ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL104464

Key Points:

- While PBMs are physics-based, the complexity of uncertainties and the high computational burden have limited their utility for predictions
- The developed novel framework integrates process-based models, surrogate, and machine learning (ML) models to predict ensemble flood attributes with error quantification
- A novel probabilistic ML model partitions the errors into reducible and irreducible types, also quantifying their distributions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

V. Y. Ivanov,
ivanov@umich.edu

Citation:

Tran, V. N., Ivanov, V. Y., Xu, D., & Kim, J. (2023). Closing in on hydrologic predictive accuracy: Combining the strengths of high-fidelity and physics-agnostic models. *Geophysical Research Letters*, *50*, e2023GL104464. https://doi.org/10.1029/2023GL104464

Received 12 MAY 2023
Accepted 21 AUG 2023

Author Contributions:

Conceptualization: Vinh Ngoc Tran, Valeriy Y. Ivanov
Data curation: Vinh Ngoc Tran, Donghui Xu
Formal analysis: Vinh Ngoc Tran, Valeriy Y. Ivanov
Funding acquisition: Valeriy Y. Ivanov, Jongho Kim
Investigation: Vinh Ngoc Tran, Valeriy Y. Ivanov
Methodology: Vinh Ngoc Tran
Software: Vinh Ngoc Tran, Donghui Xu

Closing in on Hydrologic Predictive Accuracy: Combining the Strengths of High-Fidelity and Physics-Agnostic Models

Vinh Ngoc Tran[1], Valeriy Y. Ivanov[1], Donghui Xu[2], and Jongho Kim[3]

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA, [2]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA, [3]School of Civil and Environmental Engineering, University of Ulsan, Ulsan, South Korea

Abstract Applications of process-based models (PBM) for predictions are confounded by multiple uncertainties and computational burdens, resulting in appreciable errors. A novel modeling framework combining a high-fidelity PBM with surrogate and machine learning (ML) models is developed to tackle these challenges and applied for streamflow prediction. A surrogate model permits high computational efficiency of a PBM solution at a minimum loss of its accuracy. A novel probabilistic ML model partitions the PBM-surrogate prediction errors into reducible and irreducible types, quantifying their distributions that arise due to both explicitly perceived uncertainties (such as parametric) or those that are entirely hidden to the modeler (not included or unexpected). Using this approach, we demonstrate a substantial improvement of streamflow predictive accuracy for a case study urbanized watershed. Such a framework provides an efficient solution combining the strengths of high-fidelity and physics-agnostic models for a wide range of prediction problems in geosciences.

Plain Language Summary This study proposes a new framework that combines three different modeling techniques to make flood forecasting more accurate. The framework combines the strengths of (a) complex models (or process-based models, PBMs) based on our understanding of relevant processes that can reproduce measurable quantities; (b) simpler models that are designed to mimic PBM's solutions—known as surrogate models—and make predictions within a few seconds; and (c) machine learning models that can detect relationships among variables using only data, improve the accuracy of prediction, and provide estimates of prediction uncertainty. The framework is tested in an urbanized watershed and shows a significant improvement in both computational efficiency and accuracy of streamflow prediction. Ultimately, the proposed framework is a novel powerful solution that combines the latest advances in different types of modeling approaches to solve prediction problems in geosciences. Its adaptability and efficiency make it suitable for a wide range of situations.

1. Introduction

Progress in computing technology and availability of process-level and watershed data have led to advances in the development of process-based models (PBM) (Beven, 1989; Fatichi et al., 2016; Ivanov et al., 2004a; Maxwell et al., 2014). PBMs are viewed to be models that can mimic the distribution of different forms of mechanical energy that propel the hydrologic and hydraulic dynamics, leading to the conservation of energy, momentum, and mass. Thus, such models have the capacity to provide estimates of internal states and fluxes that are measurable (Fatichi et al., 2016). With a growing physical representativeness of PBMs, it has been an expectation that the prediction error would shrink for well-constrained problems (Figure 1a) (Carpenter & Georgakakos, 2006; Fatichi et al., 2016).

Even the most advanced PBM frameworks however have drawbacks. First, they exhibit extreme computational burden when using high spatial-temporal resolutions (Fatichi et al., 2016; Smith et al., 2004). This is regarded as one of the most formidable challenges to their successful implementation for real-time or uncertainty-informed applications (Ivanov et al., 2021). Even if high-performance/parallel computing is possible, there are limitations in the computational scalability for complex PBMs that may include free surface flow formulations (Micaletto et al., 2022; Vivoni et al., 2011). Second, PBM's predictions often exhibit appreciable differences with respect to observations (Camporese et al., 2010; Maxwell et al., 2014; Smith et al., 2012a) due to the parameter assumptions (Fatichi et al., 2016), "the curse of dimensionality" (Dwelle et al., 2019), complexities of natural phenomena

Supervision: Valeriy Y. Ivanov
Validation: Vinh Ngoc Tran, Valeriy Y. Ivanov
Visualization: Vinh Ngoc Tran
Writing – original draft: Vinh Ngoc Tran, Valeriy Y. Ivanov
Writing – review & editing: Vinh Ngoc Tran, Valeriy Y. Ivanov, Donghui Xu, Jongho Kim

represented by hysteresis (Fatichi et al., 2015; Ivanov et al., 2010), nonuniqueness (Kim, Dwelle, et al., 2016; Kim & Ivanov, 2014), nonlinearity (Kim & Ivanov, 2015) and internal variability (Kim et al., 2016a, 2016b), and human activities (Wada et al., 2017). These differences can be attributed to contributions from uncertainties of various origins.

Uncertainty can be categorized into two types: epistemic and aleatory (Gong et al., 2013). Epistemic uncertainty is associated with model structure, parameters, and input data, while aleatoric uncertainty is related to processes randomness that can impact PBM's outputs of interest. Epistemic uncertainties generally arise from a lack of knowledge about values that exist and are fixed but poorly known. Aleatory uncertainties are random and theoretically irreducible, in so far as our ability to *predict* future conditions is concerned (e.g., initial conditions in flood-forecasting) (Helton et al., 2010). Uncertainty quantification (UQ) remains a conundrum in all PBM studies. UQ in streamflow prediction has been primarily concerned with epistemic sources (Beven, 2016; Moradkhani & Sorooshian, 2008; Vrugt et al., 2003). The same applies to studies with multiple quantities of interest (QoIs) (Dwelle et al., 2019). Attempts have been undertaken to assess aleatoric uncertainty (Gong et al., 2013), but they remain non-suitable for forecasting purposes. Overall, a general approach evaluating uncertainty contributions is warranted to improve the accuracy and interpretation of PBM outputs.

The "Big Data" revolution has enabled the development of sophisticated machine learning (ML) tools that can generate accurate QoI simulations based on proxy data (Kratzert et al., 2018; Shen, 2018). However, ML models are only data-driven, without any process representation or physical basis (Konapala et al., 2020; Lazer et al., 2014; Read et al., 2019). They can produce erroneous and unreliable predictions, particularly when extrapolating to regions outside of their training set space, for example, due to the changes in climate conditions, watershed land use/cover, topography, and infrastructure (Kim et al., 2016a; Ruhi et al., 2022; Winkler et al., 2021). This limitation of ML models does not deny their accuracy under well-defined conditions, however it also emphasizes the advantages of PBMs that can seamlessly incorporate such spatiotemporal changes to produce *inexact, but reasonable* predictions (Smith et al., 2004, 2012a, 2012b).

In this research, we introduce a novel hybrid framework that integrates the strengths of PBMs and ML models, while tackling two major challenges associated with PBMs: computational complexity and multiple uncertainties leading to insufficient predictive accuracy. While few studies have sought to leverage ML in combination with PBMs to improve the accuracy of hydrologic predictions (Ghaith & Li, 2020; Konapala et al., 2020; Kumanlioglu & Fistikoglu, 2019; Roy et al., 2023; Young et al., 2017), previous research has not quantified uncertainty of the joint ML-PBM prediction. We demonstrate the skill of such a framework for streamflow forecasting and exemplify how the first-principles of flood wave motion can be synergistically combined with physics-agnostic models to improve UQ-informed streamflow predictions. The framework is applied to a small case-study urbanized watershed in Houston but the machinery can be generalized for other cases with input data of comparable quality. Beyond hydrologic applications, other earth-science disciplines aiming to explore PBM-ML strengths can benefit from the conceptual integration of various modeling philosophies presented in this study.

2. Methods

2.1. Conceptualization of Model Error

Simulation of streamflow or, generally, any observable QoI ($\mathrm{QoI_{Obs}}$) can be expressed as $\mathrm{QoI_{Obs}} = \mathrm{QoI_{Sim}} + \boldsymbol{\varepsilon}$, where $\mathrm{QoI_{Sim}}$ is simulated by the PBM given its inputs (boundary and initial conditions, and model parameters), and $\boldsymbol{\varepsilon}$ is the model error that follows some distribution capturing the contributions from both epistemic and aleatory uncertainties (Gong et al., 2013; Montanari et al., 2009). For observed events, the error is quantified as $\boldsymbol{\varepsilon}^{\boldsymbol{o}} = \mathrm{QoI_{Obs}} - \mathrm{QoI_{Sim}}$ (Figure 1a). We posit that $\boldsymbol{\varepsilon} = \tilde{\boldsymbol{\varepsilon}} + \boldsymbol{\varepsilon}'$, where $\tilde{\boldsymbol{\varepsilon}}$ is a distribution of reducible error that can be inferred from the interrogation of observational and simulated data, and $\boldsymbol{\varepsilon}'$ is an irreducible error whose distribution can be quantified (Figures 1b–1d). We assert that both $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ can be represented with a ML model when given (a) outputs from a suitable PBM that has (b) appropriate model parameters, and (c) forcings/boundary conditions. Note that neither the PBM, nor its parameters, or forcings need not be error-/bias-free. The *likeliest* value of $\tilde{\boldsymbol{\varepsilon}}$ can be added to a PBM prediction to reduce its bias, while a distribution for $\boldsymbol{\varepsilon}'$ would provide the range for the noise that cannot be represented with a data-driven ML model (i.e., can be viewed as the ML aleatoric uncertainty). An a priori expectation is that ML-generated distributions for $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ would envelope observations.

10.1029/2023GL104464

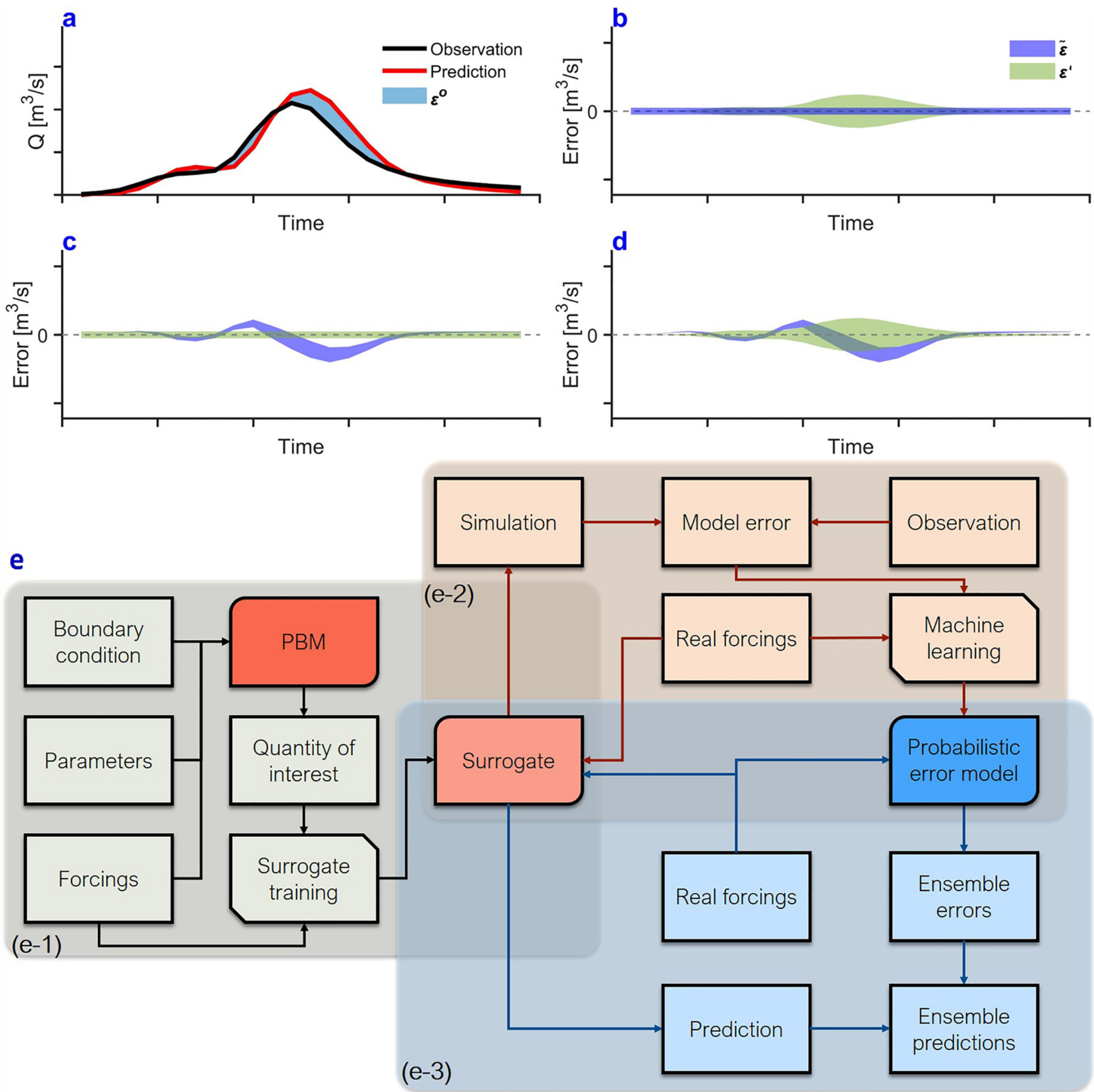

Figure 1. (a) An illustration of process-based models (PBM) model prediction, observation, and error $\boldsymbol{\varepsilon}^o$ series. Subplots (b), (c), and (d) illustrate theoretically possible permutations of the ML-model quantified error $\boldsymbol{\varepsilon}$. In (b), the PBM model prediction cannot be improved as the error is entirely irreducible ($\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'$); a confidence interval (green region) is constructed around the prediction time series in (a). In (c), the estimated error is reducible ($\boldsymbol{\varepsilon} = \tilde{\boldsymbol{\varepsilon}}$); its addition (blue region) to the prediction makes a perfect match with observation in (a). In (d), $\boldsymbol{\varepsilon} = \tilde{\boldsymbol{\varepsilon}} + \boldsymbol{\varepsilon}'$ and therefore, while the prediction can be improved, it is not fully certain and requires a confidence interval. The framework for ensemble streamflow prediction with uncertainty quantification is in (e). Box (e-1): the construction of a surrogate of PBM for simulating quantities of interest (QoIs). Various inputs are used for the PBM model. PBM forcings and outputs (i.e., QoIs) are used to construct a surrogate model(s). Box (e-2): training of a machine learning (ML) model to predict the surrogate model error. The training data are based on past rainfall series and estimates $\boldsymbol{\varepsilon}^o = Q_{\mathrm{Obs}} - Q_{\mathrm{Sim}}^{\mathrm{Surrogate}}$. Box (e-3) describes the use of the surrogate model and probabilistic ML error model to make ensemble predictions for real events. Ensemble errors derived from the ML model supplement a deterministic prediction from a surrogate model to construct ensemble predictions.

2.2. Modeling Framework

Figure 1e presents a novel modeling framework for streamflow ensemble prediction with UQ that incorporates the strengths of a state-of-the-science PBM, a probabilistic learning model, and modern advances in probabilistic ML. We describe three essential stages of the framework development and application.

The first stage is to construct a computationally efficient *surrogate* model for a PBM focusing on a specific QoI—in this study, streamflow series (Figure 1e-1). The PBM model is the TIN-Based Real Time Integrated Basin Simulator—Overland Flow Model (tRIBS-OFM) (Text S1 in Supporting Information S1) (Ivanov et al., 2004a; Kim, Warnock, et al., 2012; Kim et al., 2013) that represents both land-surface hydrology and hydrodynamics of unsteady surface flow. The surrogate model is constructed using Polynomial Chaos Expansions (PCE) (Text S3 in Supporting Information S1) (Ghanem & Spanos, 1991; Wiener, 1938) based on rainfall and streamflow (targeted QoI) data sets. Widely acknowledged as a powerful approach, PCE technique boasts a number of benefits and polynomial form also significantly reduces computational effort (Razavi et al., 2012; Sargsyan et al., 2014; Sudret, 2008).

A surrogate model was designed to simulate a variety of potential streamflow events (Tran et al., 2020; Tran & Kim, 2021). To accomplish this, we developed a synthetic data set containing rainfall and streamflow to be used as surrogate model training set (see also Section 2.4). Specifically, an ensemble of synthetic rainfall series was generated as sets of random, uncorrelated pulses. Synthetic streamflow series were then generated by tRIBS-OFM forced with the synthetic rainfall ensemble. A PCE-based surrogate model was then designed (i.e., its polynomial order is chosen) and trained using this synthetic rainfall-streamflow data set (Figure 1e-1). The inputs of the surrogate model are rainfall at the present time step and streamflow at the prior time. The simulated watershed outlet streamflow at the current time represented the QoI.

The second stage assumes the development of a probabilistic ML model to quantify $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ (Figure 1e-2). Specifically, the constructed surrogate model is run with real-world rainfall series to simulate streamflow. The errors are estimated as discrepancies with respect to observations ($\boldsymbol{\varepsilon}^o = Q_{\mathrm{Obs}} - Q_{\mathrm{Sim}}^{\mathrm{Surrogate}}$). These error estimates and the corresponding rainfall data are then used to train the probabilistic error model using a long short-term memory network (LSTM) (Hochreiter & Schmidhuber, 1997) with an attention mechanism (Vaswani et al., 2017) coupled with heteroscedastic regression (HR) (Kendall & Gal, 2017) and Monte Carlo Dropout (MCD) (Gal & Ghahramani, 2016). LSTM is viewed to be one of the most powerful MLs in hydrological applications (Alizadeh et al., 2021; Ding et al., 2020; Konapala et al., 2020; Kratzert et al., 2019). The HR serves as a supportive tool for the LSTM model to generate predictions for two distinct output variables: $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$. Ensembles of $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ are derived by aggregating the predictions from a set of LSTM models with varying architectures generated by MCD with multiple Monte Carlo draws of the dropout mask. The LSTM methodology and the framework for error ensembles are in Texts S4 and S5 in Supporting Information S1.

The process of ML training is conducted using most popular supportive techniques. Rainfall is the input variable and its look-back window size is chosen utilizing Mutual Information (Vergara & Estévez, 2014) and Hampel Test (May et al., 2008). To tackle non-stationarity of data, the discrete wavelet transform (Mallat, 1999) is implemented. Hyper-parameters are optimized through Bayesian optimization using the Gaussian process (Snoek et al., 2012) and the *k*-fold cross-validation with an early stopping technique (Yao et al., 2007). The loss function of the mean square error and the ADAM algorithm (Kingma & Ba, 2014) are adopted to optimize ML learnable parameters.

In the final stage (Figure 1e-3), the streamflow predicted by the surrogate model and the estimated ensemble errors ($\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$) diagnosed by ML are combined to provide improved streamflow estimation with UQ.

2.3. Data

In this study, we use White Oak Bayou—a highly urbanized watershed with an area of 46.9 km^2 in Harris County, TX (Figure 2a), as the case study domain. Predicting floods in this densely populated watershed is a difficult task due to the considerable impact of urbanization. A 3-m resolution Digital Elevation Model (Figure 2b) was used to develop a triangulated mesh network to capture terrain, channel, and urban features with a total of 136,423 mesh nodes and 271,215 triangular cells. The land use data from the National Land Cover Database 2016 shows predominantly developed areas: open space (13%), low intensity development (32%), medium intensity (35%), and high intensity (21%). Approximately 20,000 buildings were included in the mesh (~30% of the area). The processing of building footprints and other relevant pre-processing details are in Ivanov et al. (2021). Watershed outlet streamflow data were obtained for the USGS station 08074540. Precipitation data were obtained from Automated Surface Observing Systems (ASOS) at two stations (Houston Dunn and Houston Intercontinental).

2.4. Experimental Design

For this highly urbanized watershed, we assume the entire area is impervious. Infiltration losses are thus set to zero and all rainfall becomes surface runoff. Although tRIBS-OFM can simulate the runoff storage and flows

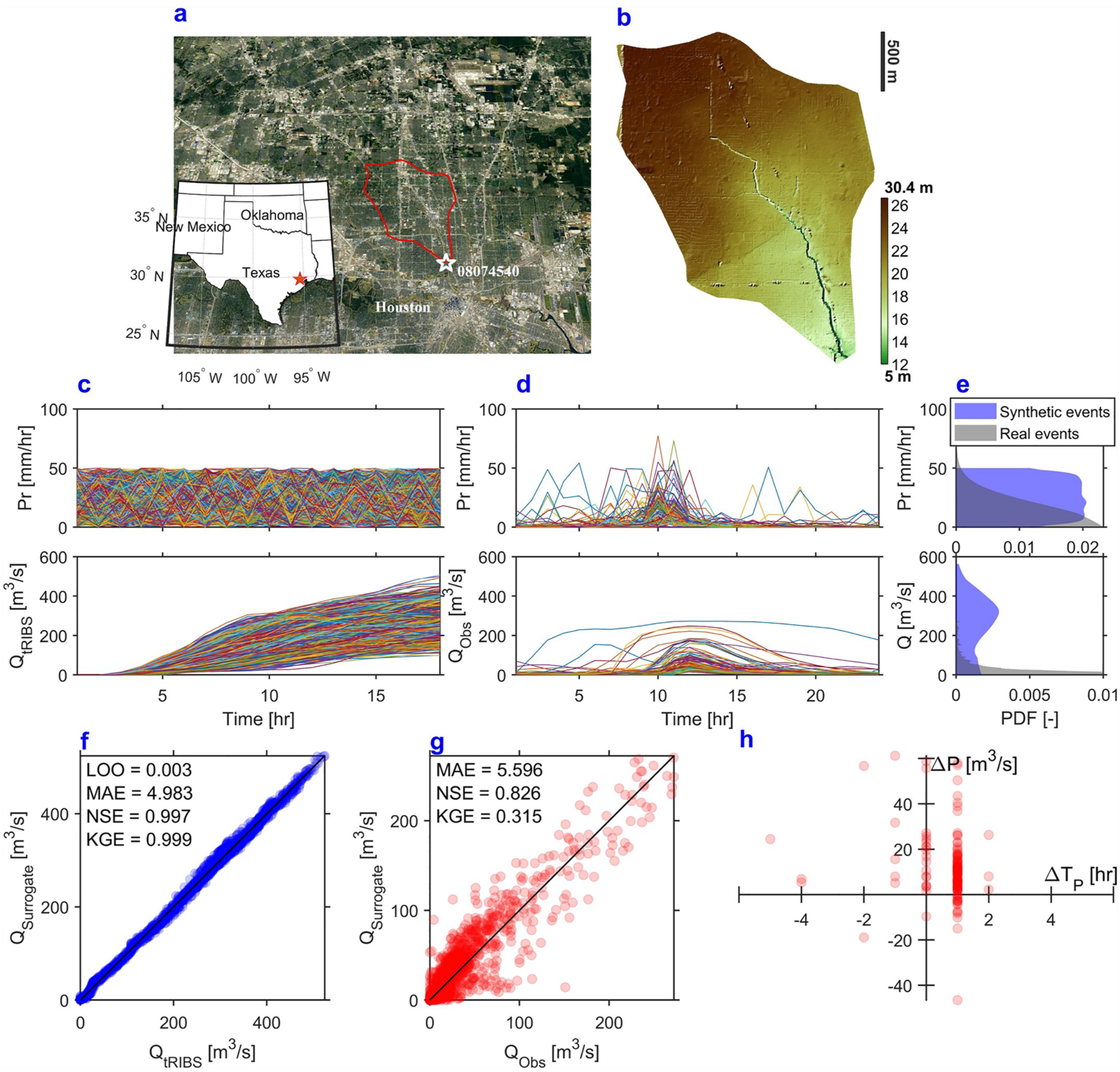

Figure 2. (a) The White Oak Bayou watershed in Harris County, Houston, TX, with a USGS stream gauge (ID = 08074540) marked by a white star. Subplots (b) shows the Digital Elevation Model. (c) 1100 realizations of 18-hourly rainfall series of uncorrelated pulses (top subplot) and the TIN-Based Real Time Integrated Basin Simulator—Overland Flow Model (tRIBS-OFM) simulated streamflows (bottom subplot) corresponding to these series. (d) The observed 24-hourly rainfall and streamflow for 138 events between 15 October 2007 and 28 November 2020. The probability density distributions of rainfall and streamflow rates for both synthetic (shaded in blue) and historic (shaded in gray) data sets are shown in subplot (e). (f) Surrogate model validation relative to tRIBS-OFM simulations for 100 synthetic events. (g) A comparison between the observed and surrogate-simulated streamflows for 138 24-hr rainfall events. (h) The peak difference ($\Delta P = Q_{\mathrm{Sim}}^{\mathrm{Peak}} - Q_{\mathrm{Obs}}^{\mathrm{Peak}}$), and time-to-peak ($\Delta T_P = t_{\mathrm{Sim}}^{\mathrm{peak}} - t_{\mathrm{Obs}}^{\mathrm{peak}}$) errors. Each circle denotes a combination of ΔP and ΔT_p for a specific event.

within stormwater infrastructure and soils, we chose to ignore these processes due to the lack of knowledge of soil water content before each simulated event and the lack of reliable data sets on stormwater infrastructure. We acknowledge that these assumptions can have a considerable impact on the simulated streamflow. Nevertheless, they also present an opportunity to ascertain whether the proposed framework can effectively capture relevant errors in $\tilde{\varepsilon}$ and $\boldsymbol{\varepsilon}'$ given these obvious inaccuracies in the description of watershed hydrology.

The Manning's roughness coefficient is the most influential parameter for overland flow (Ozdemir et al., 2013). We assume a spatially uniform value 0.015, which is the middle of the interval for concrete surfaces (0.012–0.018) (Arcement & Schneider, 1989).

This study considers rainfall as the only weather input. Other variables, such as temperature and humidity, are disregarded due to their minimal impacts on runoff generation during an event-scale period. For surrogate model training (Figure 1e-1), we use 1100 18-hr rainfall events, each of which is a series of spatially uniform pulses randomly generated from the Uniform distribution between 0 and 50 (mm/hr). This data set was designed to include extreme rates, while lacking any temporal "structure." tRIBS-OFM was thus run 1100 times to simulate streamflow series at the outlet (Figure 2c). Results of 1000 runs were deemed to be sufficient to train a surrogate model (Dwelle et al., 2019; Ivanov et al., 2021; Tran et al., 2020; Tran & Kim, 2022). The remaining 100 events were used for surrogate validation.

Once surrogate models for the QoI have been constructed and validated, they can then be applied as substitutes for tRIBS-OFM in order to simulate QoIs for any rainfall event (Figure 1e-2). In total, 138 24-hr historic rainfall events were chosen from 15 October 2007 to 28 November 2020 (Figure 2d). The maximum hourly rainfall depths range from 3.6 to 77.2 mm, and the corresponding peak streamflows range from 14.46 to 272.4 m^3/s. From Figure 2e, it is evident that the real rainfall and streamflow are mostly enclosed within the range of the training synthetic data. The surrogate model error $\boldsymbol{\varepsilon}^o$ for each of the 138 events was estimated. These error estimates and the rainfall forcing were then input to an ML model. Specifically, the first 100 events were used to train and validate the ML model (i.e., the training set), and the remaining 38 events were used to test the model (i.e., the testing set). In the testing phase, error ensemble was estimated for 38 events using the trained ML model coupled with HR and MCD. An ensemble size of 1000 was selected as a reasonable set to reflect the distribution of the errors. The distributions of $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ for each hour were obtained, as discussed in Text S5 in Supporting Information S1.

Metrics evaluating the model performance include the Mean Absolute Error (MAE), Nash-Sutcliffe model efficiency coefficient (NSE) (Nash & Sutcliffe, 1970), Kling-Gupta efficiency (KGE) (Gupta et al., 2009), peak difference (ΔP), and Time-to-Peak error (ΔT_P), which were calculated for each event. In addition, a leave-one-out error (LOO) is used in the process of constructing surrogate models and evaluating the similarity between outputs from the surrogate and original models (Blatman & Sudret, 2011).

3. Results

The 100 validation synthetic events displayed in Figure 2f exhibit excellent performance of the surrogate model in generating streamflows compared to those from tRIBS-OFM, with LOO of 0.03, MAE of 4.983 m^3/s (an error of ~2% as compared to the average streamflow in the validation data set of 241.3 m^3/s), and NSE and KGE achieve the near-perfect scores of 0.99. This result is consistent with prior studies demonstrating that PCE-based surrogates can mimic complex models with high accuracy (Dwelle et al., 2019; Ricciuto et al., 2018), even when they are trained with noisy inputs, instead of real-world rainfall (Ivanov et al., 2021; Tran & Kim, 2022). Surrogates also demonstrate their computational superiority, as they can simulate 100 events within a second. This is in stark contrast to tRIBS-OFM simulations that require ~5200 hr for the same event set (using a single processor).

The surrogate model shows a satisfactory performance in simulating streamflows for 138 actual events with the average NSE of 0.826 (Figures 2g and 2h), reflecting the skill of the high-fidelity model tRIBS-OFM. It is possible that such a high skill of the uncalibrated model might be due to the dominance of overland flow process that is well described by the full form of the Saint-Venant equation, given reasonable parameter values. However, the error is still quite considerable: MAE is 5.596 m^3/s, which is an error of 33.55% relative to the average flow of 16.67 m^3/s (across all 138 events). The KGE value is only 0.315, indicating that the model performs only moderately well. Figure 2h assesses the surrogate model performance in simulating peak flows and their timing. As seen, the surrogate model estimates higher peaks (ΔP) and delayed peak times (ΔT_P), as compared to the observations. The lack of stormwater sewer network representation in the model might be a contributing factor to the estimation of the delayed peak flows. A considerable amount of rainwater remains motionless inundating topographic lows and also, sheet overland flow may take more time to reach the watershed outlet than concentrated flows in a sewer network (Zhang et al., 2023). One my note that the surrogate model can capture these characteristics of tRIBS-OFM simulation. Hence, the surrogate model predicts higher and delayed peaks as compared to observations. A priori, the predictive accuracy can be improved through calibration of suitable parameters of the model, that is, signifying the effort of reducing their epistemic uncertainty. However, tuning parameters (e.g., the Manning's coefficient) with the aim to compensate for the impact of absent information (i.e., the knowledge of sewer system details) can be theoretically problematic: PBM's simulation content can become less physically meaningful. For example, tuning may provide unrealistic estimation for other possible variables of interest (e.g., water velocity or inundation depth) (Grayson et al., 1992; Guinot & Gourbesville, 2003).

The above dilemma can be addressed with the proposed probabilistic ML error model (Sections 2.1–2.2). Its trained version (using the error $\boldsymbol{\varepsilon}^o$ estimates from 100 events, Figure S1 in Supporting Information S1) is used to

10.1029/2023GL104464

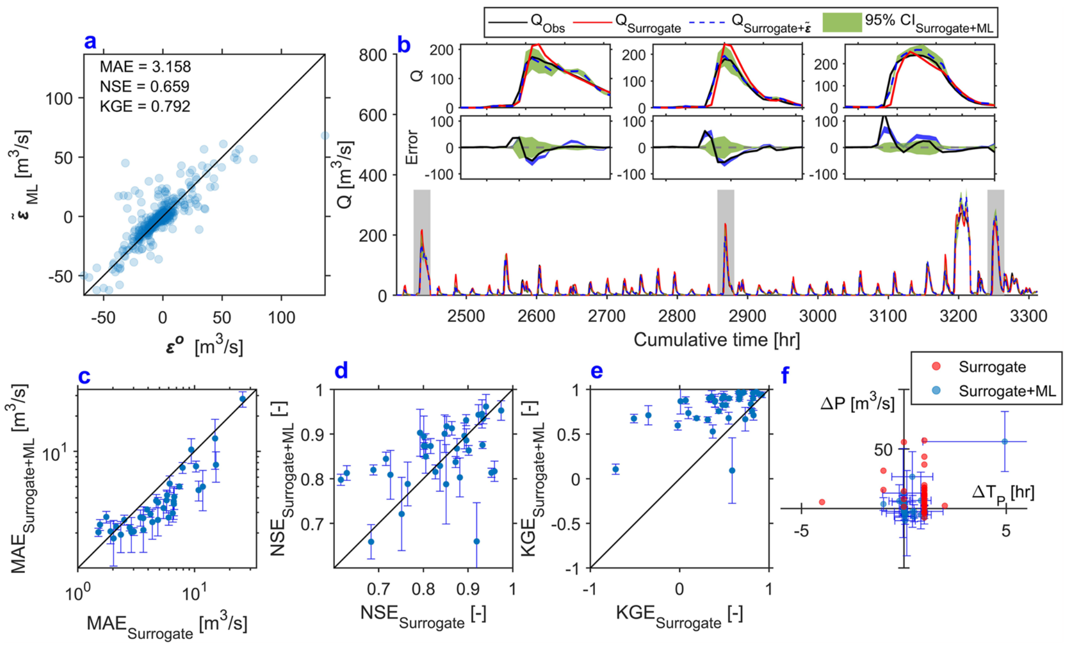

Figure 3. (a) A comparison between the surrogate errors ($\boldsymbol{\varepsilon}^{\boldsymbol{o}}$) and the means of 1000 ensemble errors $\tilde{\varepsilon}$ predicted by machine learning (ML) model for 38 rainfall events in the testing set. Results for each of 24 hr in each event are shown. Subplot (b) shows hydrographs of the 38 events. Observed streamflow (black line) is compared with the surrogate model simulation (red line), and that augmented with the mean of $\tilde{\varepsilon}$ (blue dashed line—"surrogate + $\tilde{\varepsilon}$"). The green region is the 95% distribution range for $\boldsymbol{\varepsilon}'$ obtained from 1000 ML ensemble predictions. The region is plotted by adding the hourly 2.5% and 97.5 percentiles of $\boldsymbol{\varepsilon}'$ to the hourly "surrogate + $\tilde{\varepsilon}$" hydrograph. The insets correspond to exemplary events on 17 April 2016; 17 January 2017; and 19 September 2019 (a full set is in Figures S2 and S3 in Supporting Information S1). The three upper subplots illustrate the same information as in the primary subplot (b), while the three lower subplots display the predicted errors: 95% confidence intervals (CI) for $\tilde{\varepsilon}$ (region in blue) and $\boldsymbol{\varepsilon}'$ (green) from 1000-member ML ensemble. Subplots (c), (d), and (e) compare the performance of the surrogate and combined "surrogate + ML" (blue dots with whiskers: dot = surrogate prediction + the mean of $\tilde{\varepsilon}$; whiskers = the 2.5% and 97.5% percentiles obtained from 1000 values of the "surrogate + ML" ensemble) based on mean absolute error (MAE), Nash-Sutcliffe model efficiency (NSE), and Kling-Gupta efficiency (KGE) over 38 events. Subplot (f) reproduces Figure 2i (red dots for 38 testing events) and results obtained with "surrogate + ML" (blue dots with whiskers) quantifying the errors of peak flow (ΔP) and time to peak (ΔT_P) predictions.

make 1000 error ensemble predictions for 38 testing events. Results depicted in Figure 3a are found to be fairly accurate, with an average MAE of 3.158 m^3/s for 1000 ensembles, and average NSE and KGE values of 0.659 and 0.754, respectively. The 1000-member ensemble generated with the ML model produces a 95% confidence interval (CI) for the reducible error $\tilde{\varepsilon}$ for each hour of simulation (the area shaded in blue, see insets in Figure 3b, Figures S2 and S3 in Supporting Information S1). It is relatively narrow and largely encompasses the surrogate error $\boldsymbol{\varepsilon}^{\boldsymbol{o}}$. The greatest uncertainty is typically observed during peak flow events, when model error is high. The skill exhibiting such a high accuracy is noteworthy, particularly given the fact that the variability of $\boldsymbol{\varepsilon}^{\boldsymbol{o}}$ can be quite significant and the correlation between $\boldsymbol{\varepsilon}^{\boldsymbol{o}}$ and rainfall is quite weak, with an R^2 of ~0.018. Nonetheless, the irreducible uncertainty of $\boldsymbol{\varepsilon}'$ (the area shaded in green, Figure 3b), can be quite large and larger than the uncertainty of $\tilde{\varepsilon}$ over 38 events (Figure S5 in Supporting Information S1).

The results of the ensemble streamflows combining surrogate prediction and errors quantified with ML (termed "surrogate + ML" prediction hereafter) are in Figures 3c–3f (and in Figures S2 and S3 in Supporting Information S1). Overall, the results demonstrate that the accuracy of "surrogate + ML" prediction is significantly better than that of the surrogate model alone. Inspection of Figure 3b demonstrates that the predictions become more consistent with the observed streamflow. The 95% CI encloses the observed series, albeit with higher uncertainties at the peak flows. The assessment with MAE, NSE, and KGE metrics shows that they all agree, with the average over 1000 ensemble members for 38 events approaching their theoretically ideal values (i.e., 0.0, 1.0, and 1.0, respectively) (Figures 3c–3e). Specifically, the "surrogate + ML" prediction yields an average MAE of 4.53 m^3/s, NSE of 0.85, and KGE of 0.77, which are better than the corresponding values of 6.08 m^3/s, 0.83, and 0.40, respectively, obtained from the surrogate model results alone. The ΔP and ΔT_P analysis and estimation shown in Figure 3f further demonstrate an improvement. For instance, the results in Figure 3b demonstrate how these flood peak characteristics in the "surrogate + ML" prediction become closer to the observations. The computational efficiency of the integrated PCE + ML models is high yielding 1000 ensemble results in just over 10 s. This is undeniably superior to the standard UQ approach that would adopt traditional Monte Carlo simulations with tRIBS-OFM, necessitating 52,000 hr for the same number of simulations.

4. Discussion

The results of this research serve to validate the efficacy of a novel modeling framework in Figure 1e that combines the strengths of PBM, surrogate, and ML models. A PBM offers a sound, first-principles representation of the hydrological system, a surrogate model acts as a bridge to achieve that with computational efficiency, while ML adds accuracy by providing insights into reducible and irreducible sources of prediction errors. As demonstrated, the framework is successful and the results have several major implications.

First, a state-of-the-science high-fidelity flood model that relies on *uncalibrated* lookup-table parameters can provide relatively accurate streamflow predictions when overland flow dominates the modeled response. This result emphasizes PBM's efficacy to produce not exact but skillful predictions in urban environments. Though perhaps logical, to the author's knowledge, this has not been demonstrated explicitly in prior research. As humans fractionate landscapes and change their drainage graphs (Ott & Uhlenbrook, 2004; Rogger et al., 2017), this utility of PBMs (or PBM-to-surrogate models) to flexibly incorporate the changing urbanscape characteristics will continue to be warranted, even if PBM structural alterations (and the corresponding surrogate re-training) are necessitated. In contrast, models driven exclusively by data (such as pure ML models) are handicapped in human-impacted watersheds where hydrology becomes non-stationary and past observations are rendered obsolete.

Second, previous studies have already demonstrated the power and efficiency of surrogate models (Ivanov et al., 2021; Tran et al., 2020) in their capability to capture the PBM behavior with a low level of added uncertainties. This framework posits them as an essential machinery in the forecaster's toolkit, even though surrogates require a large initial computational overhead for their training. Trained with samples supervised by PBM's high-fidelity solutions, surrogate models can mirror outcomes of a broad spectrum of physical relations. This contrasts recently novel applications of an approach of physics-informed ML (Bhasme et al., 2022; Frame et al., 2021; Lu et al., 2021; Zhong et al., 2023), in which simple constraints to a physics-ignorant solution to maintain plausible ranges for modeled variables cannot attain the same level of prediction comprehensiveness as offered by the PBM-to-surrogate modeled dynamics. Under certain conditions however surrogate model performance can become inferior. One apparent concern is the case of high dimensionality of uncertain variables such as in the spatial variability of forcings (e.g., rainfall) or antecedent conditions. In these instances, surrogate models face the challenge known as the "curse of dimensionality" (Sargsyan et al., 2014; Sudret, 2007). An analysis of QoI sensitivity to uncertain inputs (Dwelle et al., 2019) with a subsequent reduction of their set that can reduce the training overhead is one plausible direction to address this challenge. However, this topic remains an active research area.

Third, the demonstrated capacity of an ML model that can learn PBM (or its surrogate) error structure and reduce biases is a leap forward in improving predictive accuracy. Earlier intercomparison projects largely concluded that lumped/conceptual models remain superior to PBMs in terms of streamflow prediction (Smith et al., 2004, 2012a). This study shows that it is possible to relax that limitation of the PBM utility. The demonstrated ML-based error partition into $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ is agnostic to specific sources of PBM uncertainty: they can be due to both epistemic (e.g., model structure or parameter biases) or aleatoric (e.g., random biases in rainfall) uncertainties. There is no obvious way to connect them, other than perhaps for hypothetical, idealized cases of perfect geometry and homogeneity assumptions. This limitation is not however viewed as essential. The ability to "sense" and partition out the reducible error $\tilde{\boldsymbol{\varepsilon}}$—despite any obvious relationship between model inputs and observed errors (Figure S4 in Supporting Information S1) and regardless of recognizing the error source—in author's opinion, is a breakthrough in improving predictive accuracy of streamflow simulations.

Further, the uncertainty communicated by the ML model in the form of $\tilde{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varepsilon}'$ distributions offers the potential to eliminate the need for traditional UQ techniques quantifying (most commonly) PBM's epistemic uncertainty. Specifically, while a PBM modeler may always choose to carry out parameter inference using surrogate models

(Dwelle et al., 2019; Xu et al., 2022) to ensure physical realism in model QoI simulations, predictions with best-fit parameters will always remain inexact because of the multi-variate nature of error sources. The typical Monte Carlo (or "pushed-forward") simulations of posterior distributions of perceived uncertainty sources deals only with *pre-contemplated* uncertainties: they are in the PBM structure or its inputs (Sargsyan et al., 2019). The ML-based approach proposed here explores uncertainty structures *that may be entirely hidden*. These can include uncertainties not perceived as important and thus excluded (here, e.g., infiltration losses in an urbanized domain) or the "unknown unknowns" (Beven et al., 2011)—uncertainties that are unexpected and not predicted when they occur (e.g., failure of the uniform rainfall intensity assumption). Our results demonstrate that $\tilde{\boldsymbol{\varepsilon}}$ distribution correctly adjusts streamflows simulated with surrogates and the $\boldsymbol{\varepsilon}'$ confidence intervals largely envelop the observed streamflows (Figures S2 and S3 in Supporting Information S1). This is a crucial outcome demonstrating that the proposed approach can serve as an efficient UQ technique. This is compared to traditional approaches relying on the computationally intensive "pushed-forward" simulations, which nonetheless can be used a posteriori to explore the specific sources of the reducible error $\tilde{\boldsymbol{\varepsilon}}$ distribution.

The framework developed here has limitations. For a few events, the prediction accuracy was not improved (Figures 3c–3e). This demonstrates that not all reducible errors can be efficiently calculated. This is not an unexpected outcome, as only rainfall data were used to identify $\boldsymbol{\varepsilon}$—this may be insufficient, particularly in complex urban areas. Experimentation with more diverse data sets can provide a more comprehensive understanding of the importance of various factors influencing the flow regime. Further, it should be noted that the success of ML model is contingent on data availability. In cases where there is a limited amount of training data, ML may not be advantageous, or even fail to improve simulations due to the issue of underfitting. This highlights the importance of obtaining comprehensive and informative data, so that ML performance can be enhanced thus ensuring viability of the proposed integrated framework.

Overall, the proposed framework provides a comprehensive, adaptive, and highly efficient solution. Issues of computational efficiency and the lack of predictive accuracy are pervasive in many Earth sciences that rely on modeling. We anticipate that this solution can be applicable in other domains of geophysical research.

Data Availability Statement

The land use information was obtained from the National Land Cover Database 2016 (https://www.mrlc.gov/national-land-cover-database-nlcd-2016). Building footprints were downloaded from an open source (https://koordinates.com/layer/12890-houston-texas-building-footprints/). Streamflow at the USGS gauge 08074540 was obtained from https://waterdata.usgs.gov/monitoring-location/08074540/. Precipitation data was obtained from Automated Surface Observing Systems (ASOS) at two stations, Houston Dunn and Houston Intercontinental (https://www.ncei.noaa.gov/products/land-based-station/automated-surface-weather-observing-systems). The synthetic data sets for surrogate training are available through the Zenodo digital repository accessible at https://doi.org/10.5281/zenodo.8092949. The source code for the surrogate model can be obtained from https://www.uqlab.com/. The machine learning model was trained using the Keras-Tensorflow library (https://github.com/lerimaggio/deep-learning-keras-tensorflow). Code for the Bayesian optimization, wavelet transform, and mutual information were obtained from open sources, including https://github.com/thuijskens/bayesian-optimization, https://github.com/PyWavelets/pywt, https://github.com/scikit-learn/scikit-learn, respectively.

Acknowledgments
Jongho Kim was supported by a 2022-MOIS63-002 of Cooperative Research Method and Safety Management Technology in National Disaster (South Korea). Valeriy Y. Ivanov acknowledges the support of the U.S. National Science Foundation CMMI program award # 2053429. Donghui Xu was supported by the Scientific Discovery through Advanced Computing 5 (Capturing the Dynamics of Compound Flooding in E3SM), funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

References

Alizadeh, B., Ghaderi Bafti, A., Kamangir, H., Zhang, Y., Wright, D. B., & Franz, K. J. (2021). A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *Journal of Hydrology*, *601*, 126526. https://doi.org/10.1016/j.jhydrol.2021.126526

Arcement, G. J., & Schneider, V. R. (1989). *Guide for selecting Manning's roughness coefficients for natural channels and flood plains* (Vol. 239). US Government Printing Office.

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. (2016). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, *61*(9), 1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven, K., Smith, P. J., & Wood, A. (2011). On the colour and spin of epistemic error (and what we might do about it). *Hydrology and Earth System Sciences*, *15*(10), 3123–3133. https://doi.org/10.5194/hess-15-3123-2011

Bhasme, P., Vagadiya, J., & Bhatia, U. (2022). Enhancing predictive skills in physically-consistent way: Physics informed machine learning for hydrological processes. *Journal of Hydrology*, *615*, 128618. https://doi.org/10.1016/j.jhydrol.2022.128618

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

Camporese, M., Paniconi, C., Putti, M., & Orlandini, S. (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, *46*(2), W02512. https://doi.org/10.1029/2008WR007536

Carpenter, T. M., & Georgakakos, K. P. (2006). Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology*, *329*(1–2), 174–185. https://doi.org/10.1016/j.jhydrol.2006.02.013

Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*, *403*, 348–359. https://doi.org/10.1016/j.neucom.2020.04.110

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*, *125*, 13–31. https://doi.org/10.1016/j.advwatres.2019.01.002

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*(5), 3505–3524. https://doi.org/10.1002/2014wr016102

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Frame, J., Kratzert, F., Klotz, D., Gauch, M., Shelev, G., Gilon, O., et al. (2021). Deep learning rainfall-runoff predictions of extreme events. *Hydrology and Earth System Sciences Discussions*, *2021*, 1–20. https://doi.org/10.5194/hess-2021-423

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Paper presented at international conference on machine learning, PMLR*.

Ghaith, M., & Li, Z. (2020). Propagation of parameter uncertainty in SWAT: A probabilistic forecasting method based on polynomial chaos expansion and machine learning. *Journal of Hydrology*, *586*, 124854. https://doi.org/10.1016/j.jhydrol.2020.124854

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

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., & Hero, A. O., III. (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, *49*(4), 2253–2273. https://doi.org/10.1002/wrcr.20161

Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 2. Is the concept realistic? *Water Resources Research*, *28*(10), 2659–2666. https://doi.org/10.1029/92WR01259

Guinot, V., & Gourbesville, P. (2003). Calibration of physically based models: Back to basics. *Journal of Hydroinformatics*, *5*(4), 233–244. https://doi.org/10.2166/hydro.2003.0020

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Helton, J. C., Johnson, J. D., Oberkampf, W. L., & Sallaberry, C. J. (2010). Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, *39*(6), 605–646. https://doi.org/10.1080/03081079.2010.486664

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

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*(9), W09521. https://doi.org/10.1029/2009wr008611

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., & Entekhabi, D. (2004a). Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resources Research*, *40*(11), W11102. https://doi.org/10.1029/2004wr003218

Ivanov, V. Y., Xu, D., Dwelle, M. C., Sargsyan, K., Wright, D. B., Katopodes, N., et al. (2021). Breaking down the computational barriers to real-time urban flood forecasting. *Geophysical Research Letters*, *48*(20), e2021GL093585. https://doi.org/10.1029/2021gl093585

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer. In *Paper presented at thirsty-first conference on neural information processing systems*.

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*(2), 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*(1), 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*(24), 12446–12456. https://doi.org/10.1002/2016gl071480

Kim, J., Ivanov, V. Y., & Katopodes, N. D. (2012). Hydraulic resistance to overland flow on surfaces with partially submerged vegetation. *Water Resources Research*, *48*(10), W10540. https://doi.org/10.1029/2012wr012047

Kim, J., Ivanov, V. Y., & Katopodes, N. D. (2013). Modeling erosion and sedimentation coupled with hydrological and overland flow processes at the watershed scale. *Water Resources Research*, *49*(9), 5134–5154. https://doi.org/10.1002/wrcr.20373

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kumanlioglu, A. A., & Fistikoglu, O. (2019). Performance enhancement of a conceptual hydrological model by integrating artificial intelligence. *Journal of Hydrologic Engineering*, *24*(11), 04019047. https://doi.org/10.1061/(asce)he.1943-5584.0001850

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lu, D., Konapala, G., Painter, S. L., Kao, S.-C., & Gangrade, S. (2021). Streamflow simulation in data-scarce basins using Bayesian and physics-informed machine learning models. *Journal of Hydrometeorology*, *22*(6), 1421–1438. https://doi.org/10.1175/jhm-d-20-0082.1

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J., Ferguson, I. M., Ivanov, V., et al. (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *50*(2), 1531–1549. https://doi.org/10.1002/2013wr013725

May, R. J., Maier, H. R., Dandy, G. C., & Fernando, T. M. K. G. (2008). Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, *23*(10), 1312–1326. https://doi.org/10.1016/j.envsoft.2008.03.007

Micaletto, G., Barletta, I., Mocavero, S., Federico, I., Epicoco, I., Verri, G., et al. (2022). Parallel implementation of the SHYFEM (System of HydrodYnamic Finite Element Modules) model. *Geoscientific Model Development*, *15*(15), 6025–6046. https://doi.org/10.5194/gmd-15-6025-2022

Montanari, A., Shoemaker, C. A., & Van de Giesen, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges. *Water Resources Research*, *45*(12), W00B00. https://doi.org/10.1029/2009wr008471

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). Springer. https://doi.org/10.1007/978-3-540-77843-1_1

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Ott, B., & Uhlenbrook, S. (2004). Quantifying the impact of land-use changes at the event and seasonal time scale using a process-oriented catchment model. *Hydrology and Earth System Sciences*, *8*(1), 62–78. https://doi.org/10.5194/hess-8-62-2004

Ozdemir, H., Sampson, C. C., de Almeida, G. A. M., & Bates, P. D. (2013). Evaluating scale and roughness effects in urban flood modelling using terrestrial LIDAR data. *Hydrology and Earth System Sciences*, *17*(10), 4015–4030. https://doi.org/10.5194/hess-17-4015-2013

Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*(7), W07401. https://doi.org/10.1029/2011wr011527

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019WR024922

Ricciuto, D., Sargsyan, K., & Thornton, P. (2018). The impact of parametric uncertainties on biogeochemistry in the E3SM land model. *Journal of Advances in Modeling Earth Systems*, *10*(2), 297–319. https://doi.org/10.1002/2017ms000962

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., et al. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research*, *53*(7), 5209–5219. https://doi.org/10.1002/2017WR020723

Roy, A., Kasiviswanathan, K. S., Patidar, S., Adeloye, A. J., Soundharajan, B.-S., & Ojha, C. S. P. (2023). A novel physics-aware machine learning-based dynamic error correction model for improving streamflow forecast accuracy. *Water Resources Research*, *59*(2), e2022WR033318. https://doi.org/10.1029/2022WR033318

Ruhi, A., Hwang, J., Devineni, N., Mukhopadhyay, S., Kumar, H., Comte, L., et al. (2022). How does flow alteration propagate across a large, highly regulated basin? Dam attributes, network context, and implications for biodiversity. *Earth's Future*, *10*(6), e2021EF002490. https://doi.org/10.1029/2021EF002490

Sargsyan, K., Huan, X., & Najm, H. N. (2019). Embedded model error representation for Bayesian model calibration. *International Journal for Uncertainty Quantification*, *9*(4), 365–394. https://doi.org/10.1615/int.j.uncertaintyquantification.2019027384

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

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018wr022643

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., et al. (2012a). The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments. *Journal of Hydrology*, *418–419*, 3–16. https://doi.org/10.1016/j.jhydrol.2011.08.055

Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S. M., Cui, Z., et al. (2012b). Results of the DMIP 2 Oklahoma experiments. *Journal of Hydrology*, *418–419*, 17–48. https://doi.org/10.1016/j.jhydrol.2011.08.056

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., et al. (2004). The distributed model intercomparison project (DMIP): Motivation and experiment design. *Journal of Hydrology*, *298*(1–4), 4–26. https://doi.org/10.1016/j.jhydrol.2004.03.040

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems, 25*.

Sudret, B. (2007). *Uncertainty propagation and sensitivity analysis in mechanical models–Contributions to structural reliability and stochastic spectral methods*. Habilitationa diriger des recherches (p. 18). Université Blaise Pascal.

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

Tran, V. N., Dwelle, M. C., Sargsyan, K., Ivanov, V. Y., & Kim, J. (2020). A novel modeling framework for computationally efficient and accurate real-time ensemble flood forecasting with uncertainty quantification. *Water Resources Research*, *56*(3), e2019WR025727. https://doi.org/10.1029/2019WR025727

Tran, V. N., & Kim, J. (2021). A robust surrogate data assimilation approach to real-time forecasting using polynomial chaos expansion. *Journal of Hydrology*, *598*, 126367. https://doi.org/10.1016/j.jhydrol.2021.126367

Tran, V. N., & Kim, J. (2022). Robust and efficient uncertainty quantification for extreme events that deviate significantly from the training dataset using polynomial chaos-kriging. *Journal of Hydrology*, *609*, 127716. https://doi.org/10.1016/j.jhydrol.2022.127716

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems, 30*.

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing & Applications*, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

Vivoni, E. R., Mascaro, G., Mniszewski, S., Fasel, P., Springer, E. P., Ivanov, V. Y., & Bras, R. L. (2011). Real-world hydrologic assessment of a fully-distributed hydrological model in a parallel computing environment. *Journal of Hydrology*, *409*(1–2), 483–496. https://doi.org/10.1016/j.jhydrol.2011.08.053

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8). https://doi.org/10.1029/2002wr001642

Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., et al. (2017). Human–water interface in hydrological modelling: Current status and future directions. *Hydrology and Earth System Sciences*, *21*(8), 4169–4193. https://doi.org/10.5194/hess-21-4169-2017

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

Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, *12*(1), 2501. https://doi.org/10.1038/s41467-021-22702-2

Xu, D., Bisht, G., Sargsyan, K., Liao, C., & Leung, L. R. (2022). Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1. *Geoscientific Model Development*, *15*(12), 5021–5043. https://doi.org/10.5194/gmd-15-5021-2022

Yao, Y., Rosasco, L., & Caponnetto, A. (2007). On early stopping in gradient descent learning. *Constructive Approximation*, *26*(2), 289–315. https://doi.org/10.1007/s00365-006-0663-2

Young, C.-C., Liu, W.-C., & Wu, M.-C. (2017). A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Applied Soft Computing*, *53*, 205–216. https://doi.org/10.1016/j.asoc.2016.12.052

Zhang, K., Sebo, S., McDonald, W., Bhaskar, A., Shuster, W., Stewart, R. D., & Parolari, A. J. (2023). The role of inflow and infiltration (I/I) in urban water balances and streamflow regimes: A hydrograph analysis along the sewershed-watershed continuum. *Water Resources Research*, *n/a*(n/a), e2022WR032529. https://doi.org/10.1029/2022WR032529

Zhong, L., Lei, H., & Gao, B. (2023). Developing a physics-informed deep learning model to simulate runoff response to climate change in alpine catchments. *Water Resources Research*, *59*(6), e2022WR034118. https://doi.org/10.1029/2022WR034118

References From the Supporting Information

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, *390*(1–2), 85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*(1), 3–31. https://doi.org/10.1016/s0167-7012(00)00201-3

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

Budu, K. (2014). Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. *Journal of Hydrologic Engineering*, *19*(7), 1385–1400. https://doi.org/10.1061/(asce)he.1943-5584.0000892

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

Feng, D., Lawson, K., & Shen, C. (2021). Mitigating prediction error of deep learning streamflow models in large data-sparse regions with ensemble modeling and soft data. *Geophysical Research Letters*, *48*(14), e2021GL092999. https://doi.org/10.1029/2021gl092999

Ivanov, V. Y., Bras, R. L., & Vivoni, E. R. (2008). Vegetation-hydrology dynamics in complex terrain of semiarid areas: 2. Energy-Water controls of vegetation spatiotemporal dynamics and topographic niches of favorability. *Water Resources Research*, *44*(3), W03430. https://doi.org/10.1029/2006WR005595

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., & Entekhabi, D. (2004b). Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: A fully-distributed physically-based approach. *Journal of Hydrology*, *298*(1), 80–111. https://doi.org/10.1016/j.jhydrol.2004.03.041

Kim, J., Warnock, A., Ivanov, V. Y., & Katopodes, N. D. (2012). Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow. *Advances in Water Resources*, *37*, 104–126. https://doi.org/10.1016/j.advwatres.2011.11.009

Li, Y., Zhu, Z., Kong, D., Han, H., & Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowledge-Based Systems*, *181*, 104785. https://doi.org/10.1016/j.knosys.2019.05.028

Maheswaran, R., & Khosa, R. (2012). Comparative study of different wavelets for hydrologic forecasting. *Computers & Geosciences*, *46*, 284–295. https://doi.org/10.1016/j.cageo.2011.12.015

May, R., Dandy, G., & Maier, H. (2011). Review of input variable selection methods for artificial neural networks. *Artificial Neural Networks-Methodological Advances and Biomedical Applications*, *10*(1), 19–45.

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*, *54*(10), 7733–7751. https://doi.org/10.1029/2018wr022676

Nayak, P. C., Venkatesh, B., Krishna, B., & Jain, S. K. (2013). Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. *Journal of Hydrology*, *493*, 57–67. https://doi.org/10.1016/j.jhydrol.2013.04.016

Ni, L., Wang, D., Singh, V. P., Wu, J., Wang, Y., Tao, Y., & Zhang, J. (2020). Streamflow and rainfall forecasting by two long short-term memory-based models. *Journal of Hydrology*, *583*, 124296. https://doi.org/10.1016/j.jhydrol.2019.124296

Nourani, V., Komasi, M., & Mano, A. (2009). A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water Resources Management*, *23*(14), 2877–2894. https://doi.org/10.1007/s11269-009-9414-5

Quilty, J., & Adamowski, J. (2018). Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *Journal of Hydrology*, *563*, 336–353. https://doi.org/10.1016/j.jhydrol.2018.05.003

Quilty, J., Adamowski, J., & Boucher, M. A. (2019). A stochastic data-driven ensemble forecasting framework for water resources: A case study using ensemble members derived from a database of deterministic wavelet-based models. *Water Resources Research*, *55*(1), 175–202. https://doi.org/10.1029/2018wr023205

Seeger, M. (2004). Gaussian processes for machine learning. *International Journal of Neural Systems*, *14*(02), 69–106. https://doi.org/10.1142/s0129065704001899

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175. https://doi.org/10.1109/jproc.2015.2494218

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shensa, M. J. (1992). The discrete wavelet transform: Wedding the a trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, *40*(10), 2464–2482. https://doi.org/10.1109/78.157290

Shi, L., Yang, J., Zhang, D., & Li, H. (2009). Probabilistic collocation method for unconfined flow in heterogeneous media. *Journal of Hydrology*, *365*(1–2), 4–10. https://doi.org/10.1016/j.jhydrol.2008.11.012

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

Tao, L., He, X., Li, J., & Yang, D. (2021). A multiscale long short-term memory model with attention mechanism for improving monthly precipitation prediction. *Journal of Hydrology*, *602*, 126815. https://doi.org/10.1016/j.jhydrol.2021.126815

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

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

Venkata Ramana, R., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly rainfall prediction using wavelet neural network analysis. *Water Resources Management*, *27*(10), 3697–3711. https://doi.org/10.1007/s11269-013-0374-4

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, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. In *Paper presented at Proceedings of the 2016 conference on empirical methods in natural language processing*.

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