Journal of Hydrology 613 (2022) 128323

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Development of a physics-informed data-driven model for gaining insights into hydrological processes in irrigated watersheds

Kailong Li [a,b], Guohe Huang [a,*], Shuo Wang [c], Saman Razavi [b]

[a] Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan, Canada
[b] School of Environment and Sustainability, Department of Civil, Geological and Environmental Engineering, and Global Institute for Water Security, University of Saskatchewan, Canada
[c] Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief

Keywords:
Data-driven model
Hydrological inference
Importance ranking
Baseflow separation
Irrigated watershed

ABSTRACT

Data-driven hydrological modeling has seen rapid development in recent years owing to its flexibility to approximate the complex relationships between driving forces and hydrological fluxes. However, traditional data-driven models typically cannot simultaneously capture the processes that pose both chronic and acute impacts on streamflow, thus impeding further inference. Therefore, this study presents a baseflow-filtered hydrological inference model to gain insights into hydrological processes in irrigated watersheds. The proposed model starts with separating the streamflow process into two sub-processes using a process-based baseflow separation method. Each sub-process is simulated through a new interpretable data-driven model. The resulting hydrological inferences facilitate the identification of the dominant factors influencing flows in saturated and unsaturated zones. The proposed model is applied to three irrigated watersheds, and the evaluation metrics show that the proposed model outperforms two conventional data-driven models. Our findings reveal that predictors associated with air temperature and long-term (i.e., monthly) irrigation are mainly responsible for characterizing baseflow dynamics, while precipitation and short-term (i.e., semi-weekly or weekly) irrigation are primarily responsible for describing overland flow and interflow dynamics. The fidelity of the derived hydrological inference is further demonstrated through sensitivity analysis. The results show that the relative importance of predictors not only reflects their significance on model performance, but also influence the changes on streamflow.

1. Introduction

Farming has altered the hydrological processes and threatened the ecosystems in many regions worldwide (Clement et al., 2009; Dewandel et al., 2008; Dong et al., 2015; Gosain et al., 2005). In particular, flood irrigation is a prevailing irrigation method that has been widely used in many countries (Das Bhowmik et al., 2020; Mottaleb et al., 2019). On one hand, such an irrigation method intensifies the rainfall-runoff process by increasing surface runoff, thereby raising the risk of flooding (Gu et al., 2019). On the other hand, it increases groundwater levels and causes salinization (Kong et al., 2021; Waleeittikul et al., 2019). To help mitigate these negative impacts and to develop sound irrigation plans, it is necessary to examine the contribution of relevant driving forces (e.g., precipitation and irrigation) to surface and subsurface flows.

Hydrological processes in irrigated watersheds are more complex than those in pristine watersheds due to crop evapotranspiration (Vishwakarma et al., 2022) and irrigation schemes (Li et al., 2022a), which should consider both flood control and optimal allocation of often insufficient irrigation water available spatially and temporally (Li et al., 2019; Razavi et al., 2020). Such complexities lead to a significant challenge in quantifying the relative contributions of irrigation and precipitation to the rainfall-runoff processes. Extensive efforts have been made in the past decades to quantify associated driving forces, such as the use of isotope-based methods (Kong et al., 2021; Lee et al., 2021; Lv et al., 2018), correlation-based methods (Młyński et al., 2021; Qing et al., 2022; You and Wang, 2021), sensitivity-based methods (Kumar et al., 2022), and process-based simulation models (Ditthakit et al., 2021; Liu et al., 2020; Ramireddygari et al., 2000; Traylor and Zlotnik, 2016; Zeng and Cai, 2014). Although previous approaches advanced our understanding of the underlying hydrological mechanisms in irrigated watersheds, they required tremendous time to acquire and process the

* Corresponding author at: 3737 Wascana Pkwy, Regina, Saskatchewan, Canada.
E-mail addresses: Kailong.li@usask.ca (K. Li), huangg@uregina.ca (G. Huang).

https://doi.org/10.1016/j.jhydrol.2022.128323
Received 17 February 2022; Received in revised form 6 August 2022; Accepted 7 August 2022
Available online 20 August 2022

Nomenclature

BFSCE	Baseflow-filtered stepwise clustered ensemble
KGE	Kling-Gupta efficiency
MAE	Mean absolute error
MFA	Multilevel factorial analysis
MK	Mann-Kendall trend test
ML	Machine learning
NSE	Nash–Sutcliffe efficiency
OOB	Out-of-bag samples
O&I	Overland flow and interflow
R^2	Coefficient of determination
RC	Relative contribution
RF	Random Forest
RMSE	Root mean squared error
RTE	Regression tree ensemble
SCA	Stepwise cluster analysis
SI	Spring irrigation
VE	Volumetric efficiency
WFI	Wilks feature importance
WI	Winter irrigation

data (Lee et al., 2021; Liu et al., 2020). For instance, the limited accessibility of high-resolution spatially distributed irrigated water may hinder the application of process-based simulation models. Moreover, irrigated watersheds are often characterized by a flat terrain, which is particularly difficult to delineate an appropriate watershed domain for running distributed hydrological models (Rahman et al., 2010).

As an alternative to the above-mentioned approaches, data-driven models, such as neural networks (Yang et al., 2021; ASCE Task Committee, 2000; Baek et al., 2020; Chen et al., 2014; Hsu et al., 1995; Kratzert et al., 2019a), genetic programming (Babovic and Keijzer, 2002; Chadalawada et al., 2020; Meshgi et al., 2015) and regression tree ensemble (RTE) (Galelli and Castelletti, 2013; Li et al., 2021a; Schnier and Cai, 2014; Zhang et al., 2019), have received increasing attention in recent years, especially in the quantification of the contributions of driving forces. For instance, Kratzert et al. (2019b) investigated the relative contributions of climatic and non-climatic factors to runoff generation through long short-term memory networks. Konapala and Mishra (2020) examined 60 variables through a random forest model (Breiman, 2001) to understand their relative contributions to hydrological drought development. Schmidt et al. (2020) explored the relative contribution of multiple driving forces to flood events in Germany through three data-driven models. Although previous studies have effectively quantified the relationship between model inputs and outputs, it is still unclear whether the relative contributions of driving forces are physically consistent.

"Physically consistent" means that the obtained relative contributions not only reflect the importance of predictors influencing the goodness-of-fit of a model but also influence the changes on streamflow (Razavi, 2021). In recent years, the need for physically consistency of data-driven models has led to research efforts on hybridizing process-based and data-driven approaches in hydrological sciences and beyond – refer to Razavi et al. (2022) for a review and perspective of such hybridization efforts. There are various ways of incorporating hydrological knowledge into data-driven models (Chadalawada et al., 2020; Herath et al., 2021; Karpatne et al., 2018; Nearing et al., 2021; Singh et al., 2017). For instance, Daw et al. (2017) used physics-based loss functions to guide the learning of neural networks. Liang et al. (2019) and Lu et al. (2021) incorporated the outputs from a process-based model into data-driven models to improve predictive skills. However, since the above-mentioned approaches did not take into account the physical processes of hydrology, the interpretability is still limited.

In recent years, attempts have been made to incorporate data-driven models into the structure of process-driven models (Bhasme et al., 2021; Khandelwal, et al., 2020). One of the successful attempts is to use data-driven models to emulate sub-hydrological processes (e.g., overland flow and baseflow) generated from a process-based model (Meshgi et al., 2015; Tongal and Booij, 2018). Such a method allows the simulation from a data-driven model to be explained under the paradigm of a two-layer hydrological model structure (Badrzadeh et al., 2016; Corzo and Solomatine, 2007; Tongal and Booij, 2018; Wu et al., 2009). The rationale behind this method is that the underlying mechanisms of streamflow dynamics are more likely to lead to distinct behaviors in terms of flow magnitudes and seasonal characteristics (Dralle et al., 2016; Newman et al., 2015). Specifically, the baseflow mainly contributes to low-flow events, whereas intense storms give rise to high-flow events. A single global data-driven model could hardly capture both high- and low-runoff dynamics (Wu et al., 2009). Consequently, modular models (i.e., building separate models for several sub-hydrological processes) are desired for hydrological simulation and inference (Solomatine and Ostfeld, 2008). Nevertheless, most of the existing data-driven models with modular model considerations mainly focus on improving predictive accuracy (Tongal and Booij, 2018) or quantifying the contribution of different land uses (Meshgi et al., 2015). The relative contribution of relevant driving forces to surface and subsurface flow generations has not yet been studied. This requires a formal sensitivity analysis to peer inside data-driven models to understand their internals and improve the interpretability of their results – see Section 3.4 of Razavi et al. (2021) for a discussion. In irrigated watersheds, quantifying such contribution could be particularly useful in facilitating the conjunctive operation of surface water and groundwater resources, thereby reducing the risks of flooding and salinization.

In light of the above considerations, the objective of this study is to develop a baseflow-filtered stepwise clustered ensemble (BFSCE) to improve the quantification of the relative contribution of associated driving forces (e.g., precipitation and irrigation) to sub-hydrological processes (i.e., surface and subsurface flows). This study entails (i) analyzing sub-hydrological processes (including surface and subsurface flows) of three irrigated watersheds in the upper reach of the Yellow River Basin, China; (ii) identifying the relative contribution of associated driving forces to each of the sub-hydrological processes through Wilks feature importance (WFI) method; (iii) running a sensitivity analysis to examine the impact of precipitation, irrigation, and their interactions through multilevel factorial analysis (MFA) to support the relevant inference; (iv) gaining insights into runoff generation mechanisms based on the results derived from WFI and MFA.

2. Development of the baseflow filtered stepwise clustered ensemble

In this study, the baseflow in this study is defined as the groundwater contribution to streamflow, which can be affected by many factors such as changes in watershed characteristics, soil, land use, and climate (Price, 2011). The overland flow is defined as water flowing over the land surface, excluding streams and rivers. The interflow is defined as the lateral movement of water in the unsaturated zone or vadose zone, which first returns to the surface or enters a stream before becoming groundwater (Ward and Trimble, 2003). The proposed BFSCE model improved upon the previous modular models by enabling an inference module, which allows the in-depth understanding of sub-hydrological processes. The framework of the BFSCE (panel ***a*** in Fig. 1) consists of a process-based baseflow separation module, a simulation module, and an inference module. The baseflow separation module divides the streamflow into two components: baseflow and the sum of overland flow and interflow (hereinafter referred to as O&I flows) based on groundwater level fluctuations. The simulation module separately estimates the daily flow rates of baseflow, and O&I flows. The total streamflow rates

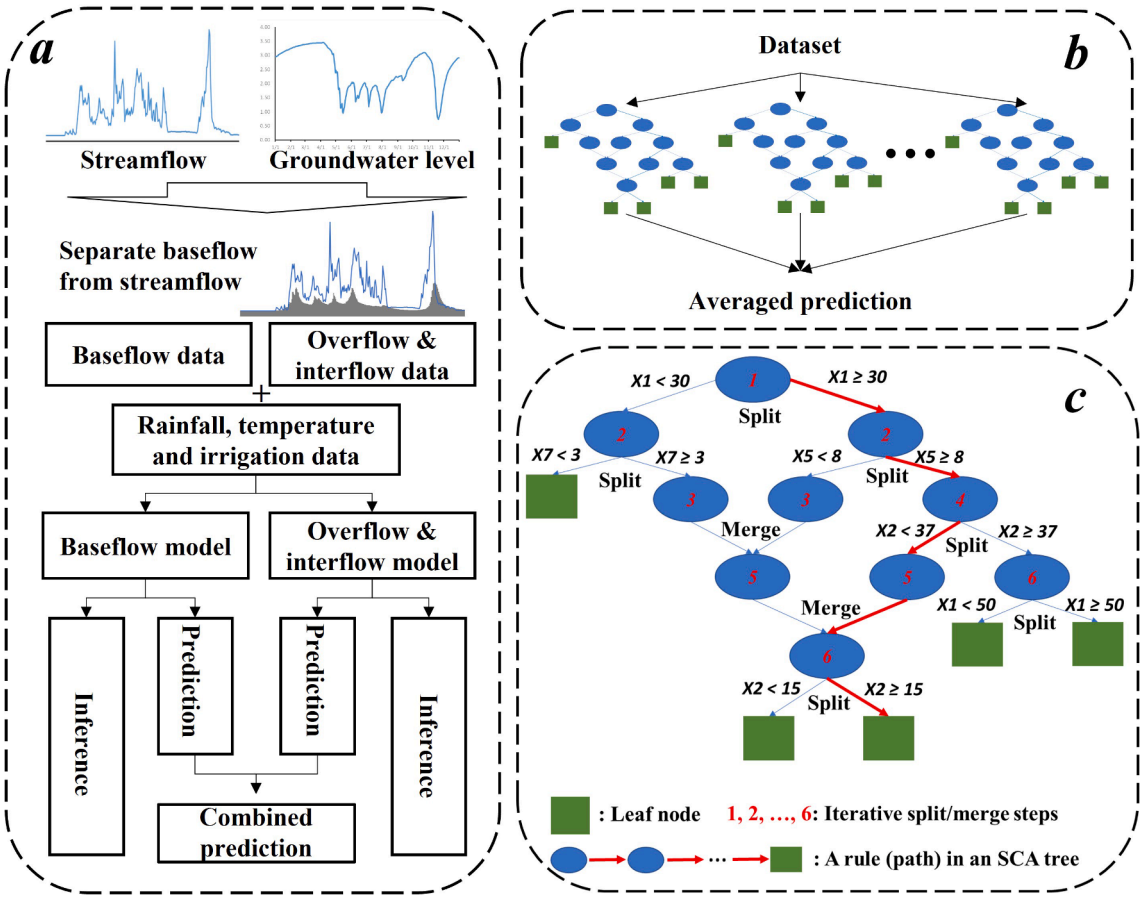

Fig. 1. The framework of the proposed BFSCE model (panel ***a***); illustration of the stepwise clustered ensemble (panel ***b***); an example of SCA tree (panel ***c***). Note that ***X1***, …, ***X7*** in panel ***c*** denote predictor variables used for the node splitting process. SCA is the ensemble member of SCE.

are obtained as the sum of these two flow components. The inference module quantifies the relative contributions of associated model predictors to sub-hydrological processes (i.e., baseflow and O&I flows).

2.1. Process-Driven baseflow separation

The baseflow separation module follows the study of Meshgi et al. (2014), which estimates the daily baseflow volume through the fluctuation of groundwater level:

$$Q_{B(t)} = Q_{B(\min)} + \sqrt{bA}\Delta h^2_{p(t+k)} \quad (1)$$

where $Q_{B(t)}$ represents the daily baseflow volume (m^3/day) at time t; $Q_{B(\min)}$ is the minimum daily baseflow volume for the entire training dataset (m^3/day); b is the coefficient related to the saturated hydraulic conductivity (Ks) ($b = 0.1Ks$); A is the total unpaved surface area in the catchment (m^2); k is the lag time between rainfall events and groundwater table responses; $\Delta h_{p(t+k)}$ is the normalized daily average of pressure head (m) ($\Delta h_{p(t+k)} = h_{(t+k)} - h_{min}$) in which $h_{(t+k)}$ is the daily averaged pressure head at time $t + k$ and h_{min} is the minimum daily averaged pressure head (m) observed for the entire dataset.

In Equation (1), the first term $Q_{B(\min)}$ represents the minimum baseflow corresponding to the deepest groundwater table in the dry period; the second term $\sqrt{bA}\Delta h^2_{p(t+k)}$ approximates the additional baseflow due to the rise in the groundwater table. Similar to Darcy's law ($q = -KA\frac{\partial h}{\partial x}$) that relates discharge to pressure head gradient, the only variable in Equation (1) is the pressure head h, which is correlated to the saturated flow. The effectiveness of this method was firstly verified in a semi-urban catchment in Singapore and then was further proved successful in a cross-site, cross-scale application in a northeastern US watershed (Meshgi et al., 2014). Once the baseflow volume is calculated through Equation (1), it is then converted to the daily mean baseflow rate (m^3/s). The overland flow rate can be calculated as the difference between the streamflow (m^3/s) and the baseflow rate (m^3/s).

Compared with other baseflow estimation methods such as the digital filter method, which removes the high-frequency signal from a streamflow time series in order to obtain the low-frequency baseflow signal (Eckhardt, 2005), Meshgi's method includes more physically meaningful information (e.g., hydrologic conductivity and groundwater level fluctuation) to derive the baseflow (Meshgi et al., 2014). This method can thus avoid using empirical parameters that require calibration. Moreover, since Meshgi's method does not require information about streamflow, the obtained baseflow time series contains signals independent of streamflow time series. These independent signals are expected to provide more information than dependent ones (derived using digital filter method) for hydrological inference. One limitation of Meshgi's method is that it requires data on groundwater fluctuation as input, which may be unavailable in many pristine watersheds.

2.2. SCE-based hydrological simulation

The baseflow and O&I flows are separately simulated through Stepwise Clustered Ensemble (SCE) (Li et al., 2021a). Stepwise Clustered Ensemble (SCE) is an ensemble of tree-structured models, and it has been used as a useful alternative to the well-known random forest (RF) (Breiman, 2001). In particular, SCE has been reported as a promising data-driven approach for streamflow simulation in terms of predictive accuracy and interpretability (Li et al., 2021a). The basic concept of regression tree ensemble is that when the number of ensemble members

(i.e., SCA trees) increases, the variation of the unstable predictions is reduced, thereby increasing the predictive accuracy.

Each SCA tree grows in accordance with a random subset of predictors sampled without replacement and a bootstrapped version of the training set, drawn randomly from the initial training dataset with replacement. This bootstrap sampling process leaves about 1/3 of the training samples as out-of-bag (OOB) samples not involved in the training process. These OOB samples can be used for validation. The tree deduction process for each SCE ensemble member follows the recursive binary splitting and merging process. As illustrated in panel ***c*** of Fig. 1, the tree deduction process begins from the top of the tree and then iteratively splits or merges the dataset (i.e., predictors and predictands) according to a certain criterion; each split is indicated via two child nodes further down the tree. A merge action is followed after each split action to reduce the probabilities that the two child nodes are split due to chance, thereby reducing the risk of overfitting. Such splitting and merging actions are repeated until no nodes can be further split or merged. Once a tree is built, the mean value of the training observations in an undividable node can be used to estimate the predicted value. For model prediction, the predictors are used as a reference to determine which node to enter for identifying the corresponding predicted values. In such a tree deduction process, the *F* test is used to determine whether (a) node(s) can be split (or merged). The calculation process for determining the splitting and merging criteria follows Li et al. (2021a) and Huang (1992). The hydrological implications of the process of an SCA tree can be explained as follows.

Suppose two groups of streamflow records (in a node) are statistically different from each other; the underlying streamflow generation mechanisms between the two groups could be different. Traditionally, when precipitation exceeds the soil infiltration capacity, infiltration-excess overland flow occurs. When rain falls on the saturated soil, saturation-excess overland flow occurs. The criterion used for splitting such a node is then interpretable because it contains the information of the possible predictor (e.g., daily precipitation >30 mm) that distinguishes the two mechanisms. The cascade sequences of the criterion throughout an SCA tree help explain the flow generation from a combination of events.

2.3. SCE-based importance rankings

The SCE-based importance ranking is achieved through the Wilks Feature Importance (WFI) method (Li et al., 2021a). Feature importance is widely used in statistical models to quantify and rank the variable importance. The obtained importance scores can be used to measure the contribution of each predictor in predicting the model output. Even though many feature importance methods have been proposed previously (Breiman, 2001; Friedman, 2001; Lundberg and Lee, 2017), they suffer from issues such as stability (i.e., a small perturbation of training data may significantly change the relative contribution of predictors) (Bénard et al., 2021; Li et al., 2021a; Schmidt et al., 2020) and fail to identify the most relevant predictors (Li et al., 2021a). Since previous methods often rely on certain objective functions (e.g., minimizing the mean squared error) to estimate the importance rankings, which mainly reflect the relative contribution of achieving the best model predictive accuracy. Due to the existence of equifinality (i.e., different importance rankings lead to the same or similar predictive accuracy), it is difficult to perform a consistent inference from the results with varying rankings. Another challenge of the existing feature importance methods is that they may fail to identify the most relevant predictors (i.e., predictors with the top ***N*** highest importance scores) (Li et al., 2021a). Our previous experiments suggested that the most relevant predictors identified by the WFI method can be universally fitted by other ML models with the highest predictive accuracy, demonstrating the robustness of the WFI method (Li et al., 2021a).

The essence of the WFI method is to quantify the effectiveness of a predictor in differentiating two groups of streamflow records across an SCA tree through the Wilks Λ statistics (Nath and Pavur, 1985; Wilks, 1967). In general, a Λ value can directly indicate the significance of the difference between two groups of data (i.e., a smaller Λ value indicates a larger difference). A detailed definition of Wilks Λ statistics can be found in Li et al. (2021a) and Li et al. (2022a). In an SCA tree, each node (***s***) with a splitting action is associated with a Λ value (**Λ**) and a predictor ($\mathbf{X_j}$). The contribution of $\mathbf{X_j}$ across the tree is calculated as:

$$WFI(X_j) = \sum_{s \in S} P_s \cdot (1 - \Lambda(s, X_j)) \quad (2)$$

where *S* is the total set of nodes, $\boldsymbol{P_s}$ is the fraction of samples in node ***s***. The estimated contributions of all predictors are then scaled to the [0,1] range. In this way, the more effective a predictor ($\mathbf{X_j}$) can differentiate two groups, the higher relative contribution (i.e., importance score) (%) can be expected to explain the variability of streamflow. Compared with other methods, the WFI method quantifies the difference between two groups of data, rather than the effectiveness of a predictor in improving model performance, and the estimated relative contributions are expected to be less biased and could potentially lead to more robust variable rankings (Li et al., 2021a).

In this study, the SCE follows the configuration of Li et al. (2021a) for the optimum model performance. Specifically, the number of ensemble members (*Ntree*) was set to 200; the minimum samples to be split in a node (*Nmin*) were set to 5; the percentage of predictors selected for each tree (*Mtry*) was set to 0.5 (i.e., indicating half of the predictors were selected for training each SCA tree); the significance level (α) for the *F* test was set to 0.05.

3. Study of hydrological processes in the ancient Yellow River irrigation system of Ningxia

3.1. Overview of the ancient Yellow River irrigation system in Ningxia

The ancient Yellow River irrigation system in Ningxia dates back >2000 years and still plays a crucial role in Ningxia province. The irrigation system is located in an arid and semi-arid region, with annual precipitation of less than 200 mm and annual potential evaporation of >1,100 mm (gauged by E601). Over 90 % of water use in this region comes from the Yellow River, which flows through the province for about 400 km. By taking advantage of the special geological feature that the river elevation is higher than the elevation of surrounding fields, water has been diverted naturally from the Yellow River to the crop fields through irrigation canals by gravity. Until 2020, it was the largest irrigation system in Northwest China, with an irrigated area of 552,000

Table 1
Volumes of local water diverted from the Yellow River and returned to the Yellow River in period 2003 to 2015. Note: return ratio is calculated as the water returning to the Yellow River divided by the water diverted from it. The data were obtained from the Ningxia Water Resources Bulletin from 2003 to 2015 (Ningxia Water Conservancy, 2015).

Year	From Local (10^9 m^3)	From the Yellow River (10^9 m^3)	Return to the Yellow River (10^9 m^3)	Return ratio
2003	8.3	55.7	25.7	46 %
2004	6.7	67.3	33.1	49 %
2005	6.9	71.1	33.5	47 %
2006	6.8	70.8	35.6	50 %
2007	6.9	64.1	29.9	47 %
2008	6.7	67.5	31.1	46 %
2009	7.2	65.1	30.2	46 %
2010	7.8	64.6	33.2	51 %
2011	8.3	65.2	34.0	52 %
2012	8.4	60.9	33.5	55 %
2013	8.8	63.3	33.4	53 %
2014	8.7	61.6	31.4	51 %
2015	8.3	62.0	31.0	50 %

ha. Table 1 shows the water diverted from and returned to the Yellow River. About half of the irrigation water diverted from the Yellow River will eventually return through the irrigation return flow (Table 1).

The study area is located at the West alluvial plain of the Yellow River, crisscrossed by irrigation canals and drainages. Fig. 2 shows the detailed layout of the irrigation and drainage systems, including six main irrigation canals and five main drainages. The water diversion usually starts in April when the water is mainly used to recharge lakes and reservoirs as well as to sustain the local ecosystem. The crop water demands begin to surge in May when the three main crops (i.e., rice, wheat, and corn) are at their critical growth stages and thus are most sensitive to water deficit. The irrigation stops at the end of September when corn is harvested and restarts at the end of October, lasting until December. Irrigation before winter is to freeze the topsoil to retain water content and facilitate the next growing season.

The soil type of the study area varies from loamy sand in the south to sandy loam in the north. The saturated hydraulic conductivity varies from 20 to 50 m/day in the south to 8–20 m/day in the north. The groundwater table depth reaches a maximum of 30 m on the west edge of the study area near mountains and is thinning towards the river plain with less than 1 m. On the other hand, the salinity increases from 0.5 g/L to 3 g/L as the aquifer thickness decreases.

3.2. Data collection and preprocessing

The datasets used in this study include daily mean streamflow rate, daily mean water diversion flow rate, daily precipitation, daily mean air temperature, daily mean evaporation, and daily mean groundwater level. All daily time series span from 2003 to 2015. As shown in Fig. 2, all gauging stations are evenly distributed within the study area. In recent years, urban development has changed the regime of streamflow in some drainages. For instance, the drainage systems of Yinchuan (i.e., the capital city of Ningxia) had gone through pipeline upgrades and tributary adaptations (e.g., merge, redirect and expand). A recent study (Li et al., 2021b) suggested that the surrounding farmlands of Yinchuan has largely vanished due to urban expansion, leading to a significant reduction in irrigation water and an increase in groundwater extraction. Mi et al. (2020) also suggested that several water-saving strategies such as irrigation canal lining and the practice of deficit irrigation have substantially changed the regime of the hydrological process in this area. From the perspective of flood risk mitigation, the third drainage was reinforced as floodways with connections to flood storage and lakes to improve flood control capacity. Therefore, the above-mentioned non-stationarities could be a primary factor compromising the performance of data-driven models.

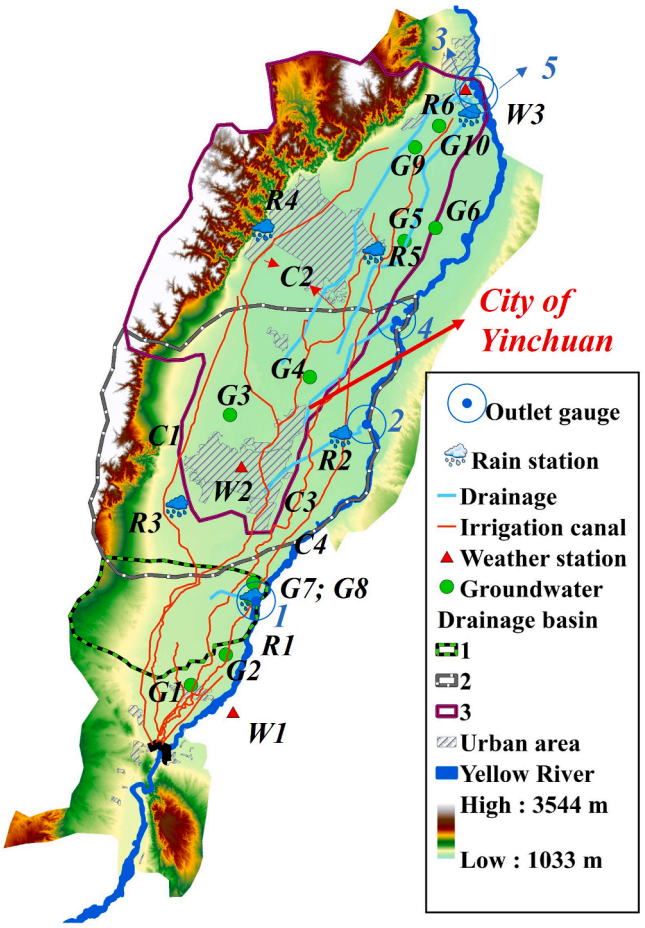

Fig. 2. Map of the study area.

Detrending is an effective way to address the issue of non-stationarity. Mann Kendall trend (MK) test (Kendall, 1948; Mann, 1945) was performed to examine the trend of streamflow time series before detrending. Based on the types of runoff generation mechanisms, the hydrological process was divided into two Spring irrigation (SI) and Winter irrigation (WI) periods. Separate models were built for these two periods. The SI spans from April to September, while WI spans from October to March. Fig. 3 illustrates the stationarity analysis for the drainage streamflow derived from the two-sided seasonal MK test, which ran a separate MK trend test for each of *m* seasons (i.e., SI and WI in our case). As shown in Fig. 3, the issue of non-stationarity exists in the streamflow time series of second, fourth, and fifth drainages at the significance level of 0.05. In particular, the second drainage indicates a significant monotonic decreasing trend for the SI periods. The fourth and fifth drainages show a significant monotonic increasing trend for the WI periods.

To address the issue of non-stationarity, the interwoven drainages with strong hydrological connections were counted as one single drainage. Specifically, the time series for the sum of second and fourth drainage flows suggest an increased stationarity for both SI and WI periods. The increased stationarity is also found for the sum of third and fifth drainage flows. Therefore, streamflow time series representing three drainage basins will be modeled. As shown in Fig. 2, the first drainage basin includes one drainage, the second drainage basin consists of the second and the fourth drainages, and the third drainage basin consists of the third and the fifth drainages. The summary of basin properties is provided in Table 2.

The combined streamflow time series for each basin are further detrended using a linear model. Specifically, daily time series were divided into two subsets, including one from 01/01/2003 to 31/12/2011 for model training and the other from 01/01/2012 to 31/12/2015 for testing. The streamflow time series over the training period were

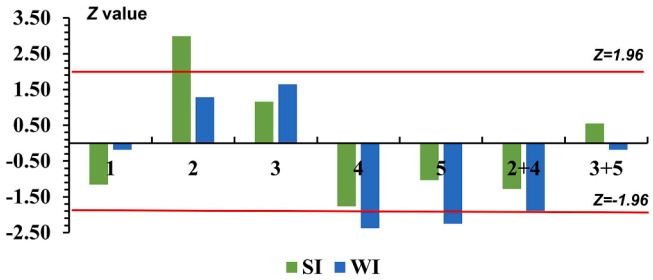

Fig. 3. The values of Z_{MK} from the two-sided seasonal Mann Kendal test for single and combined drainages (***1***,***2***, …, ***5*** indicate the ID number of drainages. "***2*** + ***4***" and "***3*** + ***5***" denotes the combined drainages). SI represents Z_{MK} values for the monthly streamflow time series of April to September; WI represents Z_{MK} values for October to March. The decision rule is: Reject H_0 if $Z_{MK} < -1.96$ or if $Z_{MK} > 1.96$.

Table 2
Summary of basin properties. Note that soil type was obtained based on field investigation; the Ks values were derived from previous studies by Gao et al. (2003); the median Ks value was used to calculate the baseflow in Equation (1).

Drainage Basin ID	Approximated area (m^2)	Soil type	Ks (m/day)
First	8.0×10^8	loamy sand	20 ~ 50
Second	3.2×10^9	sandy loam	8 ~ 20
Third	4.5×10^9	sandy loam	8 ~ 20

then used to calculate a regression line. After that, the deviations from the least-squares fit line were subtracted from the streamflow time series for both training and testing. The generated time series were used for the modeling process.

The daily averaged pressure head for each drainage basin was calculated as the mean value from all the monitoring wells in the basin. The lag time between the rainfall/irrigation events and groundwater table responses for each basin was determined through correlation analysis, as shown in Table 3. This table indicates that the lag time for all the basins is less than one day (i.e., lag = 0 has the highest absolute correlation coefficient). The lag time agrees with previous studies by Meshgi et al. (2014), who found that shallower groundwater tables had shorter lag times with rainfall events.

The input structure of the SCE was determined in consideration of antecedent watershed conditions, which are critical to streamflow predictions (Li et al., 2022; Woodhouse et al., 2016). In this study, the moving sum temperature, precipitation, and irrigation ranging up to 30 days before the date of predictions represent the antecedent watershed conditions for each of the three drainage basins. Such a preprocessing method hypothesizes that the different window lengths of moving sums can act as surrogates representing different dynamics of watershed conditions such as soil moisture, dryness of the soil, as well as the water content in the unsaturated and saturated zones. The generalized expression for such an SCE can be written as:

$$Q = f\left(P_{i1}, P_{i3}, P_{i5}, I_{j1}, I_{j3}, I_{j5}, I_{j7}, I_{15}, I_{30}, T_{k3}, T_{k5}\right) \quad (3)$$

where Q denotes the daily streamflow at time t; P, I, and T denote precipitation, irrigation, and temperature, respectively; P_{i1} denotes the precipitation for time t at the i^{th} rain station. P_{i3} denotes the mean precipitation for time *t-2, t-1,* and t at the i^{th} rain station; the number "3" indicate the length of moving window. The same denotation applies to other predictor variables. I_{15} and I_{30} denote the mean daily irrigation flow rate for the past 15 and 30 days, respectively. Subscripts *i, j,* and *k* denote the rain station, irrigation canal, and climate station index, respectively. Table 4 lists the forcing data of BFSCE selected for each of the drainage basins.

3.3. Connecting importance rankings with physical effects

Once the importance rankings are obtained using the WFI method, they can be associated with the physical effect of hydrological processes. In general, the importance score for a predictor indicates to-what-extent the predictor may affect the flow simulation. Given the physics we know that the longer the water is retained in the soil, the more likely it could reach the saturated zone. On the contrary, the shorter the water is retained in the soil, the more likely it could evaporate or release as interflow. Thus, the importance score can tell which soil layer (saturated and unsaturated) the associated predictor is more likely to contribute to. Suppose the importance score for I15 (i.e., irrigation volume with a 15-day moving window) is higher for O&I flow simulation than baseflow. It indicates that the water in unsaturated zone is more likely to reach a dynamic equilibrium state (i.e., released water is a function of irrigation water) with a 15-day irrigation window. Such a dynamic equilibrium state provides stronger signals to facilitate the model capture the O&I flow dynamics. In contrast, the water released from the saturated zone is less helpful in capturing the baseflow dynamics.

Table 3
Correlations between daily rainfall/irrigation volumes and daily groundwater level for each studied basin.

Drainage basin ID	lag = 0	lag = 1	lag = 2	lag = 3	lag = 4	lag = 5
First	**−0.71**	−0.69	−0.67	−0.65	−0.62	−0.60
Second	**−0.64**	−0.63	−0.61	−0.59	−0.57	−0.55
Third	**−0.69**	−0.68	−0.67	−0.65	−0.63	−0.62

Note that lag = 1 indicates the correlation between rainfall/irrigation event at time $\boldsymbol{t}$ and groundwater level at time $\boldsymbol{t+1}$, and so forth.

Table 4
Forcing data for each of drainage basins.

Dataset	Canal/ station	1st drainage basin	2nd drainage basin	3rd drainage basin
Canal	C1	√	√	√
	C2	√	√	
	C3		√	√
	C4	√	√	√
Rain station	R1	√		
	R2		√	
	R3		√	
	R4			√
	R5		√	√
	R6			√
Weather station	W1	√		
	W2		√	√
	W3			√
Groundwater station	G1			
	G2			
	G3		√	
	G4		√	
	G5			√
	G6			√
	G7	√		
	G8	√		
	G9			√
	G10			√

3.4. Evaluation of the BFSCE model performance

Before investigating the interpretability of BFSCE, we first evaluate model performance. Various performance metrics were used to provide a comprehensive assessment of model performance. These metrics include the coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), Kling-Gupta efficiency (KGE) (Gupta et al., 2009), Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and Volumetric efficiency (VE) (Criss and Winston, 2008). These indices have been widely used in hydrological studies since they are reliable and easily interpretable (Tongal and Booij, 2018; Song et al., 2022). The proposed BFSCE was benchmarked against SCE and random forest (RF) (Breiman, 2001). Table 5 shows the features of BFSCE model and benchmark models.

Table 5
Features of BFSCE model and benchmark models.

Model	Feature
RF	• The classic regression tree ensemble model, which is known for its high predictive accuracy (Fernández-Delgado et al., 2014). • Equipped with two feature importance methods (Breiman, 2001).
SCE	• Advanced regression tree ensemble with improved predictive accuracy and interpretability. • Equipped with an innovative feature importance method (i.e., Wilks feature importance).
BFSCE	• Combines SCE and a process-based baseflow separation module. • Can help trace the origin of streamflow from the perspective of sub-hydrological processes.

In the presence of baseflow separation, BFSCE generally outperformed SCE and RF for the testing period (Table 6). The results suggest that the subtraction of baseflow enables data-driven models to identify the critical information reflecting the overland runoff process and thus improve model performance (even though, in many cases, the improvement is limited, with the MAE differences being around 0.01). Notably, slightly lower performances were observed for the SI period compared to the WI period in the first and third drainages. One reason is that the irrigation scheduling for SI period is more complicated than for WI period due to flood control and optimal irrigation water allocation. Such complex irrigation scheduling in SI period was further compounded by precipitation. As a result, the training data for the SI period would be less representative than for the WI period, leading to a weaker model than the one built for the WI period. In the second drainage basin, the model complexity was elevated due to non-stationarity, leading to lower performance in the WI period.

Figs. 4 to 6 illustrate the BFSCE hydrographs for the three drainage basins in the testing period. The training and validation hydrographs are shown in Figures S1 to S3. The simulated streamflows generally agreed with daily runoff observations. Specifically, the NSE, R^2, and KGE values ranged from 0.7 to 0.9 for all three basins. However, some peak flows were underestimated by the model. One of the reasons is that the SCE is a rule-based nonparametric simulation approach, implying that the model prediction cannot extrapolate beyond the range of the training dataset. As a result, the BFSCE might underestimate the flow records higher than the highest values in the training dataset. For instance, the highest O&I flow over the training period was mainly caused by the compounded effect of irrigation and precipitation in the first drainage basin. The model thus considers both irrigation and precipitation (e.g., $X_1 > 70$ m^3/s and $X_2 > 20$ mm/day; X_1 and X_2 represent the 5-day average irrigation flow rate and precipitation, respectively) to classify the flow as the highest one. However, the O&I flow of a similar magnitude was outside the intensive irrigation season (with $X_1 < 30$ m^3/s) and was mainly caused by a rare precipitation event (with $X_2 > 40$ mm/day) in the testing period. In such case, even though the precipitation rate is the highest among the historical records, the model may not consider the resulting flow rate as high as the highest record over the training period. Such an underestimation is also occurred for the baseflow simulation, suggesting that rare and unseen events (including the compounded ones) could be a challenge to such models in terms of predictive accuracy. The second drainage basin, however, experienced a non-stationary change during the testing period, leading to increased groundwater levels (Fig. 7). The non-stationarity was primarily due to more irrigation water being infiltrated into the unsaturated zone rather than turning into streamflow. Since the model was built based on the training dataset, it led to an overestimation of streamflow.

3.5. Hydrological inference for sub-hydrological processes

The importance scores of predictors used for simulating the two sub-hydrological processes (i.e., baseflow and the O&I flows) were investigated, as shown in Fig. 8 (SI) and Fig. 9 (WI). In general, the importance scores for the two sub-hydrological processes showed significant differences. For the SI period in the first drainage basin, the predictors associated with 7-day and 5-day averaged irrigation water (of C2 canal) were ranked as the highest two predictors for simulating the O&I flows. In comparison, these two predictors were considered non-critical for baseflow simulation. Instead, the 30-day averaged irrigation water (*I30*) was considered the most important predictor for baseflow simulation. On the contrary, O&I flows had a more acute response to rainfall and irrigation. This is can also be demonstrated by the fact that precipitation-related predictors (e.g., R1(3), R1(5), R2(3), …, and R6(5)) achieved higher importance scores for O&I flow simulation than the baseflow one. The predictors *TW15* and *TW25* were evidenced as the second most important predictor for the baseflow simulation of the SI period in the first and second drainage basins, respectively. This was because the intensive evaporation coupled with a shallow groundwater table and high permeable soil enabled water in the saturated zone to transport quickly to the unsaturated zone due to the capillary action and thus affected baseflow.

The results also show that importance score for the SI period is higher than the WI. The median values of importance scores for simulating overland and interflows in SI periods were 3.6 %, 2.6 %, and 3.1 % for the first, second, and third drainage basins, respectively, whilst the values in WI were 3.2 %, 1.7 %, and 1.7 %, respectively. This was because the irrigation practices in the SI period were often compounded with rainfall and intensive evapotranspiration, which increased the system complexity. In addition, we found that *I30* achieved higher importance rankings for simulating O&I flows in the SI period than the WI. Such a difference could be caused by a high daily irrigation flow rate for the WI period, allowing more water discharged from interflow rather than baseflow. Owing to the retention effect of the porous soil, *I30* always achieved the highest importance score for the baseflow simulation.

3.6. Tracing the origin of streamflow

Fig. 10 shows that the importance scores representing different stations/canals were aggregated (i.e., summed), which allowed a direct comparison among three basins. The red bars indicate the importance scores from streamflow simulation, while the blue and black bars

Table 6
Comparison of model performance between BFSCE, SCE, and RF. Note that SCE was trained using the same settings (i.e., hyperparameters and predictors) with BFSCE; the hyperparameters for RF were turned using the grid search method (Gilli et al., 2019).

		BFSCE	SCE	RF			BFSCE	SCE	RF
1st SI	MAE	**1.71**	1.75	1.78	1st WI	MAE	**0.68**	**0.68**	0.71
	RMSE	**2.43**	2.49	2.57		RMSE	**1.27**	1.28	1.31
	NSE	**0.81**	0.80	0.78		NSE	**0.92**	**0.92**	0.91
	R^2	**0.82**	0.81	0.81		R^2	**0.92**	**0.92**	**0.92**
	KGE	**0.77**	**0.77**	0.75		KGE	**0.95**	**0.95**	**0.95**
	VE	**0.81**	0.80	0.80		VE	**0.82**	**0.82**	0.81
2nd SI	MAE	**2.25**	2.26	2.41	2nd WI	MAE	**1.68**	1.68	1.71
	RMSE	**2.91**	2.94	3.11		RMSE	**2.31**	2.32	**2.31**
	NSE	**0.67**	0.66	0.62		NSE	**0.53**	**0.53**	**0.53**
	R^2	**0.77**	**0.77**	0.75		R^2	**0.68**	0.67	**0.68**
	KGE	**0.80**	0.79	0.77		KGE	**0.74**	0.74	0.73
	VE	**0.83**	0.82	0.81		VE	**0.74**	**0.74**	**0.74**
3rd SI	MAE	**4.24**	4.25	4.27	3rd WI	MAE	**2.09**	2.10	2.08
	RMSE	**6.02**	6.04	6.11		RMSE	**2.99**	3.02	3.02
	NSE	**0.55**	**0.55**	0.53		NSE	**0.79**	0.78	0.78
	R^2	**0.58**	0.57	0.56		R^2	**0.79**	**0.79**	**0.79**
	KGE	**0.53**	**0.53**	**0.53**		KGE	**0.79**	0.78	**0.79**
	VE	**0.64**	**0.64**	**0.64**		VE	**0.54**	**0.54**	**0.54**

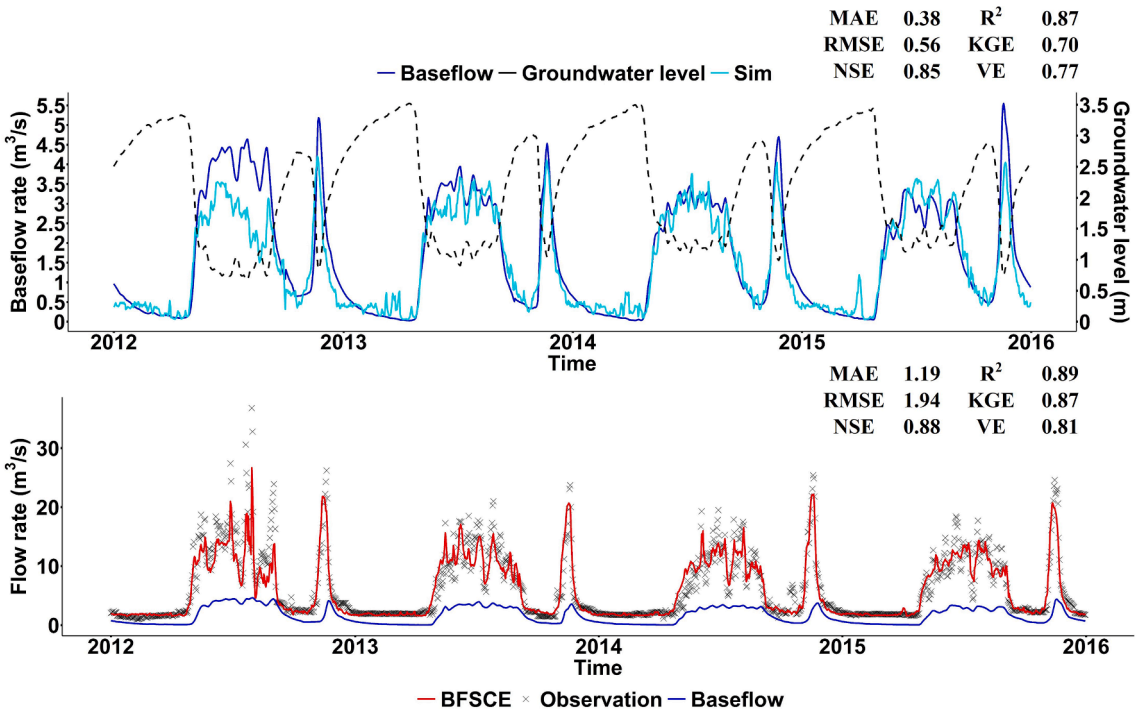

Fig. 4. Hydrographs for the testing period in the first drainage basin.

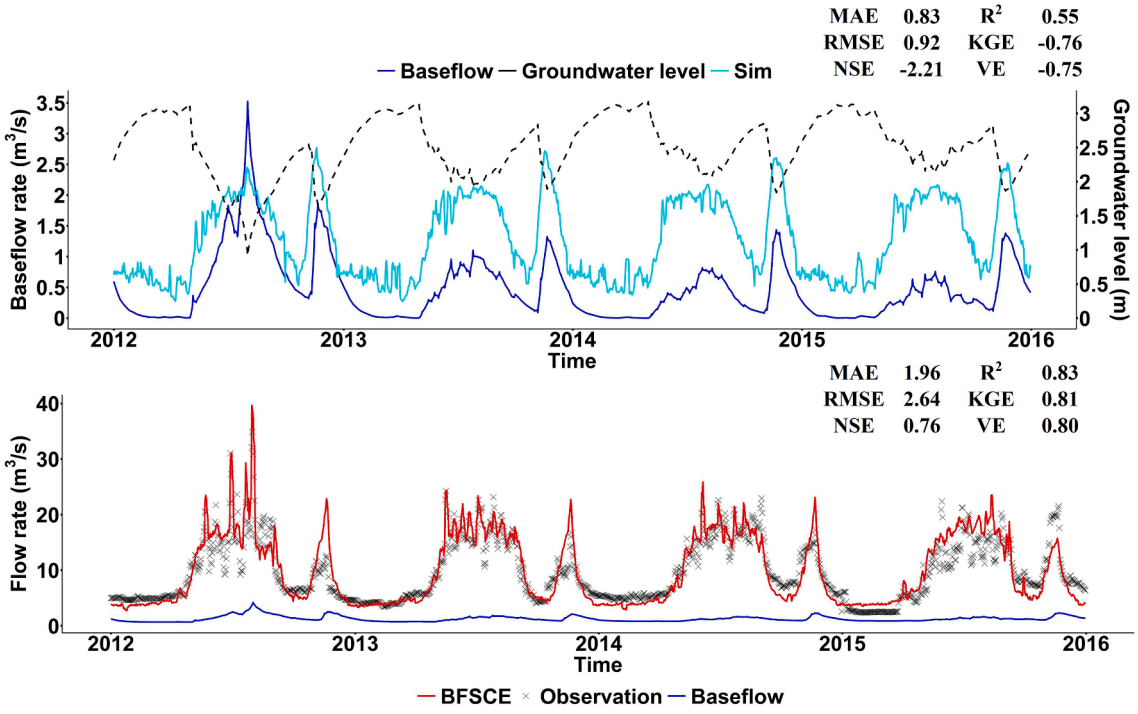

Fig. 5. Hydrographs for the testing period in the second drainage basin.

represent those from O&I flow and baseflow simulation, respectively. The results showed that compared with hydrological inference from the entire streamflow process, the inference from sub-processes provided more physics-based insights into the role a predictor plays in streamflow generation. For example, weekly (7-day) averaged irrigation water was considered the most important predictor of streamflow simulation during the SI period of the first drainage basin. This was mainly due to the substantial contribution of O&I flows (i.e., the highest blue bar). Monthly (30-day) averaged irrigation was ranked as the second-highest important predictor of streamflow simulation. Such a model decision was associated with the substantial baseflow contribution (i.e., the highest black bar). The above inference suggested that weekly averaged irrigation water mainly facilitated the model to capture the dynamics of O&I flows, while monthly averaged irrigation water facilitated the model to capture the groundwater dynamics. This was because the unsaturated zone reached a dynamic equilibrium state with weekly irrigation, leading to a constant water release that facilitated the model to capture the dynamics of O&I flows. Since the unsaturated zone was not

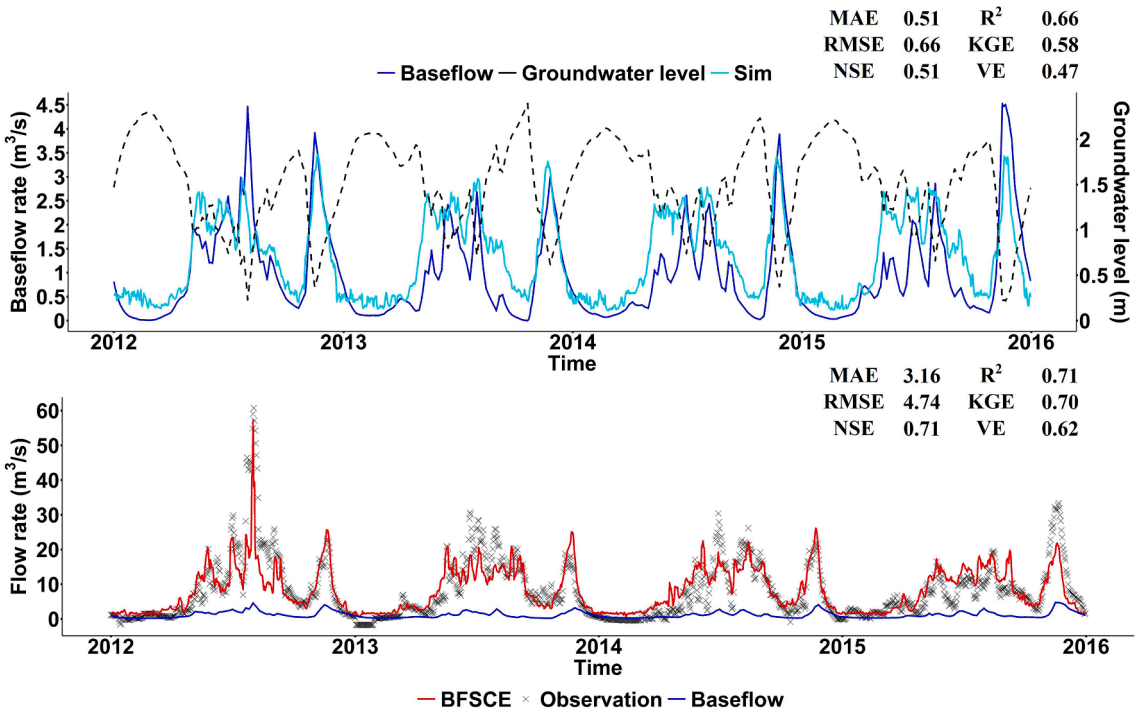

Fig. 6. Hydrographs for the testing period in the third drainage basin.

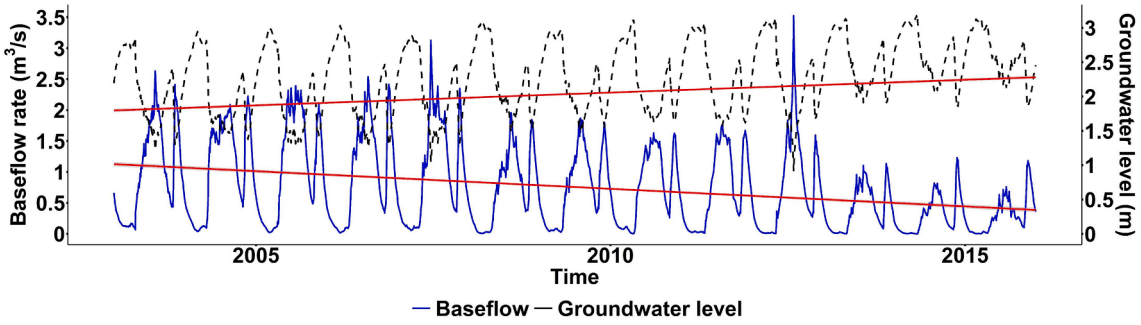

Fig. 7. Groundwater level (dashed line) and observed baseflow hydrographs (blue line) in the second drainage basin for the period from 2003 to 2015.

uniform due to the variations in landscape, irrigation, and climatic factors, the dynamic equilibrium state also varied correspondingly. Consequently, irrigation with different window lengths suggested varying importance scores in capturing the dynamics of O&I flows. In the third drainage basin, the 5-day averaged precipitation was responsible for capturing the dynamics of O&I flows, while monthly irrigation water facilitated the model to capture the groundwater dynamics. These findings agreed with our field investigation that the third drainage in recent years had been reinforced as floodways connected to flood storage and lakes to embrace the floods from the mountains on the West (Fig. 2).

Even though the monthly averaged irrigation showed significant importance for baseflow simulation during the WI period, it did not contribute much to streamflow simulation. Meanwhile, 3-day, 7-day, and 15-day averaged irrigation had the highest contributions to O&I flows and streamflow simulation for the first, second, and third drainages, respectively. These results suggested that O&I flows were the dominant flow components during the WI period, while baseflow had a minimal contribution to streamflow. These results agreed with our investigation of winter irrigation schedules. The high volume of irrigation water flushes over the irrigation field (bare ground) in a short period to freeze the topsoil. Therefore, the irrigation water quickly percolates into the saturated zone without being uptook by crops, thereby leading to substantial O&I flows.

4. Discussions

4.1. Does the importance score lead to physically consistent results?

Traditionally, importance scores mainly reflect the relative contribution of predictors to model predictive accuracy, rather than physically influence the model predictand (e.g., volume of streamflow) (Molnar, 2020; Rajbahadur et al., 2021). A frequently asked question in statistical-based hydrological inference is whether the importance scores can lead to physically consistent results. In this study, even though the physical knowledge suggested that the obtained inferences were reasonable, we still need more rigorous evidence to test whether the relative importance of predictors achieved from BFSCE also reflects their relative significance on the changes of streamflow. To this end, a multilevel factorial analysis (MFA) was performed as follows.

(1) Reprocess the testing dataset of SI period by multiplying its irrigation and precipitation associated time series by vectors of (1.1,

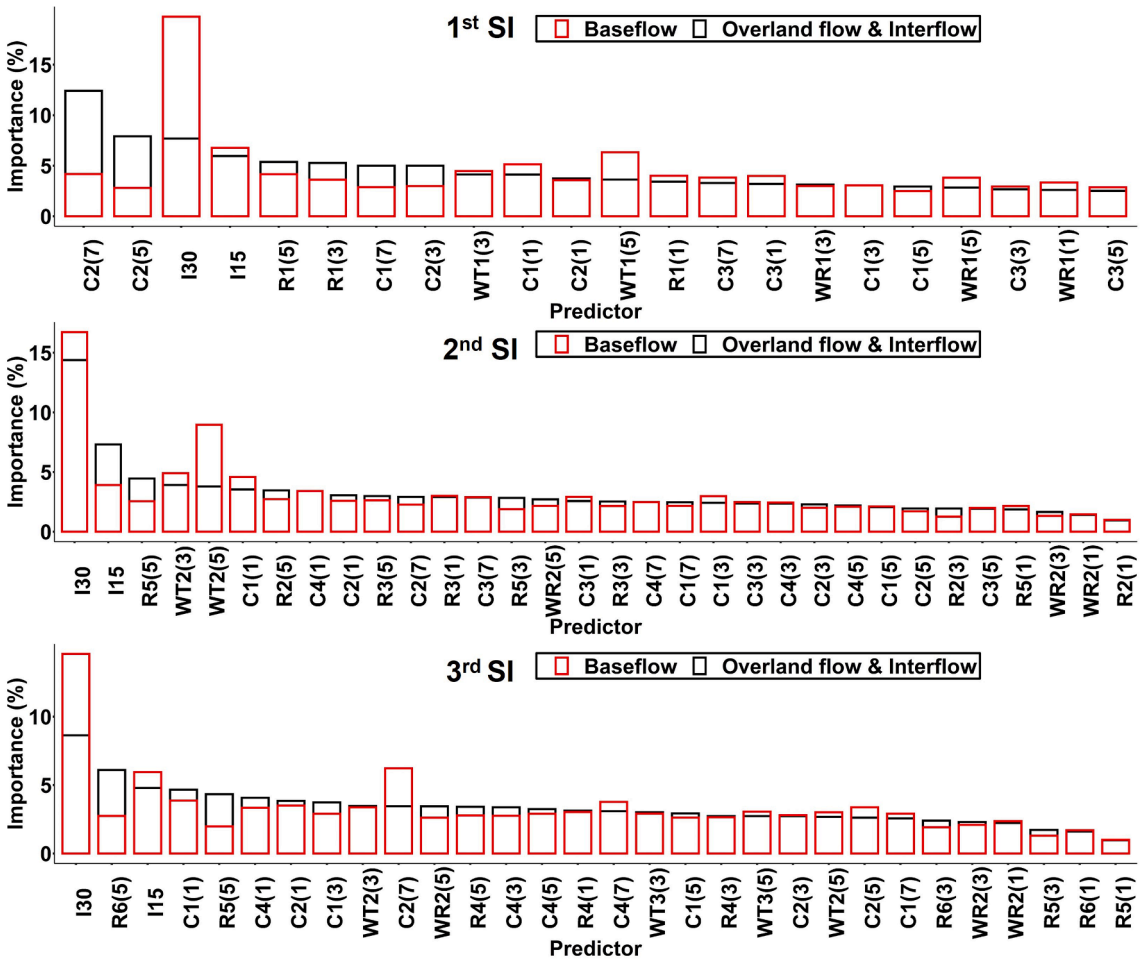

Fig. 8. Importance scores of SCE model predictors for simulating the two sub-hydrological processes in the SI period. Note that the same set of predictors was used for simulating baseflow and O&I flows; ***Ci(n)*** denotes the predictor representing the mean daily irrigation flow rate during the period [*t-n, t*] gauged at irrigation canal ***Ci***; similar expressions also apply to predictors for rain station ***Ri(n)***, the rain gauge at weather station ***WRi(n)***, and air temperature gauge at weather station ***WTi(n)***; ***I15*** and ***I30*** denotes the mean daily irrigation flow rate for the past 15 and 30 days, respectively.

0.9), (1.1, 1.0), (1.1, 1.1), (1.0, 0.9), (1.0, 1.0), (1.0, 1.1), (0.9, 0.9), (0.9, 1.0) and (0.9, 1.1), separately.

(2) Train 10 sets of BFSCE and SCE models (for the SI period of the training dataset) with different random seeds.

(3) Feed each of the trained models with the reprocessed testing datasets.

(4) Calculate the mean daily streamflow, baseflow, and O&I flow volumes (m^3) based on the results obtained from each model.

(5) Use flow volumes as response variables to calculate the main effects of individual factors (i.e., irrigation and precipitation), their interactions, and model error.

(6) The contribution of each factor can be calculated using its sum of squares divided by the total sum of squares (Montgomery, 2017).

Fig. 11 presents the main effects of irrigation and precipitation on streamflow, O&I flow, and baseflow, with detailed information shown in Tables S1 to S3. The results suggested that precipitation for the first drainage (in the red box of Fig. 11) had little contribution to the streamflow volume during the SI period. In addition, precipitation contributed little to baseflow volume for all three basins (shown in the blue box). The differences between the relative contributions (importance score) obtained from WFI (hereinafter shown as $\boldsymbol{RC}_{\mathrm{WFI}}$) and those obtained from MFA (hereinafter shown as $\boldsymbol{RC}_{\mathrm{MFA}}$) are shown in Fig. 12. The $\boldsymbol{RC}_{\mathrm{WFI}}$ of irrigation (or precipitation) for simulating streamflow or its sub-processes was calculated as the mean value of the three highest importance scores (as indicated in Fig. 10) associated with the corresponding factor. $\boldsymbol{RC}_{\mathrm{WFI}}$ reflects the relative contribution of precipitation and irrigation in describing the flow dynamics (i.e., model performance), while $\boldsymbol{RC}_{\mathrm{MFA}}$ reflects the sensitivity of flow volume change (m^3) in response to precipitation and irrigation. The hypothesis is that if a predictor is more important in the modeling process, it will be more influential to the changes on streamflow or its sub-processes. Fig. 12 showed that the $\boldsymbol{RC}_{\mathrm{WFI}}$ (the top panel) agreed with the $\boldsymbol{RC}_{\mathrm{MFA}}$ (the bottom panel). Precipitation accounted for the highest $\boldsymbol{RC}_{\mathrm{MFA}}$ to streamflow generation in the third drainage basin, and it also had the highest $\boldsymbol{RC}_{\mathrm{WFI}}$ in the modeling process. The results suggested that compared with the first and second drainage basins, precipitation in the third drainage basin showed higher importance in both flow simulation (from the model performance perspective) and formulation (from the physics perspective). This is because the streamflow in the third basin contains flash floods from the mountains (shown in Fig. 2) during the rainy season, making precipitation a dominant factor influencing the streamflow. Compared with the second and the third drainage basins, precipitation in the first drainage basin achieved the lowest $\boldsymbol{RC}_{\mathrm{MFA}}$ to streamflow generation, corresponding to the lowest $\boldsymbol{RC}_{\mathrm{WFI}}$ among the three basins. This is because the unit area irrigation volume in the first drainage basin is the highest among all three basins, leading to lower importance in precipitation.

Precipitation showed higher $\boldsymbol{RC}_{\mathrm{MFA}}$ for the O&I flow generation, and the same pattern could also be achieved from the results of $\boldsymbol{RC}_{\mathrm{WFI}}$. This is because precipitation had a higher rate (mm/hour) than irrigation in general and was more likely to generate infiltration excess flow,

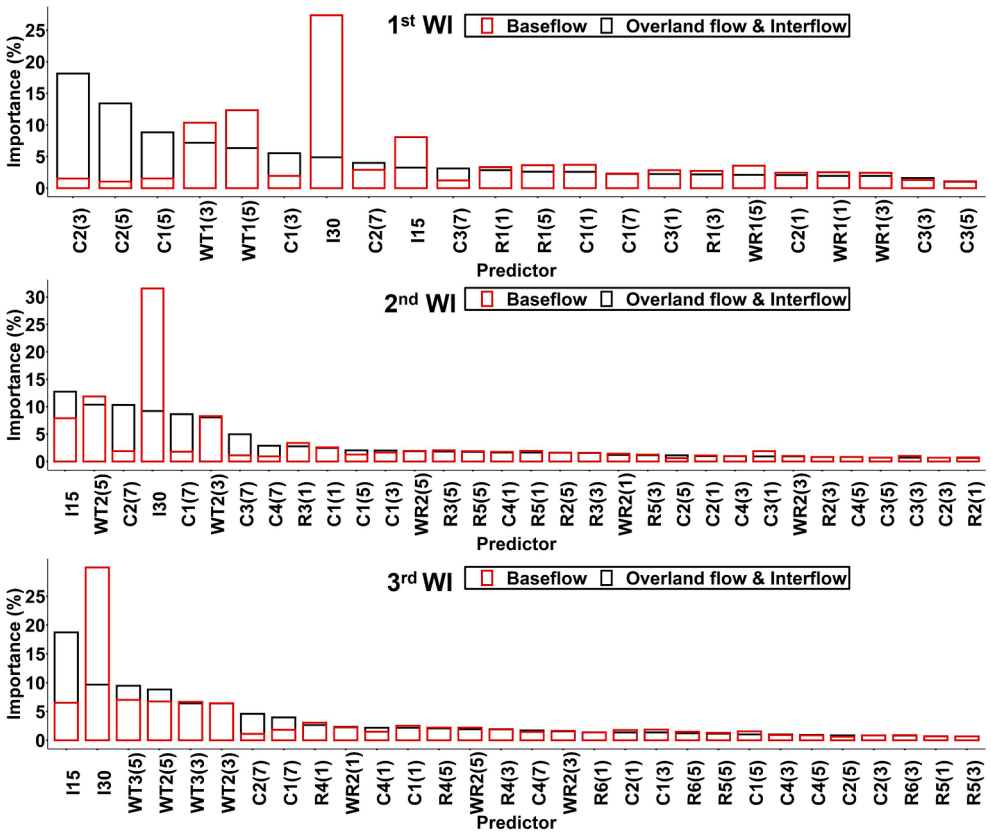

Fig. 9. Importance scores of SCE model predictors for simulating the two sub-hydrological processes in the WI period. The notations are as same as in Fig. 8.

contributing to the O&I flows. On the contrary, irrigation water was more likely to percolate and become interflows and baseflows, leading to a lower contribution to O&I flow generation. These findings proved our hypothesis and suggested that model inference from SCE could explain the hydrological process using the knowledge from process-based models.

It should be noted that precipitation in the first drainage basin barely influenced the flow volume. Nevertheless, its contribution still accounted for 7.6 % (over half of the contribution for irrigation) in the modeling process. The results indicated that precipitation might not significantly change flow volume but influence other the flow characteristics such as timing of flow peaks.

4.2. *What role does relative importance of driving forces play in understanding hydrological processes?*

One may wonder what role relative importance of driving forces plays in gaining insights into hydrological processes, which was associated with process-based models. Even though the process-based hydrological models have been studied for more than a century (Mulvany, 1850), hydrological processes under some complicated conditions, such as in coastal basins where the hydrological processes are influenced by tide and current fluxes (Armenio et al., 2017; Bevacqua et al., 2017), as well as watersheds with intensive human activities (Dewandel et al., 2008; Gosain et al., 2005) are still less-studied and need additional knowledge to gain a thorough understanding. Even for the most commonly studied watersheds, a process-based model may not consider every aspect of its driving forces, such as large-scale climatic indices (e.g., El Niño) (Li et al., 2021a).

In BFSCE, complex relationships between streamflow (or its sub-components) and its driving forces can be examined without an assumption of any linear or non-linear functions. With the provision of the state-of-the-art dataset (e.g., soil moisture data (O'Neill et al., 2019), surface mass change data (for the estimation of groundwater change) (Landerer and Swenson, 2012), vegetation growth data (Myneni et al., 2015), and snowmelt data (Hersbach et al., 2018)), the resulting relative contribution thus carries the critical information that may be neglected by a process-based model. Such information can help us gain a more in-depth understanding of hydrological sub-processes.

In terms of the irrigated watersheds, it is difficult for a conventional data-driven model to achieve a robust quantification of how irrigation and precipitation influence the sub-hydrological processes. The cutting-edge feature importance method in the BFSCE model suggested that the hydrological inferences were valid for both the entire hydrological process and its sub-processes. The modular models specialized for sub-hydrological processes can help identify the dominant factor in the saturated and unsaturated zones. Such an understanding can further help decision-makers implement the best management practices for mitigating the risks of flooding and salinization.

4.3. *Challenges and opportunities*

The proposed BFSCE model has some limitations. Since it cannot extrapolate beyond the historical observations, it may face difficulties in evaluating the impact of changing climate or other non-stationary dynamics. In this study, the non-stationary dynamics in the second drainage basins led to a significant overestimation of baseflow. A possible way to improve the simulation is to extend the coverage of training data. Such approaches include but are not limited to using process-based models as an emulator of the real system to generate synthetic data in parts of the problem space where actual data is sparse or unavailable (Razavi, 2021) and training one model with big data covering hundreds of basins with the assumption that data from one basin can inform inference in other basins (Kratzert et al., 2019b).

The selection of predictors is another challenge to the proposed BFSCE model. In this study, we used irrigation and precipitation of

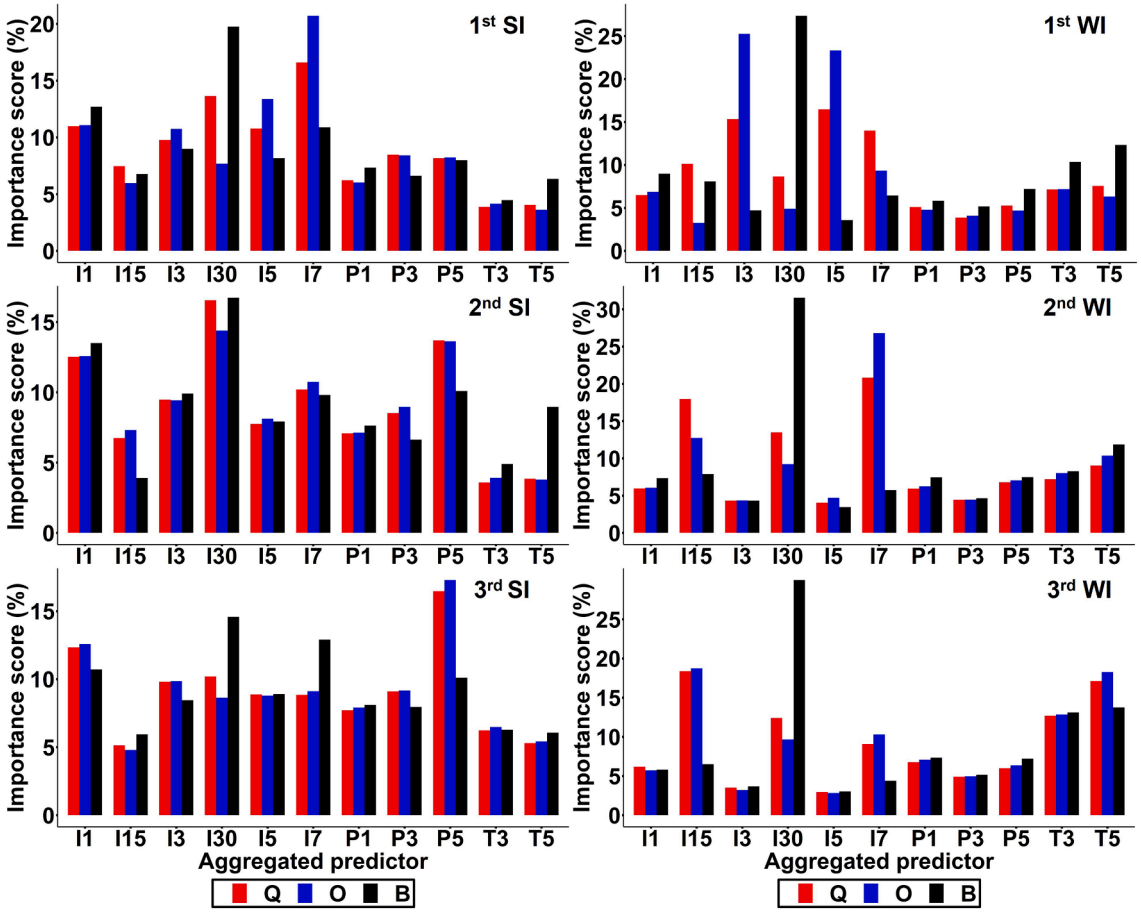

Fig. 10. Aggregated importance scores for the baseflow (B), O&I flows (O), and streamflow (Q). Note that ***P, I,*** and ***T*** denote precipitation, irrigation, and temperature, respectively; the numbers in aggregated predictor represent the length of moving windows.

various moving windows as surrogates to the basin antecedent conditions. The length of these moving windows was designed in accordance with the basin characteristics. Nevertheless, it is challenging to fully characterize the basin conditions. Therefore, this is of particular importance to know that the robustness of the proposed BFSCE is expected to improve with additional sources of data such as soil moisture and snowmelt, as well as more data samples.

The baseflow simulation module embedded in BFSCE uses constant coefficients, which may face difficulties addressing uncertainties in complex hydrological systems (Balogun et al., 2020; Yang et al., 2020). For the proposed BFSCE model, empirical equations proposed by Meshgi et al. (2014) were used to estimate the baseflow. However, previous studies suggested that the coefficients in the empirical equations, optimized globally to work across many watersheds, might not be optimal for any specific watershed. Future studies should take into account the uncertainties of process-based modules such that more physically reliable results can be expected.

5. Conclusions

This study proposed a BFSCE for improving the capacity of hydrological inference with the aid of a process-based baseflow separation method. It is the first attempt to incorporate a feature importance method and a process-based baseflow separation method for quantifying the relative contributions of associated model predictors to sub-hydrological processes (i.e., baseflow and O&I flows). The results revealed that the predictors associated with air temperature and long-term (i.e., monthly) irrigation mainly characterized baseflow dynamics, while precipitation and short-term (i.e., weekly or semi-weekly) irrigation mainly captured overland flow and interflow dynamics. Such information could help determine whether the watershed is precipitation or irrigation dominated. Such a better understanding could further facilitate the conjunctive operation of surface water and groundwater resources, thereby reducing the risks of flooding and salinization.

The fidelity of hydrological inference through the feature importance method was further demonstrated through MFA. The results revealed that the relative importance of predictors reflected their significance on model performance and their influence on explaining the changes of streamflow. In general, the higher importance of a predictor, the more changes on streamflow can be explained by this predictor. The proposed BFSCE model is a step closer to understanding sub-hydrological processes through data-driven models. As physics-informed data-driven models continue to develop in the future, the improved hydrological inference through data-driven modelling frameworks will draw increasing attention from the hydrological community.

CRediT authorship contribution statement

Kailong Li: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Guohe Huang:** Supervision. **Shuo Wang:** Writing – review & editing. **Saman Razavi:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

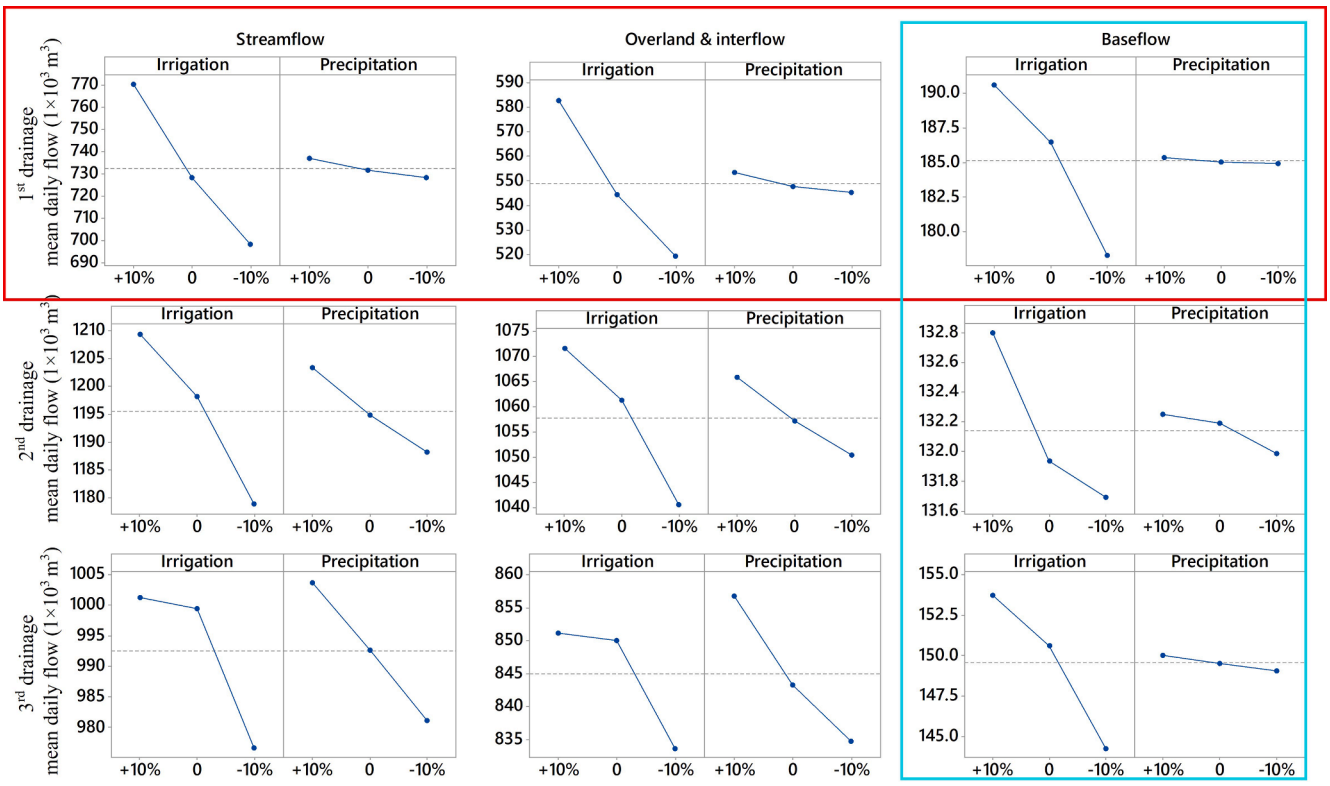

Fig. 11. Main effect of individual factors. Note that each row represents a drainage basin.

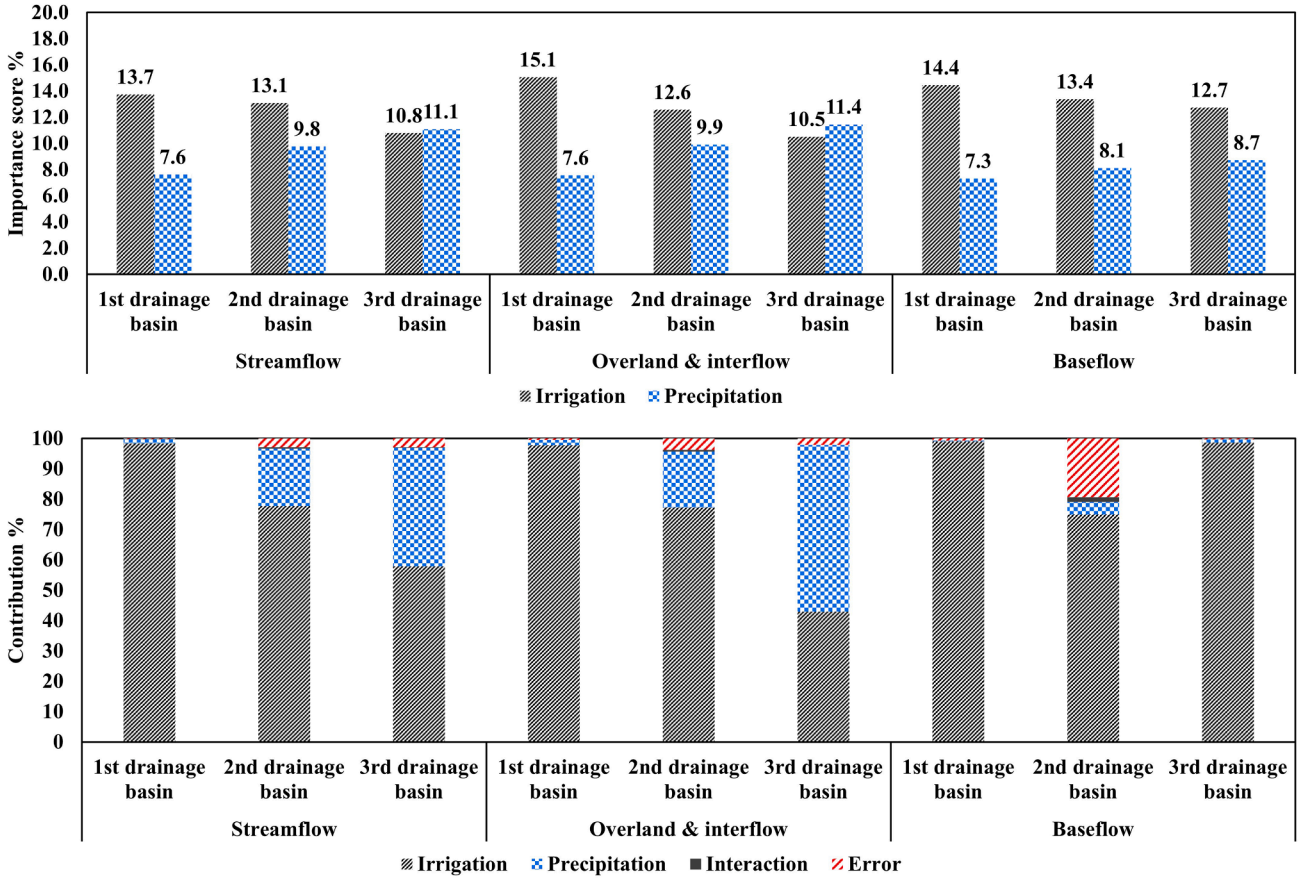

Fig. 12. Relative contributions obtained from WFI (top) and MFA (bottom).

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by Canada Research Chair Program, Natural Science and Engineering Research Council of Canada, and MITACS. We appreciate Ningxia Water Conservancy for offering the streamflow, groundwater and irrigation data, and related help. We are also very grateful for the helpful inputs from the Editor and anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2022.128323.

References

Armenio, E., De Serio, F., Mossa, M., 2017. Analysis of data characterizing tide and current fluxes in coastal basins. Hydrol. Earth Syst. Sci. 21 (7), 3441–3454.
ASCE Task Committee, 2000. Artificial neural networks in hydrology. II: Hydrologic applications. J. Hydrol. Eng. 5 (2), 124–137.
Babovic, V., Keijzer, M., 2002. Rainfall runoff modelling based on genetic programming. Hydrol. Res. 33 (5), 331–346.
Badrzadeh, H., Sarukkalige, R., Jayawardena, A., 2016. Improving ann-based short-term and long-term seasonal river flow forecasting with signal processing techniques. River Res. Appl. 32 (3), 245–256.
Baek, S.-S., Pyo, J., Chun, J.A., 2020. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. Water 12 (12), 3399.
Balogun, A., Quan, S., Pradhan, B., Dano, U., Yekeen, S., 2020. An improved flood susceptibility model for assessing the correlation of flood hazard and property prices using geospatial technology and fuzzy-ANP. J. Environ. Inform. 37 (2), 107–121.
Bénard, C., Biau, G., Veiga, S., Scornet, E., 2021. Interpretable random forests via rule extraction. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp. 937–945.
Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., Vrac, M., 2017. Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). Hydrol. Earth Syst. Sci. 21 (6), 2701–2723.
Bhasme, P., Vagadiya, J., Bhatia, U., 2021. Enhancing predictive skills in physically-consistent way: Physics Informed Machine Learning for Hydrological Processes. arXiv preprint arXiv:2104.11009.
Breiman, L., 2001. Random forests. Machine learning, 45(1): 5-32.
Chadalawada, J., Herath, H., Babovic, V., 2020. Hydrologically informed machine learning for rainfall-runoff modeling: a genetic programming-based toolkit for automatic model induction. Water Resour. Res. 56 (4).
Chen, L., Singh, V.P., Guo, S., Zhou, J., Ye, L., 2014. Copula entropy coupled with artificial neural network for rainfall–runoff simulation. Stoch. Env. Res. Risk Assess. 28 (7), 1755–1767.
Clement, F., Orange, D., Williams, M., Mulley, C., Epprecht, M., 2009. Drivers of afforestation in Northern Vietnam: assessing local variations using geographically weighted regression. Appl. Geogr. 29 (4), 561–576.
Corzo, G., Solomatine, D., 2007. Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. Hydrol. Sci. J. 52 (3), 491–507.
Criss, R.E., Winston, W.E., 2008. Do Nash values have value? Discussion and alternate proposals. Hydrol. Processes: Int. J. 22 (14), 2723–2725.
Das Bhowmik, R., Seo, S.B., Das, P., Sankarasubramanian, A., 2020. Synthesis of irrigation water use in the United States: spatiotemporal patterns. J. Water Resour. Plann. Manage. 146 (7), 04020050.
Daw, A., Karpatne, A., Watkins, W., Read, J., Kumar, V., 2017. Physics-guided neural networks (pgnn): An application in lake temperature modeling. arXiv preprint arXiv: 1710.11431.
Dewandel, B., Gandolfi, J.M., De Condappa, D., Ahmed, S., 2008. An efficient methodology for estimating irrigation return flow coefficients of irrigated crops at watershed and seasonal scale. Hydrol. Processes: Int. J. 22 (11), 1700–1712.
Ditthakit, P., Pinthong, S., Salaeh, N., Binnui, F., Khwanchum, L., Kuriqi, A., Khedher, K. M., Pham, Q.B., 2021. Performance evaluation of a two-parameters monthly rainfall-runoff model in the Southern Basin of Thailand. Water 13 (9), 1226.
Dong, L., Yu, D., Zhang, H., Zhang, M., Jin, W., Liu, Y., Shi, X., 2015. Long-term effect of sediment laden Yellow River irrigation water on soil organic carbon stocks in Ningxia, China. Soil Tillage Res. 145, 148–156.
Dralle, D.N., Karst, N.J., Thompson, S.E., 2016. Dry season streamflow persistence in seasonal climates. Water Resour. Res. 52 (1), 90–107.
Eckhardt, K., 2005. How to construct recursive digital filters for baseflow separation. Hydrol. Processes: Int. J. 19 (2), 507–515.
Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Machine Learn. Res. 15 (1), 3133–3181.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Galelli, S., Castelletti, A., 2013. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. Hydrol. Earth Syst. Sci. 17 (7), 2669.
Gao, Z., Gu, J., Xu, J., 2003. Hydrological and hydrogeological parameters of the western irrigation district of Qingtongxia River in Ningxia. Water Resour. Protect. (in Chinese)(2): 14-16.
Gilli, M., Maringer, D., Schumann, E., 2019. Numerical Methods and Optimization in Finance. Academic Press.
Gosain, A., Rao, S., Srinivasan, R., Reddy, N.G., 2005. Return-flow assessment for irrigation command in the Palleru River basin using SWAT model. Hydrol. Processes: Int. J. 19 (3), 673–682.
Gu, T.-F., Zhang, M.-S., Wang, J.-D., Wang, C.-X., Xu, Y.-J., Wang, X., 2019. The effect of irrigation on slope stability in the Heifangtai Platform, Gansu Province, China. Eng. Geol. 248, 346–356.
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. J. Hydrol. 377 (1–2), 80–91.
Herath, H.M.V.V., Chadalawada, J., Babovic, V., 2021. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. Hydrol. Earth Syst. Sci. 25 (8), 4373–4401.
Hersbach, H. et al., 2018. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 10.
Hsu, K.l., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res., 31(10): 2517-2530.
Huang, G., 1992. A stepwise cluster analysis method for predicting air quality in an urban environment. Atmos. Environ. Part B 26 (3), 349–357.
Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2018. Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. 31 (8), 1544–1554.
Kendall, M.G., 1948. Rank Correlation Methods, 4th ed. Charles Griffin and Company, London, p. 1948.
Khandelwal, A. et al., 2020. Physics guided machine learning methods for hydrology. arXiv preprint arXiv:2012.02854.
Konapala, G., Mishra, A., 2020. Quantifying climate and catchment control on hydrological drought in the continental United States. Water Resour. Res. 56 (1).
Kong, X., et al., 2021. Quantification of surface water and groundwater salinity sources in irrigated lowland area of North China Plain. Hydrol. Process. 35 (4), e14037.
Kratzert, F., et al., 2019a. Benchmarking a catchment-aware long short-term memory network (LSTM) for large-scale hydrological modeling. Hydrol. Earth Syst. Sci. Discuss. 1–32.
Kratzert, F., et al., 2019b. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23 (12).
Kumar, R., et al., 2022. Assessment of climate change impact on snowmelt runoff in himalayan region. Sustainability 14 (3), 1150.
Landerer, F.W., Swenson, S., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48 (4).
Lee, K.-S., Park, Y., Shin, W.-J., 2021. Hydrograph separation for a small agricultural watershed: The role of irrigation return flow. J. Hydrol. 593, 125831.
Li, K., Huang, G., Wang, S., 2019. Market-based stochastic optimization of water resources systems for improving drought resilience and economic efficiency in arid regions. J. Cleaner Prod. 233, 522–537.
Li, K., Huang, G., Baetz, B., 2021a. Development of a Wilks feature importance method with improved variable rankings for supporting hydrological inference and modelling. Hydrol. Earth Syst. Sci. 25, 4947–4966.
Li, K., Huang, G., Zhang, X., Lu, C., Wang, S., 2021b. Temporal-spatial changes of monthly vegetation growth and their driving forces in the ancient yellow river irrigation system, China. J. Contaminant Hydrol. 243, 103911.
Li, K., Huang, G., Wang, S., Razavi, S., Zhang, X., 2022. Development of a joint probabilistic rainfall-runoff model for high-to-extreme flow projections under changing climatic conditions. Water Resour. Res. 58 (6).
Li, Kailong, Huang, Guohe, Wang, Shuo, Baetz, Brian, Xu, Weihuang, 2022a. A Stepwise Clustered Hydrological Model for Addressing the Temporal Autocorrelation of Daily Streamflows in Irrigated Watersheds. Water Resources Research 58 (2), e2021WR031065. https://doi.org/10.1029/2021WR031065. In this issue.
Liang, J., Li, W., Bradford, S.A., Šimůnek, J., 2019. Physics-informed data-driven models to predict surface runoff water quantity and quality in agricultural fields. Water 11 (2), 200.
Liu, W., et al., 2020. Quantifying the streamflow response to groundwater abstractions for irrigation or drinking water at catchment scale using SWAT and SWAT–MODFLOW. Environ. Sci. Eur. 32 (1), 1–25.
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. J. Hydrometeorol. 22 (6), 1421–1438.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 4765–4774.
Lv, Y., Gao, L., Geris, J., Verrot, L., Peng, X., 2018. Assessment of water sources and their contributions to streamflow by end-member mixing analysis in a subtropical mixed agricultural catchment. Agric. Water Manag. 203, 411–422.
Mann, H.B., 1945. Nonparametric tests against trend. Econometr.: J. Econom. Soc. 245–259.
Meshgi, A., Schmitter, P., Babovic, V., Chui, T.F.M., 2014. An empirical method for approximating stream baseflow time series using groundwater table fluctuations. J. Hydrol. 519, 1031–1041.

Meshgi, A., Schmitter, P., Chui, T.F.M., Babovic, V., 2015. Development of a modular streamflow model to quantify runoff contributions from different land uses in tropical urban environments using genetic programming. J. Hydrol. 525, 711–723.
Mi, L., et al., 2020. Evolution of groundwater in Yinchuan Oasis at the upper reaches of the yellow river after water-saving transformation and its driving factors. Int. J. Environ. Res. Public Health 17 (4), 1304.
Młyński, D., Wałęga, A., Kuriqi, A., 2021. Influence of meteorological drought on environmental flows in mountain catchments. Ecol. Ind. 133, 108460.
Molnar, C., 2020. Interpretable Machine Learning. Lulu Press, Morrisville, North Carolina, United States, 2020.
Montgomery, D.C., 2017. Design and analysis of experiments. John Wiley & Sons.
Mottaleb, K.A., Krupnik, T.J., Keil, A., Erenstein, O., 2019. Understanding clients, providers and the institutional dimensions of irrigation services in developing countries: a study of water markets in Bangladesh. Agric. Water Manag. 222, 242–253.
Mulvany, T.J., 1850. On the use of self-registering rain and flood gauges. Making Observations of the Relations of Rain Fall and Flood Discharges in a Given Catchment. Transactions and Minutes of the Proceedings of the Institute of Civil Engineers of Ireland.
Myneni, R., Knyazikhin, Y., Park, T., 2015. MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. , Accessed 2021-10-26 from https://doi.org/10.5067/MODIS/MCD15A3H.006.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10 (3), 282–290.
Nath, R., Pavur, R., 1985. A new statistic in the one-way multivariate analysis of variance. Comput. Stat. Data Anal. 2 (4), 297–315.
Nearing, G.S., et al., 2021. What role does hydrological science play in the age of machine learning? Water Resour. Res. 57 (3).
Newman, A., et al., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrol. Earth Syst. Sci. 19 (1), 209.
O'Neill, P.E., S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell., 2019. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 6 [Data Set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/EVYDQ32FNWTH.
Ningxia Water Conservancy, 2015. Ningxia Water Resources Bulletin. Ningxia Water Conservancy, Ningxia, China.
Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. Prog. Phys. Geogr. 35 (4), 465–492.
Qing, Y., Wang, S., Ancell, B.C., Yang, Z.-L., 2022. Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. Nat. Commun. 13 (1), 1–10.
Rahman, M., Arya, D., Goel, N., 2010. Limitation of 90 m SRTM DEM in drainage network delineation using D8 method—a case study in flat terrain of Bangladesh. Appl. Geomatics 2 (2), 49–58.
Rajbahadur, G.K., Wang, S., Ansaldi, G., Kamei, Y., Hassan, A.E., 2021. The impact of feature importance methods on the interpretation of defect classifiers. IEEE Trans. Software Eng.
Ramireddygari, S., Sophocleous, M., Koelliker, J., Perkins, S., Govindaraju, R., 2000. Development and application of a comprehensive simulation model to evaluate impacts of watershed structures and irrigation water use on streamflow and groundwater: the case of Wet Walnut Creek Watershed, Kansas, USA. J. Hydrol. 236 (3–4), 223–246.
Razavi, S., 2021. Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. Environ. Modell. Software 144, 105159.
Razavi, S., et al., 2021. The future of sensitivity analysis: an essential discipline for systems modeling and policy support. Environ. Modell. Software 137, 104954.
Razavi, S., et al., 2022. Coevolution of machine learning and process-based modelling to revolutionize Earth and environmental sciences: A perspective. Hydrol. Process. 36 (6), e14596.
Razavi, S., Gober, P., Maier, H.R., Brouwer, R., Wheater, H., 2020. Anthropocene flooding: Challenges for science and society. Hydrol. Process. 34 (8), 1996–2000.
Schmidt, L., Heße, F., Attinger, S., Kumar, R., 2020. Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. Water Resour. Res. 56 (5).
Schnier, S., Cai, X., 2014. Prediction of regional streamflow frequency using model tree ensembles. J. Hydrol. 517, 298–309.
Singh, A.P., Medida, S., Duraisamy, K., 2017. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. AIAA J. 55 (7), 2215–2227.
Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinf. 10 (1), 3–22.
Song, Tangnyu, Huang, Guohe, Wang, Xiuquan, et al., 2022. Neglected Spatiotemporal Variations of Model Biases in Ensemble-Based Climate Projections. Geophysical Research Letters 49 (16), e2022GL098063. In this issue.
Tongal, H., Booij, M.J., 2018. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. J. Hydrol. 564, 266–282.
Traylor, J.P., Zlotnik, V.A., 2016. Analytical modeling of irrigation and land use effects on streamflow in semi-arid conditions. J. Hydrol. 533, 591–602.
Vishwakarma, D.K., et al., 2022. Methods to estimate evapotranspiration in humid and subtropical climate conditions. Agric. Water Manag. 261, 107378.
Waleeittikul, A., Chotpantarat, S., Ong, S.K., 2019. Impacts of salinity level and flood irrigation on Cd mobility through a Cd-contaminated soil, Thailand: experimental and modeling techniques. J. Soils Sediments 19 (5), 2357–2373.
Ward, A.D., Trimble, S.W., 2003. Environmental Hydrology. CRC Press.
Wilks, S.S., 1967. Collected Papers; Contributions to Mathematical Statistics. Wiley.
Woodhouse, C.A., Pederson, G.T., Morino, K., McAfee, S.A., McCabe, G.J., 2016. Increasing influence of air temperature on upper Colorado River streamflow. Geophys. Res. Lett. 43 (5), 2174–2181.
Wu, C., Chau, K., Li, Y., 2009. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. Water Resour. Res. 45 (8).
Yang, G., Li, M., Guo, P., 2020. Monte Carlo-Based Agricultural Water Management under Uncertainty: A Case Study of Shijin Irrigation District, China. J. Environ. Inform.
Yang, Y., Huang, T.T., Shi, Y.Z., Wendroth, O., Liu, B.Y., et al., 2021. Comparing the Performance of an Autoregressive State-Space Approach to the Linear Regression and Artificial Neural Network for Streamflow Estimation. Journal of Environmental Informatics 37 (1), 36–48. In this issue.
You, J., Wang, S., 2021. Higher probability of occurrence of hotter and shorter heat waves followed by heavy rainfall. Geophys. Res. Lett. 48 (17).
Zeng, R., Cai, X., 2014. Analyzing streamflow changes: irrigation-enhanced interaction between aquifer and streamflow in the Republican River basin. Hydrol. Earth Syst. Sci. 18 (2), 493–502.
Zhang, H., Yang, Q., Shao, J., Wang, G., 2019. Dynamic streamflow simulation via online gradient-boosted regression tree. J. Hydrol. Eng. 24 (10), 04019041.