating the posterior distribution can be performed quickly, and such a weakness can be overcome.

4.2.3. Single Versus Dual EnKF in Real-Time Flood Forecasting

Convincing evidence is presented that both single and dual EnKF can improve accuracy and predictability during real-time forecasting (with the exception of approaches using PCE-II). Both of these techniques perform well but the dual EnKF is the superior choice. As an example of the approaches using NAM, the three metrics of AE_{peak}, BS, and $\overline{UR}$ in the *Random* cases provided slightly better results: 515.64 versus 500.45, 0.75 versus 0.66, and 367.34 versus 340.12, respectively (A2 vs. A3). But, in the *Selected* cases, there was a

relatively large performance improvement for the two metrics of AE_{peak} and BS: 401.26 versus 242.62 and 0.78 versus 0.24 (A11 vs. A12). Similar trends were observed when using PCE-I, and the difference is remarkable, especially for the AE_{peak} metric (e.g., about 2.5 times for A5 vs. A6).

From the overall inspection, it can be determined that the dual EnKF can adjust the peak of a hydrograph more accurately and give a more confident result with a smaller uncertainty range. Therefore, we compared the distribution of flood peak values for 500 ensemble members in Figure 11. This figure confirms that the joint update of states and parameters improves accuracy at flood peak more effectively than a single update of states. Also, for the joint update, the expected value of the distribution was closer to the peak observation, and its variability is smaller (a narrower distribution).

Because the updating process is made twice, the dual EnKF is computationally more expensive. The computation time it takes to update states and parameters increased almost linearly. That is, the calculation time doubled or tripled for the cases of single and dual EnKF (using *Random* specification), respectively, as compared to the case without assimilations. However, for the approaches using the *Selected* specification, the calculation time did not seem to change significantly (Figure 10), not because the time required for Kalman filtering was reduced but because the time required for the parameter inference was so large that the filtering effect was masked.

5. Discussion

5.1. How Can PCE Be Constructed for Flood Forecasting?

From the simulated flood forecasting results presented in section 4.2, it is apparent that the manner of PCE construction has a significant impact on forecasting. The biggest difference in building PCE-I and PCE-II involves setting the range of the training sample (called experimental design). It is not surprising that a surrogate model trained for an event provides acceptable results only for the event trained. The flexibility to generalize to well-behaved outcomes for another event (e.g., a future event) is relatively low. This is why the calibrated model is often not appropriate for future forecasting. On the other hand, if a surrogate model can mimic the behavior of the original model to the greatest extent possible in a wide variety of situations and conditions, it will be able to capture its characteristics more comprehensively, thus playing a sufficient role in forecasting future events. Here we provide evidence the PCE-I model behaves like the NAM for the forecasting period, while the PCE-II behaves differently (despite both models behaving properly for the calibration period). To examine the robustness of both PCE model results, the Sobol′ method (detailed in Appendix D) was used to implement the variance-based measures of parameter sensitivities (Sobol′, 2001), which is commonly used as a global sensitivity analysis technique to determine the key parameters in the model (Wang et al., 2018).

First, the PCE-I posterior histograms of the nine parameters obtained from GLUE for the calibration period are similar to those of the NAM, except for Lm and TG (Figure 12). For these two parameters, a posterior histogram difference is a minor issue because the choice of the parameter values does not affect the end result, that is, the sensitivity of the parameters is low. Other parameters of CQOF (first) and CK12 (second) are the two most influential parameters to the model results, that is, their sensitivities are high. This result is consistent for both NAM and PCE-I (Figure 13). The slight difference between the results of PCE-I and NAM, observed from the investigation of the sensitivity and the posterior distribution, is because we chose an appropriate number of training samples when constructing the PCE-I model. If one greatly increases the number of training sets, the difference in the above results will essentially disappear.

Second, the failure of PCE-II to mimic the NAM for the forecasting period can be explained largely due to the fact that PCE-II was trained using the only 100 behavioral parameter sets that were optimized for the calibration event. Model results will only vary within the boundaries that its trained data understand, and it will not be able to simulate the behavior of another event with a high skill, that is, model "overfitting" occurs. However, over the calibration period, PCE-II always shows a good predictive performance for almost all parameter sets (compare the hydrographs of A1 to A3 with A7 to A9 in Figure 6). In other words, no matter what parameter one chooses, satisfactory results are always achieved, which indicates that the influence of parameters is excluded. The posterior histograms of parameters for PCE-II (Figure 12c) are almost uniform, except for the parameter of CQOF, which is the only one that can affect the end result, especially maintain
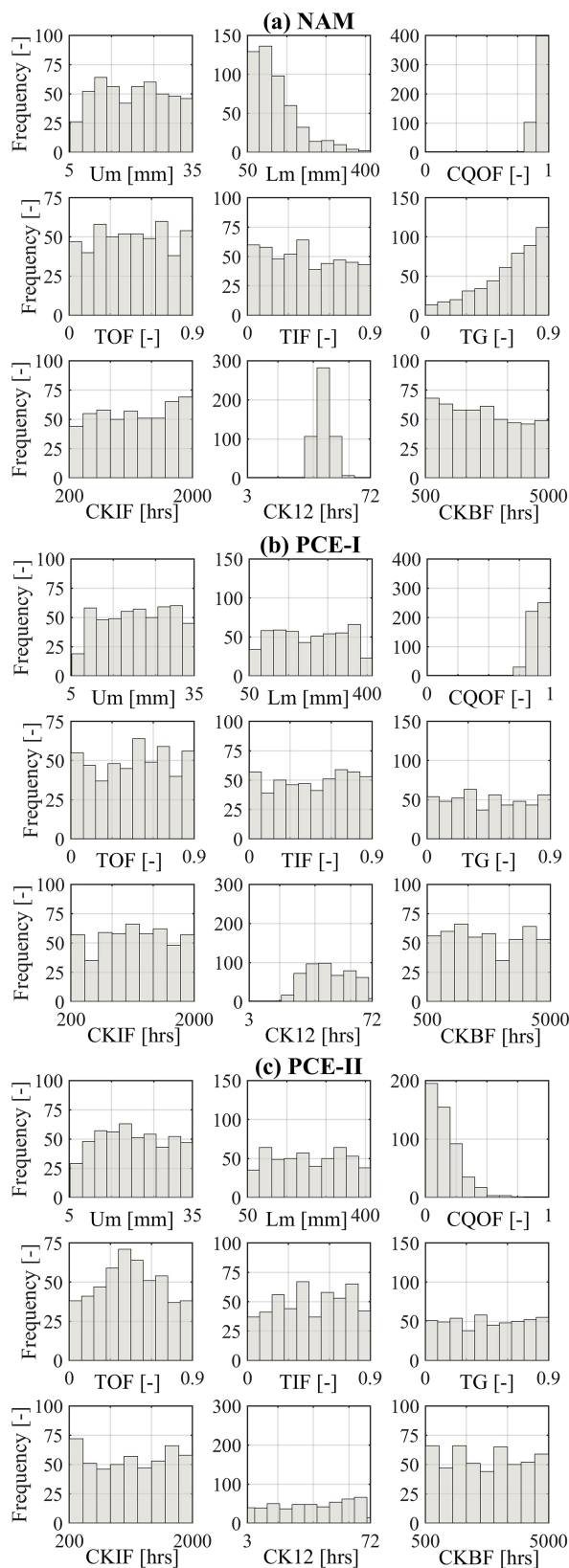
Figure 12. The posterior histograms for the nine model parameters in Table 2 from 500 behavioral sets of three models (NAM, PCE-I, and PCE-II) inferred by GLUE over the calibration period.

the accuracy of the flood peak (note that the sensitivity of this parameter for *PE* is unusually high in Figure 13c). If we change the threshold value of the likelihood function corresponding to the flood peak chosen to make the behavior set a slightly less constrained, this parameter will no longer play a role in constraining the result and follow a uniform distribution as well.

Another interesting aspect of the sensitivity test is that the sensitivity results of PCE-II differ from those of NAM and PCE-I but are similar to those of NAM-II. The sensitivities of parameters have been altered in PCE-II. The NAM-II in Figure 13d is hypothetically introduced to mimic the situations of PCE-II. Specifically, it refers to the sensitivity results when the NAM model was tested based on the posterior distributions (which are also used to select the training parameter set for building PCE-II), not the prior distributions of the parameters.

5.2. Is It Feasible to Construct a Time-Invariant PCE Model?

A long-lasting challenge in hydrologic modeling is how to estimate parameters or state vectors optimized for all external and internal conditions. This would not be an issue for estimating previously described variables if the amount of data for calibration was sufficient. However, in the case of future forecasts during which no observation for calibration is available, it poses a problem. To tackle this challenging problem, Fan et al. (2016) and Wang et al. (2017) adopted a modeling framework in equation (1), so that PCE models should be reconstructed continuously at every time step. This method is flawless in theory but requires additional computational resources (see efficiency comparisons in the supporting). That is, the time to configure the PCE at every time step must be added to the total model simulation time, that is, making the slope in Figure 10 steeper. This disadvantage can be more pronounced when constructing surrogate models for complex, process-based deterministic models.

Unlike previous efforts, this study adopted an alternative modeling framework such as equation (2); that is, the PCE model is time invariant and thus developed only once over the calibration period. Therefore, during real-time forecasting, the total run time consists only of computational intervals needed for data assimilation of all ensemble members. This enhances computational efficiency significantly (see efficiency comparisons in the supporting information). This framework is not perfect, but the potential error that can occur by using the time-independent PCE model is minimized by coupling the data assimilation technique, thus complementing accuracy. From a comparison of the results of 18 approaches, we confirmed that the modeling framework needed for building a PCE model (especially PCE-I) is feasible. This embraces the notion that the PCE construction does not require information for future conditions but can be made with historically available data available prior to the forecasting period.

5.3. Do Surrogate and Specification Sacrifice Efficiency?

Our results indicate that a sophisticated combination of three independent techniques (i.e., surrogate modeling, parameter inference, and data assimilation) supplies superior predictive performance for real-time ensemble flood forecasting. The combination of many methods however leads to an essential reduction in efficiency. Because data assimilation has been shown to be necessary, we must accept efficiency deterioration.
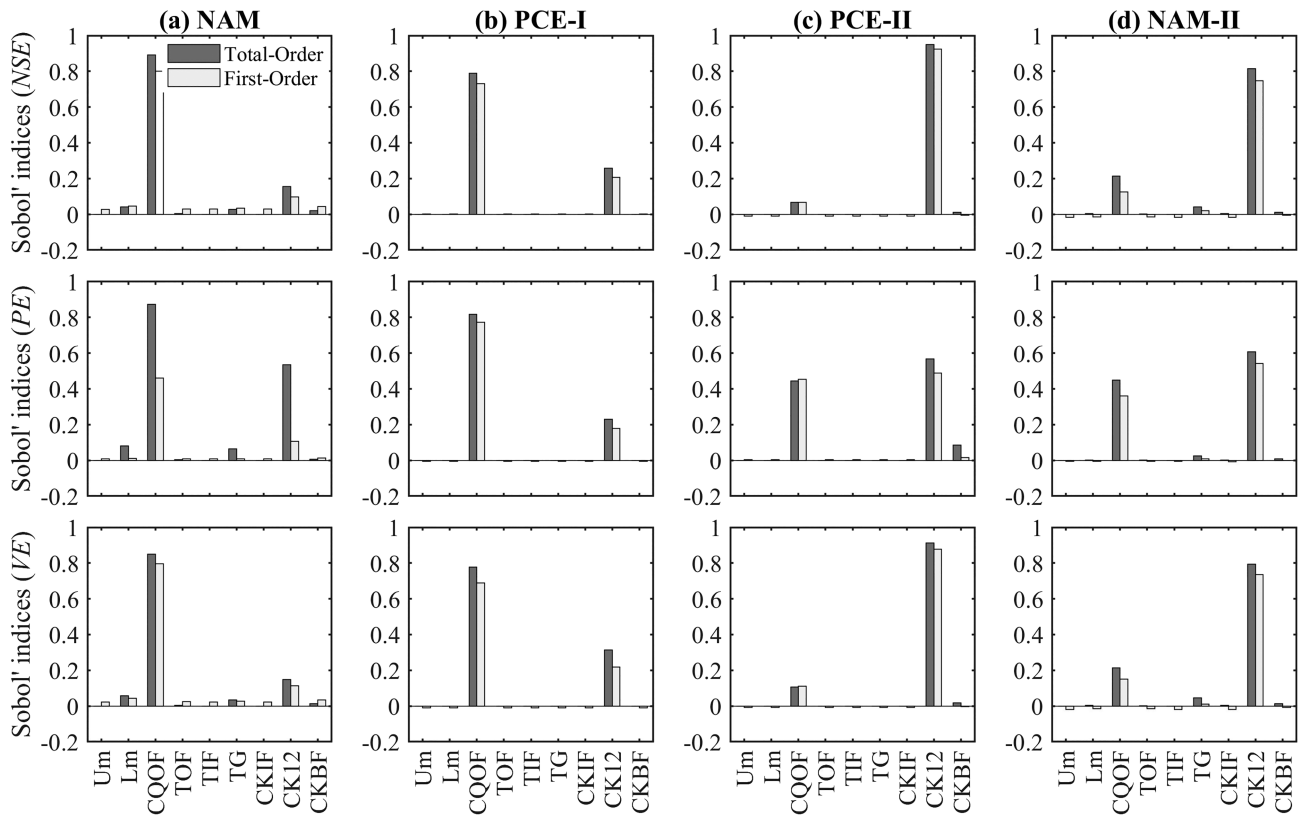
Figure 13. Sobol' sensitivity analysis for the nine parameters, computed for the three likelihood functions of (top) *NSE*, (middle) *PE*, and (bottom) *VE* over the calibration period. The sensitivity results are attained based on (a–c) the prior distributions of parameters for the three models of NAM, PCE-I, and PCE-II, respectively; and (d) the posterior distributions of parameters for NAM model. The posterior is also used to select the training parameter set for building PCE-II.

However, for surrogate modeling and parameter specification, it remains to be determined whether the additional time required by the technique combination leads to efficiency deterioration. First, for construction of the surrogate model, particularly PCE-I, the efficiency issue may not be relevant because the task does not require any observations for calibration and can be completed before the flooding season. In contrast, obtaining an ensemble of parameter sets from posterior distributions should be carried out immediately prior to the flood forecasting period, when observations are necessary. Therefore, it may take an appreciable time for completing this task, and method efficiency may be affected.

5.4. What Are the Differences Between PCE and Data-Driven Models?

Both PCE and data-driven models can provide satisfactory results for short-term forecast, but key differences between them exist. (1) PCE has a functionality of including model parameters and states as an input vector —this enables formal uncertainty quantification and model sensitivity analysis; (2) hydrologic/hydraulic model state variables (and parameters) are theoretically observable and in the case of process-based models have their own physical meaning, making it easier to physically interpret the results of PCE; (3) while purely data-driven methods are trained with observations, PCE is trained through high-fidelity samples supervised by physical relations, thus requiring fewer data samples for training; (4) data-driven models often have assumptions about the distributions governing variability of their outputs, and therefore, this can lead to nonphysical results (e.g., negative outputs quantifying mass, and streamflow) and fail to display nonnormal, bimodal, or other complex behaviors.

5.5. Can Modeling Framework Be Applied to High-Dimensional Problems?

While the implementation and analysis of experiments is valid for the presented scope of the experimental design, one needs to proceed with care when extending this approach to more complex models. The most fundamental concern that remains is whether the proposed framework can be applied to high-dimensional

problems in which fully distributed models are used. The dimension of a distributed model can be defined as the product of the number of grids cells and the number of parameters (and states). The dimension order of any truly physical models is therefore large, and extending our framework directly to such a model is not straightforward—known as the "curse of dimensionality" (Caflisch, 1998; Davis & Rabinowitz, 2007; Sudret, 2007). By examining how each of the methods mentioned in the framework resolves the problem of reducing dimensions efficiently and to what extent it has been applied, the feasibility of applying the proposed framework can be estimated.

Regarding the surrogate modelling (PCE), techniques such as Bayesian compressive sensing (Sargsyan et al., 2014) and sparse regression (Blatman & Sudret, 2008, 2010) proved capability and efficiency in many prior studies using complex models with high dimensions, up to 80 dimensions (Sargsyan et al., 2014). However, these studies avoided the calculation of fully distributed problems by assuming the spatial variability of parameters to be homogeneous. Second, for the parameter specification, any optimization technique applied to high-dimensional problems could be relevant. For example, one of the large-scale optimization algorithms, the competitive swarm optimizer (Cheng & Jin, 2015) was employed up to the dimension of 5,000. These algorithms have been successfully optimized for problems of very large scale, but their optimizations have been applied to simple analytical functions rather than (hydrologic or meteorological) models. To our knowledge, the number of dimensions has not yet been high in problems of hydrologic optimization, in which the dimension order is almost identical to the number of parameters. The spatial variability of parameters is not fully addressed in most studies, although a "multiplier" concept (Pokhrel et al., 2008). Lastly, EnKF is made possible in problems of higher dimensionality through covariance localization. It is mainly applied in meteorological models with many parameters, and the number of dimensions can be up to the order of millions, for example, 2,592,000 (Fujita et al., 2007). The localization technique was able to reduce the dimensions efficiently.

6. Conclusions

This study presents a new robust, accurate, and efficient modeling framework that consists of the novel integration of three individual techniques: surrogate modeling, parameter inference, and data assimilation. This unified framework is suited for ensemble flood forecasts quantifying prediction uncertainty. The strengths of each technique are (i) the use of PCE offers significant computational savings, (ii) the inference of parameters before data assimilation allows for faster convergence, smaller uncertainties, and greater accuracy of the end results, and (iii) the Kalman filters assimilate errors that occur in real-time flood forecasting. Based on the results of the 18 refined approaches according to the permutations of the above methods, the following conclusions can be drawn:

- Of the two methods for PCE construction, only PCE-I (constructed based on prior, uniform distributions) is acceptable for forecasting, although both methods reproduce observations of the calibration period well. Note that PCE-II (constructed based on posterior distributions) does not provide satisfactory results, even when coupled with other inference and assimilation techniques. The results obtained from PCE-I are similar and, in some cases, even superior to those based on the original deterministic NAM model. The PCE used is a single model constructed before the forecast period and thus does not change over time—this is a unique feature different from previous studies in which PCE was rebuilt at each calibration or forecasting time step.

- Especially for short-range forecasting, model parameter input and state initialization play a crucial role. In some previous studies, posterior distributions were employed to derive a parameter ensemble before forecasting, but the effect of such parameter specification was not quantified for the data assimilation. *Selected* parameter specification (made through the GLUE framework in this study) offers improved accuracy and predictability of forecast outcomes over the *Random* parameter specification. However, it is less computationally efficient, and the issue is expected to be especially problematic when using complex deterministic models.

- The usefulness of single and dual EnKFs is demonstrated through comparisons of the 18 approaches. Both techniques have excellent overall performance, but the dual EnKF showed a slightly better performance than the single EnKF. There was a remarkable improvement in reproducing the hydrograph peak values (Table 4). In the absence of assimilation, the *Selected* approach offers superior results and if it cannot be used, data assimilation must be applied.

- The computational time discussed in this study consists of three principal components: surrogate building time, parameter inference time, and data assimilation time. Our conclusions may marginally vary depending on the particular model used and the region in which it is applied, but here the efficiency improvement from using the surrogate modeling technique overwhelms any efficiency deterioration derived from the other two components. That is, the use of the metamodel makes it possible to effectively address computational efficiency. This feasibility is maximized when many ensemble outcomes are needed and when complex, physically based models should be simulated.
- From the comprehensive analyses presented above, A15 is our first choice and A14 is the second. When only a deterministic model is used, we recommend A12 (or A11). Using the unified framework developed here, real-time and ensemble flood forecasting are promising directions, allowing for satisfactory measures of accuracy, predictability, and efficiency. Ultimately, the framework developed in this study contributes to a shift in modeling paradigm arguing that complex, high-fidelity, physical hydrologic and hydraulic models should be increasingly adopted for real-time and ensemble flood forecasting

Appendix A: The Construction of PCE Surrogate Model

A.1. The Determination of PCE Coefficients

The least squares regression method is employed to establish the PCE coefficients:

$$\varepsilon = \mathrm{argmin}_{\varepsilon \in \mathbb{R}^{|A|}} \mathbb{E}\left[\left(\boldsymbol{Y_t} - \sum_{\boldsymbol{\alpha} \in A} \varepsilon_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}_t) \right)^2 \right] \tag{A1}$$

where $\mathscr{X}_t = \left\{ \mathscr{X}_t^{(1)}, ..., \mathscr{X}_t^{(N)} \right\}$ consists of N sets of input variables $\boldsymbol{X}_t$ (the set $\mathscr{X}_t$ is called the experimental design), and $\mathscr{Y}_t = \left\{ \mathscr{M}\left(\mathscr{X}_t^{(1)} \right), ..., \mathscr{M}\left(\mathscr{X}_t^{(N)} \right) \right\}$ be the corresponding model evaluations $\left\{ \mathscr{Y}_t^{(k)} = \mathscr{M}\left(\mathscr{X}_t^{(k)} \right), k = 1, ..., N \right\}$. The estimates of the PCE coefficients are thus given by

$$\widehat{\varepsilon} = \mathrm{argmin}_{\varepsilon \in \mathbb{R}^{|A|}} \frac{1}{N} \sum_{k=1}^{N} \left(\mathscr{Y}_t^{(k)} - \sum_{\boldsymbol{\alpha} \in A} \varepsilon_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}\left(\mathscr{X}_t^{(k)} \right) \right)^2 \tag{A2}$$

which is equivalent to

$$\widehat{\varepsilon} = \left(\mathbf{F}^{\mathrm{T}} \mathbf{F} \right)^{-1} \mathbf{F}^{\mathrm{T}} \mathscr{Y}_t \tag{A3}$$

where $\mathbf{F}$ is so-called the *information matrix* of size $N \times |A|$ whose elements are defined as

$$\mathbf{F}_{k,l} = \Psi_l\left(\mathscr{X}_t^{(k)} \right) \qquad k = 1, ..., N; l = 0, ..., \mathrm{card}\, A - 1 \tag{A4}$$

Once a PCE model is derived, the prediction using the model is extremely simple and straightforward: Input the values of model input to equation (2) and then obtain the values of model response $\boldsymbol{Y_t}$.

A.2. PCE Error Estimates

The leave-one-out cross-validation error (*LOO*) was designed to overcome the overfitting limitation of normalized empirical error by using cross validation (Blatman & Sudret, 2010). In this study, after the number of sets $\mathscr{X}_t$ is defined, the *LOO* is used to determine the polynomial degree. The leave-one-out cross-validation error can be written as follows:

$$LOO = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{\mathscr{M}\left(\mathscr{X}_t^{(k)} \right) - \mathscr{M}^{PCE}\left(\mathscr{X}_t^{(k)} \right)}{1 - \hbar_k} \right)^2 \tag{A5}$$

where $\hbar_k$ is the kth diagonal term of the matrix $\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$.

Several software tools are currently available for research purposes to carry a range of UQ tasks, including PCE regression, for example, the MIT Uncertainty Quantification Library (Parno et al., n.d.), the Uncertainty Quantification Toolkit (Debusschere et al., 2016), Dakota (Eldred et al., 2010), Chaospy (Feinberg & Langtangen, 2015), and the UQLab (Marelli & Sudret, 2017). The latter libraries are used in this study.

Appendix B: Likelihood Functions Used in GLUE

Nash-Sutcliffe efficiency (*NSE*, [-])

$$NSE^i = 1 - \frac{\sum_{t=1}^{T} \left(y_t^{Obs} - y_t^i\right)^2}{\sum_{t=1}^{T} \left(y_t^{Obs} - y_{mean}^i\right)^2}, \quad i = 1, ..., n \tag{B1}$$

Peak error (*PE*, [%])

$$PE^i = \frac{\left|y_{max}^{Obs} - y_{max}^i\right|}{y_{max}^{Obs}} \times 100, \qquad i = 1, ..., n \tag{B2}$$

Volume error (*VE*, [%])

$$VE^i = \frac{\left|V^{Obs} - V^i\right|}{V^{Obs}} \times 100, \qquad i = 1, ..., n \tag{B3}$$

where V is the total volume of hydrograph.

Appendix C: Ensemble Kalman Filter (EnKF)

C.1. States Updated

An ensemble of state vector, x consisting of n by M_S is propagated through **Model** of both deterministic model and PCE models, such that each state vector represents one realization of the model states. Then, the state forecast is made for each ensemble member as follows (forecast step):

$$x_t^{i-} = f\left(x_{t-1}^{i+}, u_t, \theta^i\right) + w_t^i, \qquad i = 1, ..., n \tag{C1}$$

where x_t^{i-} is the ith forecasted states vector at time t, x_{t-1}^{i+} is the ith updated states vector at time $t-1$, M_S is the number of model states $x = \{x_j, \ j = 1, ..., M_s\}$, and n is the number of ensemble members. The nonlinear propagator $f(\cdot)$ contains M_I model input vector u_t, $\{u_{1,t}, ..., u_{M_I,t}\}$ and the ith model parameter vector θ^i corresponding to the model state x_{t-1}^{i+}. The term w_t^i is the ith model error and presents all uncertainty related to model structure, forcing data, and model parameter (Moradkhani, Sorooshian, et al., 2005). In this study, the model error is represented by the uncertainty of model parameters.

Suppose that the actual observation (y_{t+1}^{Obs}) is taken at time $t+1$ and that we intend to assimilate the vector of observations into the model. The predicted output of model, y_{t+1}^i at time $t+1$ is computed with the propagator $h(\cdot)$ as a function of θ^i, u_{t+1}, and x_t^{i-}, which can be written as follows:

$$y_{t+1}^i = h\left(x_t^{i-}, u_{t+1}, \theta^i\right) \tag{C2}$$

To represent the error statistics in the forecast step, we assume that at time $t+1$, we have an ensemble of n forecasted states, $x_t^- \triangleq \left(x_t^{1-}, ..., x_t^{n-}\right)$ and an ensemble of n forecasted outputs, $y_{t+1} \triangleq \left(y_{t+1}^1, ..., y_{t+1}^n\right)$. Then the ensemble means of forecasted state $(\overline{x}_t^-)$ and the ensemble mean of forecasted output $(\overline{y}_{t+1})$ are estimated by

$$\overline{x}_t^- \triangleq \frac{1}{n} \sum_{i=1}^{n} x_t^{i-} \tag{C3}$$

$$\overline{y}_{t+1} \triangleq \frac{1}{n} \sum_{i=1}^{n} y_{t+1}^i \tag{C4}$$

Then, we define the ensemble error matrix of forecasted state, E_{t+1}^- around the ensemble mean by

$$E_{t+1}^{-} \triangleq \left[\boldsymbol{x}_t^{1-} - \overline{\boldsymbol{x}}_t^{-} \dots \boldsymbol{x}_t^{n-} - \overline{\boldsymbol{x}}_t^{-} \right] \tag{C5}$$

and the ensemble of output error matrix, E_{t+1}^{y} is as follows:

$$E_{t+1}^{y} \triangleq \left[y_{t+1}^{1} - \overline{y}_{t+1} \dots y_{t+1}^{n} - \overline{y}_{t+1} \right] \tag{C6}$$

The error covariance matrix is calculated including

- The error covariance matrix of ensemble forecast state

$$Q_{t+1}^{x} = \frac{1}{n-1} E_{t+1}^{-} \left(E_{t+1}^{-} \right)^{\mathrm{T}} \tag{C7}$$

- The error covariance matrix of model output

$$Q_{t+1}^{y} = \frac{1}{n-1} E_{t+1}^{y} \left(E_{t+1}^{y} \right)^{\mathrm{T}} \tag{C8}$$

- The forecast cross covariance of the states and output

$$Q_{t+1}^{xy} = \frac{1}{n-1} E_{t+1}^{-} \left(E_{t+1}^{y} \right)^{\mathrm{T}} \tag{C9}$$

In order for the EnKF to maintain sufficient spreads in ensemble and to prevent from filter divergence (Whitaker & Hamill, 2002), observations should be treated as random variables. At each time, an observation is perturbed by adding noise drawn from a Gaussian distribution of mean zero and predefined covariance (Burgers et al., 1998). Thus, in the updated step, the forecasted state set $\boldsymbol{x}_{t+1}^{i-}$ is updated using the Kalman gain K_{t+1}^{x} as follows:

$$\boldsymbol{x}_t^{i+} = \boldsymbol{x}_t^{i-} + K_{t+1}^{x} \left(y_{t+1}^{Obs,i} - y_{t+1}^{i} \right) \tag{C10}$$

where $y_{t+1}^{Obs,i}$ is the ith trajectory of the observation replicates generated by adding to the actual observation (y_{t+1}^{Obs}) error, η (i.e., a perturbation to observation) that has zero mean and the covariance, $E_{t+1}^{y^{Obs}}$, which is determined in section 2.1.4, as follows:

$$y_{t+1}^{Obs,i} = y_{t+1}^{Obs} + \eta_{t+1}^{i}, \quad \eta_{t+1}^{i} \sim N \left(0, E_{t+1}^{y^{Obs}} \right) \tag{C11}$$

The Kalman gain matrix can be calculated by

$$K_{t+1}^{x} = Q_{t+1}^{xy} \left[Q_{t+1}^{y} + Q_{t+1}^{Obs} \right]^{-1} \tag{C12}$$

where Q_{t+1}^{Obs} is the covariance matrix of the observation, $y_{t+1}^{Obs,i}$, which is defined similar to Q_{t+1}^{y}.

$$Q_{t+1}^{Obs} = \frac{1}{n-1} E_{t+1}^{Obs} \left(E_{t+1}^{Obs} \right)^{\mathrm{T}} \tag{C.13}$$

$$E_{t+1}^{Obs} \triangleq \left[y_{t+1}^{Obs,1} - y_{t+1}^{Obs} \dots y_{t+1}^{Obs,n} - y_{t+1}^{Obs} \right] \tag{C14}$$

C.2. Dual Parameters-States Updated

The dual EnKF requires two interactive and parallel filters for the states and parameters estimation (Moradkhani, Sorooshian, et al., 2005). The parameters are first updated and then the states. In order to extend the applicability of the single EnKF to the simultaneous parameters-states EnKF, one needs to treat the ensemble size of parameter sets similar to the model state. However, the parameter values are not changed after the forecast step:

$$\theta_{t+1}^{i-} = \theta_t^{i+} \tag{C15}$$

Using the parameters forecasted and the replicates of forcing data, states of the ensemble model and model prediction are computed as follows:

$$\boldsymbol{x}_t^{i-} = f\left(\boldsymbol{x}_{t-1}^{i+}, \boldsymbol{u}_t, \theta_{t+1}^{i-}\right) + w_t^i, \qquad i = 1, \dots n \tag{C16}$$

$$y_{t+1}^i = h\left(\boldsymbol{x}_t^{i-}, \boldsymbol{u}_{t+1}, \theta_{t+1}^{i-}\right) \tag{C17}$$

Updating the ensemble parameter member is made

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^{\theta}\left(y_{t+1}^{Obs,i} - y_{t+1}^i\right) \tag{C18}$$

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

$$K_{t+1}^{\theta} = Q_{t+1}^{\theta y}\left[Q_{t+1}^y + Q_{t+1}^{Obs}\right]^{-1} \tag{C19}$$

where $Q_{t+1}^{\theta y}$ is the cross-covariance matrix of model parameters and model output. Now use the updated parameter θ_{t+1}^{i+} to the step given in Appendix C1 to update the ensemble model states simultaneously.

Appendix D: Sobol′ Sensitivity Analysis

Sobol′ method is a variance-based sensitivity analysis that identifies parameter sensitivities by evaluating the variance of model output (y) due to the variability of individual parameters and their parameter interactions (Crestaux et al., 2009; Saltelli, 2002; Sobol′, 2001). Instead of the model output, model performance measures (e.g., *NSE*, *PE*, and *VE* in this study) can be used (Tang et al., 2007). The total variance, $D(y)$ is decomposed as follows:

$$D(y) = \sum_{a=1}^{M_P} D_a + \sum_{a < b} D_{ab} + \dots + D_{1 \dots M_P} \tag{D1}$$

where D_a is the variance of y due to the changes of ath model parameter, θ_a, denoting the first order contribution to $D(y)$; D_{ab} is the variance of y due to the pairwise interactions of ath and bth parameters, referring to the second order contribution.

The first (S_a) and total ($S_{Total,a}$) order Sobol′ sensitivity indices can be, respectively, expressed as follows:

$$S_a = \frac{D_a}{D(y)} \tag{D2}$$

$$S_{Total,a} = 1 - \frac{D_{\sim a}}{D(y)} \tag{D3}$$

where $D_{\sim a}$ is the variance averaged over the contributions resulting from all parameters except for θ_a.

References

Abbaszadeh, P., Moradkhani, H., & Yan, H. (2018). Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. *Advances in Water Resources*, *111*, 192–204. https://doi.org/10.1016/j.advwatres.2017.11.011

Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, *43*, W01403. https://doi.org/10.1029/2005wr004745

Associated Programme on Flood Management (APFM) (2013). Integrated flood management tools series: Flood forecasting and early warning. *Rep.*, Associated Programme on Flood Management.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. https://doi.org/10.1109/78.978374

Ballio, F., & Guadagnini, A. (2004). Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. *Water Resources Research*, *40*, W04603. https://doi.org/10.1029/2003wr002876

Bauser, H. H., Berg, D., Klein, O., & Roth, K. (2018). Inflation method for ensemble Kalman filter in soil hydrology. *Hydrology and Earth System Sciences Discussions*, *1-18*. https://doi.org/10.5194/hess-2018-74

Berveiller, M., Sudret, B., & Lemaire, M. (2006). Stochastic finite elements: A non intrusive approach by regression. *Eur. J. Comput. Mech.*, *15*.

Beven, K. (1989). Changing ideas in hydrology—The case of physically-based models. *Journal of Hydrology*, *105*(1-2), 157–172. https://doi.org/10.1016/0022-1694(89)90101-7

Beven, K. (2000). Uniqueness of place and non-uniqueness of models in assessing predictive uncertainty, paper presented at Computational methods in water resources—Volume 2—Computational methods, surface water systems and hydrology.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1-4), 11–29. https://doi.org/10.1016/s0022-1694(01)00421-8

Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., & Zyvoloski, G. A. (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources*, *31*(4), 630–648. https://doi.org/10.1016/j.advwatres.2007.12.003

Blatman, G., & Sudret, B. (2008). Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique*, *336*(6), 518–523. https://doi.org/10.1016/j.crme.2008.02.013

Blatman, G., & Sudret, B. (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, *25*(2), 183–197. https://doi.org/10.1016/j.probengmech.2009.10.003

Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, *230*(6), 2345–2367. https://doi.org/10.1016/j.jcp.2010.12.021

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2

Burgers, G., van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:asitek>2.0.co;2

Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, *298*(1-4), 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042

Caflisch, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, *7*, 1–49.

Cameron, D., Beven, K., Tawn, J., & Naden, P. (2000). Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation). *Hydrology and Earth System Sciences Discussions*, *4*(1), 23–34.

Chen, H., Yang, D., Hong, Y., Gourley, J. J., & Zhang, Y. (2013). Hydrological data assimilation with the ensemble square-root-filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Advances in Water Resources*, *59*, 209–220. https://doi.org/10.1016/j.advwatres.2013.06.010

Cheng, R., & Jin, Y. (2015). A competitive swarm optimizer for large scale optimization. *IEEE Trans Cybern*, *45*(2), 191–204. https://doi.org/10.1109/TCYB.2014.2322602

Choi, H. T., & Beven, K. (2007). Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology*, *332*(3-4), 316–336. https://doi.org/10.1016/j.jhydrol.2006.07.012

Cintra, R. S., & Velho, H. F. d. C. (2018). Data assimilation by artificial neural networks for an atmospheric General Circulation Model. In *Advanced Applications for Artificial Neural Networks* (Chap. 14, pp. 266–286). Rijeka, Croatia: IN-TECH. https://doi.org/10.5772/intechopen.70791

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (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. https://doi.org/10.1016/j.advwatres.2008.06.005

Crestaux, T., Le Maître, O., & Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, *94*(7), 1161–1172. https://doi.org/10.1016/j.ress.2008.10.008

Danish Hydraulic Institute (2014). *DHI Mike 11: A modelling system for rivers and channels*. Reference Manual, Danish Hydraulic Institute (DHI) Water & Environment: Hørsholm, Denmark.

Davis, P. J., & Rabinowitz, P. (2007). *Methods of numerical integration*. Courier: Corporation.

Davison, B., Fortin, V., Pietroniro, A., Yau, M. K., & Leconte, R. (2017). Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction. *Hydrology and Earth System Sciences Discussions*, *1-38*. https://doi.org/10.5194/hess-2017-482

Debusschere, B. J., Sargsyan, K., Najm, H. N., & Safta, C. (2016). The Uncertainty Quantification Toolkit (UQTk). In *in Handbook of Uncertainty Quantification* (Chap. 53, pp. 1807–1828). Livermore, CA: Sandia National Laboratories. https://doi.org/10.1007/978-3-319-11259-6_56-1

DeChant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, *48*, W04518. https://doi.org/10.1029/2011wr011011

DeChant, C. M., & Moradkhani, H. (2014). Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *Journal of Hydrology*, *519*, 2967–2977. https://doi.org/10.1016/j.jhydrol.2014.05.045

Dwelle, M. C., Kim, J., Sargsyan, K., & Ivanov, V. Y. (2019). Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification. *Advances in Water Resources*. https://doi.org/10.1016/j.advwatres.2019.01.002

Eldred, M. S., Dalbey, K. R., Bohnhoff, W. J., Adams, B. M., Swiler, L. P., Hough, P. D., et al. (2010). *Dakota, A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.2 Theory ManualRep* (Vol. *2007*). Albuquerque, NM: Sandia National Laboratories.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143. https://doi.org/10.1029/94jc00572

Fan, Y., Huang, W., Huang, G. H., Huang, K., & Zhou, X. (2014). A PCM-based stochastic hydrological model for uncertainty quantification in watershed systems. *Stochastic Environmental Research and Risk Assessment*, *29*(3), 915–927. https://doi.org/10.1007/s00477-014-0954-8

Fan, Y. R., Huang, G. H., Baetz, B. W., Li, Y. P., Huang, K., Li, Z., et al. (2016). Parameter uncertainty and temporal dynamics of sensitivity for hydrologic models: A hybrid sequential data assimilation and probabilistic collocation method. *Environmental Modelling & Software*, *86*, 30–49. https://doi.org/10.1016/j.envsoft.2016.09.012

Fatichi, S., Katul, G. G., Ivanov, V. Y., Pappas, C., Paschalis, A., Consolo, A., et al. (2015). Abiotic and biotic controls of soil moisture spatiotemporal variability and the occurrence of hysteresis. *Water Resources Research*, *51*, 3505–3524. https://doi.org/10.1002/2014wr016102

Feinberg, J., & Langtangen, H. P. (2015). Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, *11*, 46–57. https://doi.org/10.1016/j.jocs.2015.08.008

Fujita, T., Stensrud, D. J., & Dowell, D. C. (2007). Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Monthly Weather Review*, *135*(5), 1846–1868. https://doi.org/10.1175/mwr3391.1

Ghanem, R. G., & Spanos, P. D. (1991). Stochastic finite elements: A spectral approach. *Springer, Verlag New York*. https://doi.org/10.1007/978-1-4612-3094-6

Gharamti, M. E., Hoteit, I., & Valstar, J. (2013). Dual states estimation of a subsurface flow-transport coupled model using ensemble Kalman filtering. *Advances in Water Resources*, *60*, 75–88. https://doi.org/10.1016/j.advwatres.2013.07.011

Ghiocel, D. M., & Ghanem, R. G. (2002). Stochastic finite-element analysis of seismic soil-structure interaction. *Journal of Engineering Mechanics*, *128*(1), 66–77. https://doi.org/10.1061/(asce)0733-9399(2002)128:1(66)

Giannakis, D., & Majda, A. J. (2012). Quantifying the predictive skill in long-range forecasting. *Part II: Model error in coarse-grained Markov models with application to ocean-circulation regimes, Journal of Climate*, *25*(6), 1814–1826. https://doi.org/10.1175/jcli-d-11-00110.1

He, M., Hogue, T. S., Margulis, S. A., & Franz, K. J. (2012). An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. *Hydrology and Earth System Sciences*, *16*(3), 815–831. https://doi.org/10.5194/hess-16-815-2012

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, *81*(1), 23–69. https://doi.org/10.1016/S0951-8320(03)00058-9

Heo, K.-Y., Ha, K.-J., Yun, K.-S., Lee, S.-S., Kim, H.-J., & Wang, B. (2014). Methods for uncertainty assessment of climate models and model predictions over East Asia. *International Journal of Climatology*, *34*(2), 377–390. https://doi.org/10.1002/joc.3692

Hossain, F., & Anagnostou, E. N. (2005). Assessment of a stochastic interpolation based parameter sampling scheme for efficient uncertainty analyses of hydrologic models. *Computers & Geosciences*, *31*(4), 497–512. https://doi.org/10.1016/j.cageo.2004.11.001

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. https://doi.org/10.1175/mwr-d-15-0440.1

Ivanov, V. Y., Fatichi, S., Jenerette, G. D., Espeleta, J. F., Troch, P. A., & Huxman, T. E. (2010). Hysteresis of soil moisture spatial heterogeneity and the "homogenizing" effect of vegetation. *Water Resources Research*, *46*, W09521. https://doi.org/10.1029/2009wr008611

Jin, X., Xu, C.-Y., Zhang, Q., & Singh, V. P. (2010). Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, *383*(3-4), 147–155. https://doi.org/10.1016/j.jhydrol.2009.12.028

Kim, J., Dwelle, M. C., Kampf, S. K., Fatichi, S., & Ivanov, V. Y. (2016). On the non-uniqueness of the hydro-geomorphic responses in a zero-order catchment with respect to soil moisture. *Advances in Water Resources*, *92*, 73–89. https://doi.org/10.1016/j.advwatres.2016.03.019

Kim, J., & Ivanov, V. Y. (2014). On the nonuniqueness of sediment yield at the catchment scale: The effects of soil antecedent conditions and surface shield. *Water Resources Research*, *50*, 1025–1045. https://doi.org/10.1002/2013wr014580

Kim, J., & Ivanov, V. Y. (2015). A holistic, multi-scale dynamic downscaling framework for climate impact assessments and challenges of addressing finer-scale watershed dynamics. *Journal of Hydrology*, *522*, 645–660. https://doi.org/10.1016/j.jhydrol.2015.01.025

Kim, J., Ivanov, V. Y., & Fatichi, S. (2016a). Environmental stochasticity controls soil erosion variability. *Scientific Reports*, *6*, 22065. https://doi.org/10.1038/srep22065

Kim, J., Ivanov, V. Y., & Fatichi, S. (2016b). Soil erosion assessment—Mind the gap. *Geophysical Research Letters*, *43*, 12,446–412,456. https://doi.org/10.1002/2016gl071480

Kim, J., Tanveer, M. E., & Bae, D.-H. (2018). Quantifying climate internal variability using an hourly ensemble generator over South Korea. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3037–3051. https://doi.org/10.1007/s00477-018-1607-0

Kitanidis, P. K., & Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 1. *Analysis of uncertainty, Water Resources Research*, *16*(6), 1025–1033. https://doi.org/10.1029/WR016i006p01025

Kleeman, R. (2002). Measuring dynamical prediction utility using relative entropy, *Journal of the Atmospheric Sciences*, *59*(13), 2057-2072, doi:10.1175/1520-0469(2002)059 < 2057:mdpuur>2.0.co;2.

Kullback, S. (1997). *Information theory and statistics*. North Chelmsford, MA: Courier Corporation.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., & Terray, L. (2014). Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin. *Water Resources Research*, *50*, 3317–3341. https://doi.org/10.1002/2013wr014897

Le Maître, O. P., Reagan, M. T., Najm, H. N., Ghanem, R. G., & Knio, O. M. (2002). A stochastic projection method for fluid flow. *Journal of Computational Physics*, *181*(1), 9–44. https://doi.org/10.1006/jcph.2002.7104

Li, Y., Ryu, D., Western, A. W., Wang, Q. J., Robertson, D. E., & Crow, W. T. (2014). An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting. *Journal of Hydrology*, *519*, 2722–2736. https://doi.org/10.1016/j.jhydrol.2014.08.009

Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., et al. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, *16*(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, *235*(3-4), 276–288. https://doi.org/10.1016/s0022-1694(00)00279-1

Madsen, H., & Skotner, C. (2005). Adaptive state updating in real-time river flow forecasting—A combined filtering and error forecasting procedure. *Journal of Hydrology*, *308*(1-4), 302–312. https://doi.org/10.1016/j.jhydrol.2004.10.030

Marelli, S., & Sudret, B. (2017). *UQLab user manual—Polynomial chaos expansions Rep*. Safety & Uncertainty Quantification, ETH Zurich: Chair of Risk.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239. https://doi.org/10.2307/1268522

McKenna, S. A., Doherty, J., & Hart, D. B. (2003). Non-uniqueness of inverse transmissivity field calibration and predictive transport modeling. *Journal of Hydrology*, *281*(4), 265–280. https://doi.org/10.1016/s0022-1694(03)00194-x

Mendoza, P. A., McPhee, J., & Vargas, X. (2012). Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment. *Water Resources Research*, *48*, W09532. https://doi.org/10.1029/2011wr011089

Meng, J., & Li, H. (2018). Uncertainty quantification for subsurface flow and transport: Coping with nonlinearity/irregularity via polynomial chaos surrogate and machine learning. *Water Resources Research*. https://doi.org/10.1029/2018wr022676

Mockler, E. M., Chun, K. P., Sapriza-Azuri, G., Bruen, M., & Wheater, H. S. (2016). Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations. *Advances in Water Resources*, *97*, 299–313. https://doi.org/10.1016/j.advwatres.2016.10.008

Mondal, A., & Mujumdar, P. P. (2012). On the basin-scale detection and attribution of human-induced climate change in monsoon precipitation and streamflow. *Water Resources Research*, *48*. https://doi.org/10.1029/2011wr011468

Moradkhani, H., DeChant, C. M., & Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, *48*(12). https://doi.org/10.1029/2012wr012144

Moradkhani, H., Hsu, K.-L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*. https://doi.org/10.1029/2004WR003604

Moradkhani, H., & Sorooshian, S. (2008). General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In S. Sorooshian, K.-L. Hsu, E. Coppola, B. Tomassetti, M. Verdecchia, & G. Visconti (Eds.), *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrologic Models* (pp. 1–24). Berlin: Springer. https://doi.org/10.1007/978-3-540-77843-1_1

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Nga, P. H., Takara, K., & Son, N. H. (2015). Flood hazard impact analysis in the downstream of Vu Gia-Thu Bon River System, Quang Nam Province, Central Vietnam. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, *71*(4), I_157–I_162. https://doi.org/10.2208/jscejhe.71.I_157

Nielsen, S. A., & Hansen, E. (1973). Numerical simulation of the rainfall-runoff process on a daily basis. *Hydrology Research*, *4*(3), 171–190.

Nikiema, O., & Laprise, R. (2011). Budget study of the internal variability in ensemble simulations of the Canadian Regional Climate Model at the seasonal scale. *Journal of Geophysical Research*, *116*(D16). https://doi.org/10.1029/2011jd015841

O'Brien, R. J., Misstear, B. D., Gill, L. W., Deakin, J. L., & Flynn, R. (2013). Developing an integrated hydrograph separation and lumped modelling approach to quantifying hydrological pathways in Irish river catchments. *Journal of Hydrology*, *486*, 259–270. https://doi.org/10.1016/j.jhydrol.2013.01.034

Parno, M., Davis, A., & Conrad, P. n.d. MIT Uncertainty Quantification (MUQ) library, edited.

Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*, *44*. https://doi.org/10.1029/2007wr006615

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, *22*(3), 579–590. https://doi.org/10.1111/0272-4332.00040

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., & Thornton, P. (2014). Dimensionality reduction for complex models via Bayesian compressive sensing. *International Journal for Uncertainty Quantification*, *4*(1), 63–93. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2013006821

Sene, K. (2008). Flood warning, forecasting and emergency response. *Springer Science & Business Media*. https://doi.org/10.1007/978-3-540-77853-0

Shen, Z. Y., Chen, L., & Chen, T. (2012). Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: A case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrology and Earth System Sciences*, *16*(1), 121–132. https://doi.org/10.5194/hess-16-121-2012

Shukla, J., DelSole, T., Fennessy, M., Kinter, J., & Paolino, D. (2006). Climate model fidelity and projections of climate change. *Geophysical Research Letters*, *33*. https://doi.org/10.1029/2005gl025579

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1-3), 271–280. https://doi.org/10.1016/s0378-4754(00)00270-6

Sochala, P., & Le Maître, O. P. (2013). Polynomial Chaos expansion for subsurface flows with uncertain soil parameters. *Advances in Water Resources*, *62*, 139–154. https://doi.org/10.1016/j.advwatres.2013.10.003

Sudret, B. (2007). *Uncertainty propagation and sensitivity analysis in mechanical models. Contributions to structural reliability and stochastic spectral methods*. Habilitation thesis: Universite Blaise Pascal, Clermont-Ferrand, France.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, *93*(7), 964–979. https://doi.org/10.1016/j.ress.2007.04.002

Tang, Y., Reed, P., van Werkhoven, K., & Wagener, T. (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, *43*(6). https://doi.org/10.1029/2006wr005813

Thompson, J. R., Sørenson, H. R., Gavin, H., & Refsgaard, A. (2004). Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England. *Journal of Hydrology*, *293*(1-4), 151–179. https://doi.org/10.1016/j.jhydrol.2004.01.017

Todini, E. (1999). Using phase-state modelling for inferring forecasting uncertainty in nonlinear stochastic decision schemes. *Journal of Hydroinformatics*, *1*(2), 75–82. https://doi.org/10.2166/hydro.1999.0007

Todini, E. (2004). Role and treatment of uncertainty in real-time flood forecasting. *Hydrological Processes*, *18*(14), 2743–2746. https://doi.org/10.1002/hyp.5687

Tran, V. N., & Kim, J. (2019). Quantification of predictive uncertainty with a metamodel: Toward more efficient hydrologic simulations. *Stochastic Environmental Research and Risk Assessment*, *33*(7), 1453–1476. https://doi.org/10.1007/s00477-019-01703-0

United Nations Development Programme (UNDP) (1999). *Viet Nam: Flood damage summary 06 Nov 1999, edited, ReliefWeb*. Hanoi, Vietnam.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, *41*(1). https://doi.org/10.1029/2004wr003059

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2008). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026. https://doi.org/10.1007/s00477-008-0274-y

Wang, D., Chen, Y., & Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research*, *45*(11). https://doi.org/10.1029/2008wr007401

Wang, S., Ancell, B. C., Huang, G. H., & Baetz, B. W. (2018). Improving robustness of hydrologic ensemble predictions through probabilistic pre- and post-processing in sequential data assimilation. *Water Resources Research*, *54*(3), 2129–2151. https://doi.org/10.1002/2018wr022546

Wang, S., Huang, G. H., Baetz, B. W., & Ancell, B. C. (2017). Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle Markov chain Monte Carlo and factorial polynomial chaos expansion. *Journal of Hydrology*, *548*, 484–497. https://doi.org/10.1016/j.jhydrol.2017.03.027

Wang, S., Huang, G. H., Baetz, B. W., & Huang, W. (2015). A polynomial chaos ensemble hydrologic prediction system for efficient parameter inference and robust uncertainty assessment. *Journal of Hydrology*, *530*, 716–733. https://doi.org/10.1016/j.jhydrol.2015.10.021

Weerts, A. H., & El Serafy, G. Y. H. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, *42*(9). https://doi.org/10.1029/2005wr004093

Wei, C., & Dewoolkar, M. M. (2006). Formulation of capillary hysteresis with internal state variables. *Water Resources Research*, *42*. https://doi.org/10.1029/2005wr004594

Whitaker, J. S. (2012). Developments in ensemble data assimilation, paper presented at Proceedings of the Seminar on Data assimilation for atmosphere and ocean, ECMWF, 6-9. *September*, *2011*.

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:edawpo>2.0.co;2

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, *60*(4), 897. https://doi.org/10.2307/2371268

Wu, B., Zheng, Y., Tian, Y., Wu, X., Yao, Y., Han, F., et al. (2014). Systematic assessment of the uncertainty in integrated surface water-groundwater modeling based on the probabilistic collocation method. *Water Resources Research*, *50*, 5848–5865. https://doi.org/10.1002/2014wr015366

Xie, X., & Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources*, *33*(6), 678–690. https://doi.org/10.1016/j.advwatres.2010.03.012

Xie, X., & Zhang, D. (2013). A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. *Water Resources Research*, *49*, 7350–7365. https://doi.org/10.1002/2012wr012853

Xiu, D., & Karniadakis, G. E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, *24*(2), 619–644. https://doi.org/10.1137/s1064827501387826

Young, P. C. (2002). Advances in real-time flood forecasting. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *360*(1796), 1433–1450. https://doi.org/10.1098/rsta.2002.1008

Zahmatkesh, Z., Karamouz, M., & Nazif, S. (2015). Uncertainty based modeling of rainfall-runoff: Combined differential evolution adaptive Metropolis (DREAM) and K-means clustering. *Advances in Water Resources*, *83*, 405–420. https://doi.org/10.1016/j.advwatres.2015.06.012

Zhang, H., Hendricks Franssen, H.-J., Han, X., Vrugt, J. A., & Vereecken, H. (2017). State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrology and Earth System Sciences*, *21*(9), 4927–4958. https://doi.org/10.5194/hess-21-4927-2017

Zhang, X., Liu, P., Cheng, L., Liu, Z., & Zhao, Y. (2018). A back-fitting algorithm to improve real-time flood forecasting. *Journal of Hydrology*, *562*, 140–150. https://doi.org/10.1016/j.jhydrol.2018.04.051