Environmental Modelling and Software 112 (2019) 157–165

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Dynamic runoff simulation in a changing environment: A data stream approach

Qinli Yang[a,b,c], Heng Zhang[a], Guoqing Wang[c], Shasha Luo[a], Dongzi Chen[a], Wanshan Peng[a], Junming Shao[d,b,*]

[a] School of Resources and Environment, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, Chengdu, 611731, China
[b] Big Data Research Center, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, Chengdu, 611731, China
[c] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, 210029, China
[d] School of Computer Science and Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, Chengdu, 611731, China

ARTICLE INFO

Keywords:
Runoff
Climate change
Land cover change
Data stream mining

ABSTRACT

In this study, we introduce a data stream method for dynamic runoff simulation, which allows capturing the evolving relationship between runoff and its impact factors (e.g., temperature, rainfall). The basic idea is to view continuously arriving data of runoff and its impact factors as a data stream, and dynamically learn its relationship. To validate the effectiveness of the proposed method, we compare its performance with that of three data driven models (ANN, SVR, Random Forest) and six representative hydrological models (SWAT, AWBM, SimHyd, SMAR, Sacramento, and Tank) in simulating monthly runoff. The proposed method performs well with the best NSE of 0.88, being superior to comparable models. Furthermore, the data stream model also shows its advantage in the flexibility of combing various impact factors of runoff into the model. The findings demonstrate that the data stream method provides a promising way to dynamically simulate runoff in a changing environment.

1. Introduction

1.1. Challenges of runoff response simulation in the changing environment

There have been extensive studies on runoff simulation, yet more accurate runoff simulation still remains a challenge in the context of a changing environment (Montanari et al., 2013). In recent decades, the formation and variation of runoff is suggested to have been considerably impacted by climate change and increasingly intensive human activities (Wang et al., 2016; Yang et al., 2017). The runoff series and the rainfall-runoff relationship have exhibited non-stationary features (Milly et al., 2005), which result in poor performances of many existing runoff simulation models. So, for non-stationary runoff sequences and non-stationary rainfall-runoff relationships, how to establish and apply effective models to simulate runoff response to environment changes has been one of the forefront and challenging topics in water resources research (Xu et al., 2013).

At present, the methods for runoff simulation can be broadly categorized into two groups: data-driven models and hydrologic models. A data-driven model simulates runoff by constructing a mathematical relationship between runoff and affecting factors based on data mining or machine learning techniques, like Artificial Neural Network (ANN) (Lee et al., 2008), genetic programming (Mehr and Nourani, 2017), neuro fuzzy inference system (Bartoletti et al., 2018), and Support Vector Regression (SVR) (Granata et al., 2016). A hydrologic model is a simplification of a real-world water system in a given basin. It quantitatively simulates runoff by establishing relationship between rainfall and runoff based on physical mechanisms (Wang et al., 2016). To date, with the development of hydrology and computer science, numerous hydrologic models ranging from catchment scale to global scale have been developed (Li et al., 2015). They include lumped models (e.g., AWBM (Boughton, 1995), SimHyd (Chiew et al., 2002)) along with semi-distributed models (e.g., SWAT (Arnold et al., 1998)) and distributed models (e.g., VIC (Liang et al., 1994), MIKE-SHE (Refsgaard et al., 2010)). In spite of different structures, mechanisms, scale and platforms of the hydrological models, a procedure for runoff simulation can be summarized into two phases. Firstly, calibrate parameters in a given hydrologic model based on a certain period of historical data. Secondly, simulate runoff response over the validation period/future period by using the calibrated model.

* Corresponding author. School of Computer Science and Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, Chengdu 611731, China.
E-mail address: Junmshao@uestc.edu.cn (J. Shao).

https://doi.org/10.1016/j.envsoft.2018.11.007
Received 21 August 2018; Received in revised form 1 November 2018; Accepted 17 November 2018
Available online 19 November 2018
1364-8152/

Software availability

Name of software Data stream model for dynamic Runoff Simulation (DaRuS)
Developers Junming Shao, Qinli Yang
Contact address No. 2006, Xiyuan Avenue, Chengdu 611731, China
Email junmshao@uestc.edu.cn, qinli.yang@uestc.edu.cn
Year first available 2018
Required hardware and software DaRuS works in Java on Windows, Linux based computers
Cost Free
Software availability http://dm.uestc.edu.cn/wp-content/uploads/code/DaRus.zip

However, in reality, the relationship between climate, human activities and runoff may vary over time (i.e., become non-stationary) (Milly et al., 2005; Wang et al., 2016). This can be seen in the changes of dominant factors attributed to runoff variation. (Chang et al., 2015; Yin et al., 2017). For example, Chang et al. (2015) applied the VIC model to investigate the impacts of climate change and human activities on runoff change in Weihe River basin. They reported that climate change accounted for 36%, 28%, 53% and 10% of runoff change in the Weihe River in the 1970s, 1980s 1990s and 2000s, respectively. Thereby, in the non-stationary context, the calibrated model may no longer suit for the validation periods.

To adapt hydrological models to the instability of rainfall-runoff relationships, one of feasible strategies is to calibrate models at regular intervals. But, the work is time consuming and in particularly the interval length is difficult to decide. For instance, Merz et al. (2011) calibrated the HBV model for every 5 years based on historical data of 41 years in 273 basins in Austria. They found that both parameters of HBV and NSE varied obviously with time. Moreover, the selection of calibration periods is also important for hydrological model performance (Young, 2000; Gibbs et al., 2018). To the best of our knowledge, the structures and parameters of hydrologic models are rarely updated over constituent time periods (Xu et al., 2013; Ejigu et al., 2013).

More recently, how to simulate runoff more accurately under the changing environment conditions has attracted much attention (Marshall et al., 2006; Meng et al., 2016; Chang and Yeh, 2018; Gibbs et al., 2018; Qi and Liu, 2018). Data assimilation is a new emerging way to explore the dynamic properties of runoff simulation. The basic idea is to combine recent observation(s) with background estimates of a system variable (Evensen, 2006). For instance, Meng et al. (2016) proposed to trace parameter changes in hydrological models by using an ensemble Kalman filter technique with a constrained parameter evolution scheme. But the findings were based on synthetic experiments, not real practice. Pathiraja et al. (2016a) simultaneously estimated model parameters and states by using the Ensemble Kalman Filter and further improved runoff prediction for two pairs of experimental catchments in Western Australia (Pathiraja et al., 2016b). In the context of changing environment, dynamic runoff simulation by taking non-stationary rainfall-runoff relationship into account, still remains a challenge.

Beyond, most existing hydrological models are based on physical mechanisms and require a great deal of designated data. In many cases, this required data is hard to collect or unavailable. In addition, data indirectly related to runoff such as population, water withdrawals or Enhanced Vegetation Index (EVI, one of commonly used vegetation indices) are difficult to integrate into hydrological models. Therefore, current hydrological models are themselves limited in their practical extensions.

Currently, for non-stationary rainfall-runoff relationships in a changing environment, a new method is required to simulate runoff response to multiple driving factors in a dynamic manner.

1.2. Introduction to data stream and concept drift

Regarding the challenge of dynamic runoff simulation in a changing environment, data stream model provides a promising way and a new perspective. Data stream, different from traditional data, is characterized by its continuity, time sequence, huge volume and time evolution (Shao et al., 2017a). Data stream model aims at handling data streams especially focusing on dynamic evolution (Cabrera and Sànchez-Marrè, 2018). In this study, the concurrent time series of driving factors (e.g., climatic variables) and runoff can be regarded as a data stream. In addition, the change of relationship between runoff and driving factors can be viewed as abrupt or gradual changes (concept drift) in a data stream. From the new perspective, the question of runoff simulation is transferred to a data stream mining task.

During the past decades, hundreds of techniques have been proposed for data mining (Gaber et al., 2009; Shao et al., 2016, 2017b; Gomes et al., 2017) and have been widely used in water resources research (Yang et al., 2011, 2015; Yaseen et al., 2015; Tan et al., 2018) and environment science (Gibert et al., 2018). Here, only the relevant concept drift detection methods and data stream learning methods are presented.

Concept drift, is defined as the change of conditional distribution of the target variable given the input data (Widmer and Kubat, 1996). The objective of concept drift detection is to capture the changes of data patterns in a data stream. According to the speed of change, concept drift can be categorized as gradual concept drift or abrupt concept drift. To detect the gradual concept drift, the most used methods are the sliding window model and the decay function. However, the selection of window size has a decisive effect on the performance of the model. To identify the abrupt concept drift, two strategies are adopted: the distribution-based method (Kuncheva, 2008) and the error-rate based method (Ross et al., 2012). The distribution-based methods detect concept drifts by dynamically monitoring the change of distributions between two fixed or adaptive windows of data. For instance, ADWIN (Bifet and Gavaldà, 2007) maintains a time window of the data stream, separating the window into two sub-windows, and finally decides whether to shrink the window by comparing the difference of expected values of the two sub-windows. Due to the evolving nature of data streams, it is often difficult to determine the appropriate window size. Moreover, window-dependent approaches tend to detect and identify concept drifts in a relatively slow way. The error-rate based methods identify concept drifts by dynamically monitoring the performance of data stream prediction and deciding concept drift happens when the performance gets worse. One typical representative algorithm is DDM (Gama et al., 2004). However, since the prediction performance also heavily depends on the presence of noisy instances and the learning model itself, so the error-rate based method is not always a good option.

Regarding data stream learning, two main strategies have been proposed: model-based and instance-based strategies. A model-based strategy is learning from fixed or flexible recent data to update predictive models (Ikonomovska et al., 2011; Almeida et al., 2013). For example, Ikonomovska et al. (2011) proposed a decision tree-based data stream regression algorithm, which can incrementally update the decision tree to adapt to the concept drift in the data stream. Almeida et al. (2013) developed a rule-based regression algorithm named AMRules, which can adaptively update models according to the concept of drift detection, based on the Page-Hinkley test. The results are easier for interpretation. Alternatively, instance-based strategy mines data stream patterns by treating the sampling data as instances. Typically, Shaker and Huellermeier (2012) proposed a lazy learning-based algorithm (IBLStreams) for data stream regression and classification. Based on the spatial and temporal correlation and consistency of sampling data, IBLStreams deletes or adds samples so as to maintain data representing the current data pattern. Cabrera and Sànchez-Marrè (2018) proposed a

case-based stochastic learning approach for environmental data stream mining via a case-based reasoning system.

1.3. Aims and objectives

This study aims to develop a new method for dynamic runoff simulation based on data stream mining, which allows learning the non-stationary relationship between runoff and its impact factors. The objectives are as follows.

(1) simulate runoff response to climate and land cover changes in a dynamic manner;
(2) introduce a new flexible way to integrate various impact factors in runoff simulation;
(3) taking the Qingliu River catchment as a case study, to verify the superiority of the data stream model over several representative data driven models and hydrological models.

2. Study area and data acquisition

2.1. Study area

To demonstrate the procedure of our approach in runoff simulation, Qingliu River catchment (Fig. 1) is chosen as a representative case study. The reasons behind the catchment selection include (1) the authors are familiar with the catchment; (2) located in the rapidly developing south-east of China, the study area is experiencing hydro-climatic change and land use variation (Liu et al., 2010; Zhang and Pu, 2008); and (3) high quality (long term and full) data have been collected for the selected catchment.

The Qingliu River is a secondary-order tributary of the lower reaches of the Yangtze River. The Qingliu River catchment (32°13′-32°40′ N, 117°59′-118°25′ E, Fig. 1), covers a drainage area of 1070 km^2, receives an average annual precipitation of about 1000 mm and is subject to a mean annual temperature of about 16.0 °C. Seven rain gauges and one hydrological station are located in the catchment. According to the recorded hydrological data, the annual average runoff in the Qingliu River catchment is about 0.298 billion m^3. The land use in the catchment is dominated by farmland and forest. Only a small proportion of grassland has been identified.

2.2. Data acquisition

2.2.1. Hydro-climatic data

Daily precipitation, temperature, pan evaporation, wind speed, relative humidity, and solar radiation at seven rain gauges, recorded from 1989 to 2010, are acquired from the China Meteorological Administration. Monthly runoff data over the period of 1989–2010 and daily runoff data (with 20% of missing), at the Chuzhou station, are provided by Nanjing Hydraulic Research Institute. Based on the available data, the missing data of daily runoff are interpolated and supplemented, which can be further used for hydrological models.

Fig. 2 illustrates the annual runoff coefficient (runoff depth/rainfall) over the study period of 1989–2010, which also reflects the changes of the rainfall-runoff relationship.

2.2.2. Landsat data and EVI data

All available Level 1 Terrain (L1T) Landsat 5, 7 and 8 images of the study area with cloud cover less than 90% and corresponding EVI data files from 1989 to 2010 are acquired from USGS (United States Geological Survey). Briefly, the authors classify the land cover into five categories, namely farmland, forest, water body, residential area and others. Based on Landsat imagery, Continuous Change Detection and Classification of land cover (CCDC) algorithm (Zhu et al., 2014) is applied to yield the continuous classification of land use in the Qingliu River catchment. Building up on the continuous classification results, clear pixels (without cloud cover) covered by farm or forest are identified in each image, and EVI of the identified pixels are averaged to represent the vegetation coverage level of the study area. Since the time step of the averaged EVI data is 16 days, monthly EVI time series of the Qingliu River catchment is constructed by integration and interpolation. It worth noting that the EVI data not only represent vegetation cover level of the study area, but also partially reflects land cover change in the catchment. The basic statistical information of EVI can be identified.

3. Methodology

3.1. Existing hydrologic models for runoff simulation

3.1.1. SWAT model

The SWAT model is a well-established distributed hydrologic model for assessing water resource and nonpoint-source pollution problems for a wide range of scales and environmental conditions across the globe (Gassman et al., 2007). In this study, only the components relevant to runoff simulation in SWAT will be introduced. SWAT operates by dividing a given basin into multiple sub-basins and further delineating sub-basins into hydrologic response units (HRUs). The HRUs exhibit homogeneous combinations of land use, soil properties and slope. SWAT model requires a great deal of input data, mainly including meteorological data, topography, soils, and land use/land cover data. The meteorological variables commonly refer to precipitation and temperature. Depending on the potential evapotranspiration (PET) calculation method selected in SWAT, so wind speed, solar radiation

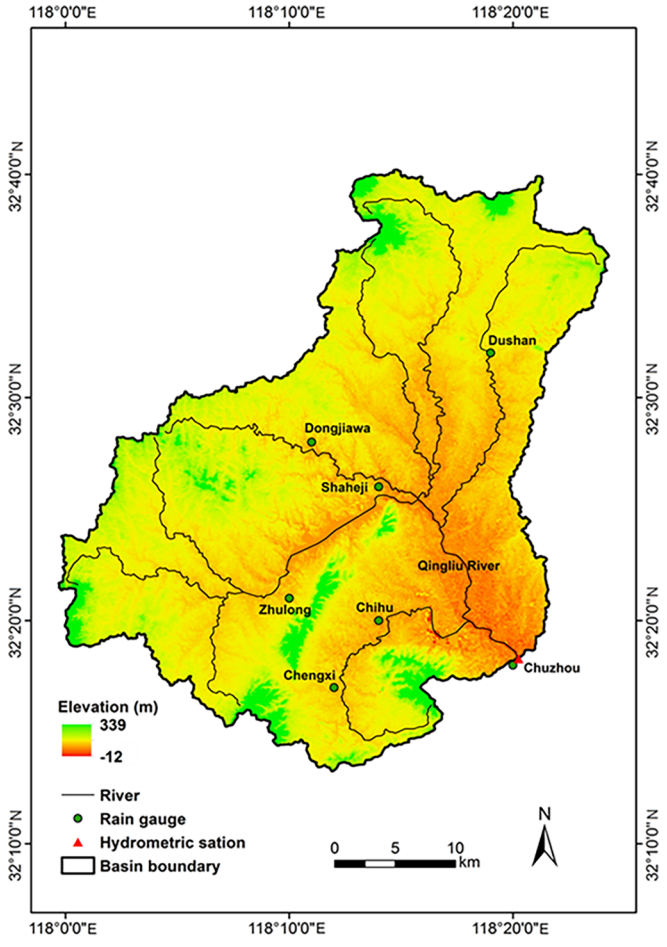

Fig. 1. River system and the locations of hydro-meteorological stations in the Qingliu River catchment, Anhui province, China.

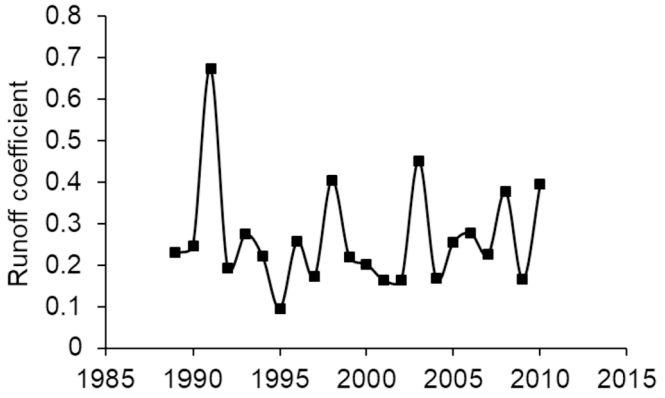

Fig. 2. Runoff coefficient over the period of 1989–2010.

and relative humidity may also be required. More details can be found in the SWAT theoretical documentation (http://swatmodel.tamu.edu).

3.1.2. Lumped hydrologic models

Five lumped conceptual rainfall-runoff models are selected, namely AWBM, Sacramento, SimHyd, SMAR, and Tank model. They are integrated in the Rainfall-Runoff Library (RRL, https://toolkit.ewater.org.au/Tools/RRL) and are suited for catchments from 10 km^2 to 10,000 km^2.

The AWBM model (Australian Water Balance Model) is a rainfall-runoff model based on the principle of water balance (Boughton, 1995). The model divides a basin into three parts with different surface water storage capacities. The structure of the model is relatively simple, and there are 8 parameters that need to be determined. AWBW requires evapotranspiration (ET) as an input.

In contrast, the Sacramento model (Finnerty et al., 1997; Gan and Burges, 2006) is relatively complex, using a total of 16 parameters to simulate the water balance. In the simulation of hydrological process, the Sacramento model divides the basin into 3 parts: permeable area, impervious area, and variable impervious area. The soil is vertically divided into two layers and five stores in the model. Runoff generated within the Sacramento model consists of three flow components: surface runoff, interflow, and base flow.

SimHyd model is a simplified lumped conceptual rainfall-runoff model on daily time step with 7 parameters (Chiew et al., 2002). It has been widely used in simulating runoff across Australia and worldwide (Chiew et al., 2009; Zhan et al., 2014). One advantage of the model is that it takes into account two types of flow generation mechanisms, namely saturation excess runoff and infiltration excess runoff. The model estimates river runoff generation from three contributions: infiltration runoff, interflow (and saturation excess runoff), and base flow.

The Soil Moisture Accounting and Routing (SMAR) model was proposed by Professor O'Connell and Nash of the National University of Ireland in the 1970s (Tan and O'Connor, 1996). After continuous improvement, it has been widely used around the world. The SMAR model has 9 parameters and consists of two parts: water balance and sink calculation. The mechanism of runoff generation in the model is based on saturation-excess.

The Tank model (Sugawara, 1974) treats the catchment as one or a few tanks. The advantage of the model is its simple principle, but the relatively large number of parameters is the major drawback of the model. In total, there are 18 parameters needing to be determined in the model.

3.2. Selected data-driven models

Support vector regression (SVR), proposed by Vapnik et al. (1997), is a regression technique based on Support Vector Machine (SVM). The performance of SVR depends on kernel functions, such as linear kernel, polynomial kernel, and radial basis kernel. The basic ideas underlying SVM for regression and function estimation can be found in Smola and Schölkopf (1998).

Random forest, proposed by Breiman (2002), is an ensemble learning based on bagging. It constructs many diverse regression trees via random sampling and determines output for predictions by majority voting.

Artificial neural network (ANN) (Lawrence, 1994) is a structure of interconnected units or nodes of large number of artificial neurons. It can learn about the nonlinear relationships between the inputs and outputs without a detailed understanding of its physical processes. ANN has been extensively used in runoff simulation and prediction in recent years (Nourani et al., 2014).

3.3. Dynamic runoff response simulation with a data stream method

In the following sections, we start with the detection of evolving relationship, and afterwards, introduce an instance-based data stream regression to model the dynamic runoff response to a changing environment. Fig. 3 presents the framework of the instance-based data stream model.

3.3.1. Evolving relationship detection and adaptation

In the context of changing environment, the relationship between runoff and its influencing factors may change smoothly or abruptly, which corresponds to gradual concept drift and sudden concept drift in data stream mining, respectively.

In handling gradual concept drift, instead of using recent data in a sliding fixed-size window to re-train the model, we employ an instance-based learning model, called IBLStreams (Shaker and Huellermeier, 2012), to optimize the composition and size of the relevant instances autonomously. In instance-based learning, instead of taking time to build a global model, the target function is approximated locally by means of selected instances. The inherent incremental nature of instance-based learning algorithms and their simple and flexible adaptation mechanism makes this type of algorithm suitable for learning in complex dynamic environments (Shao et al., 2014). Specifically, for a time step i, a new incoming example (X_i, Y_i), where $X_i = (R_i, T_i, E_i)$, $Y_i = Q_i$, and R_i, T_i, E_i, Q_i represent the rainfall, temperature, evaporation and runoff, respectively, is first added into a relevant base set. During the learning process, the relevant base set is dynamically updated, where some redundant data or outliers will be removed to make it better characterize the current relationship between runoff and its influencing factors. To this end, a set C of examples within a neighborhood of X_i are considered as candidates. Afterwards, the kc youngest examples in the neighborhood set C are used to determine a confidence interval as follows.

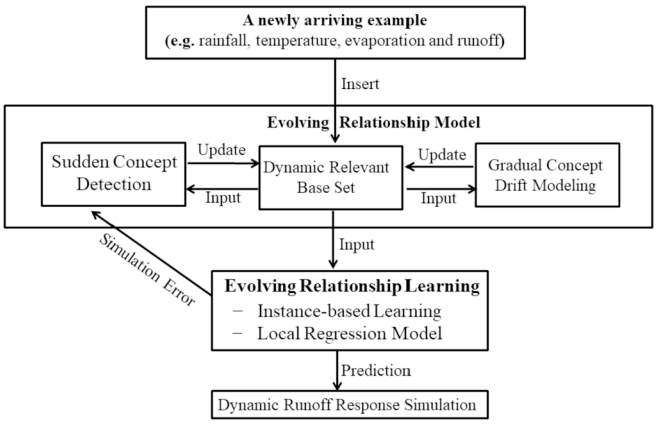

Fig. 3. Framework of the data stream method.

$$\mathrm{CI} = [\bar{y} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{k_c}},\ \bar{y} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{k_c}}] \tag{1}$$

where $\bar{y}$ is the average target value (i.e., Y_i) for the youngest kc examples and σ is the standard deviation; α is the significance level and α is set as 0.001 in this study. A candidate example in C is removed if it falls outside this confidence interval and is not one of the youngest kc instances.

For abrupt concept drift, old instances in the relevant base set should be removed instantaneously since the relationship between runoff and its affecting factors has changed significantly. To detect sudden concept drift, a statistical test (Gama et al., 2004) is applied. Specifically, we maintain the mean absolute error e and standard deviation s for the last 50 training instances. Let $e_{\min}$ denote the smallest among these errors and $s_{\min}$ be the associated standard deviation. A change is detected if the current value of e is significantly higher than e_{min} with standard Z-test. Once such an abrupt change is detected, the relative increase of error is used to determine the percentage of old examples to be removed.

3.3.2. Runoff simulation with evolving relationship learning

Building upon the evolving relationship detection and adaptation, the instance-based learning in the form of the nearest neighbor classifier is employed for runoff simulation. The most simple and typical way is to use the weighted mean of the neighbor's outputs as a prediction. Formally, given an input vector $X_i = (X_i^1, \cdots, X_i^m)$, where m is the number of impacting factors that will be considered in this study and i is the number of observations, its output Y_i can be estimated as follows.

$$Y_i^{est} = \sum_{X_j \in N_k(X_i)} w(X_j) Y_j \tag{2}$$

To gain a better prediction performance, here we assume that the relationship between runoff and its impacting factors can at least be approximated sufficiently well with a locally weighted linear regression function as follows.

$$Y_i = f(X_i^1, \cdots, X_i^m) = \beta_0 + \sum_{j=1}^{m} \beta_j X_i^j = \beta^T \begin{bmatrix} 1 \\ X_i \end{bmatrix} \tag{3}$$

Where X_i^j is the j-th dimension of an example X_i, and $\beta^T = \{\beta_0,\ \beta_1 \cdots, \beta_m\}$ is the corresponding coefficients of X_i, which can be estimated as

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \tag{4}$$

Here W is a diagonal weight matrix $diag(w_1, \cdots, w_k)$, where w_i is defined as follows.

$$w(X_j) = \frac{\frac{1}{d(X_j, X_i)}}{\sum_{X_j \in N_k(X_i)} \frac{1}{d(X_j, X_i)}} \tag{5}$$

where $d(X_j, X_i)$ is a metric function and Euclidean distance is used in this study.

Once $\hat{\beta}$ is computed with Eq. (4), the estimated runoff Y_i^{est} can be directly obtained based on Eq. (3). In case $X^T W X$ is singular and its inverse does not exist, the weighted average in Eq. (2) is used for prediction instead.

In addition, the performance of a nearest neighbor classifier is affected by the number of neighbors. With an initial value of K, K is then automatically updated by checking whether the parameter benefits by increasing or decreasing the current value by 1. To this end, the mean error in a window formed by the last 100 instances with $K - 1$ and $K + 1$ neighbors is considered. Whenever one of these two variants performs better in terms of the mean error, the current K is adapted correspondingly (Shaker and Hüllermeier, 2012).

3.4. Evaluation metrics

To evaluate the performance of different (environmenta) models, please refer to literature published by Bennett et al. (2013). In this study, the Nash–Sutcliffe model efficiency coefficient (NSE), mean absolute error (MAE), root mean square error (RMSE), Relative volume error (RE), Akaike Information Criterion (AIC) (Akaike, 1974), and Bayesian Information Criterion (BIC) (Schwarz, 1978) are selected as the evaluation criteria (equations (6)–(11)).

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N} (Y_i^{est} - Y_i)^2}{\sum_{i=1}^{N} (Y_i - \overline{Y_i})^2} \tag{6}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |Y_i^{est} - Y_i|}{N} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (Y_i^{est} - Y_i)^2}{N}} \tag{8}$$

$$\mathrm{RE} = \frac{\sum_{i=1}^{N} (Y_i^{est} - Y_i)}{\sum_{i=1}^{N} Y_i} \tag{9}$$

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{10}$$

$$\mathrm{BIC} = k\ln(N) - 2\ln(L) \tag{11}$$

where Y_i^{est} is the simulated runoff by the model at time i, Y_i represents the observed runoff at time i, $\overline{Y_i}$ indicates for the average value of observed runoff, N means the number of observations, k is the number of parameters of the model, and L is the maximum of the likelihood function.

4. Experiments

4.1. Synthetic data

For better illustration, a synthetic data stream is generated, which consists of both stationary part and non-stationary part. For non-stationary part, it further includes the gradual concept drift and sudden concept drift. Specifically, a linear function $y = w_1{*}x_1 + w_2{*}x_2 + w_3$ is considered. For illustration, x is only two-dimensional input data, y is the output. For stationary part, one thousand 2-dimensional points are randomly generated, and w is fixed as [0.5, 0.2, 2.5]. Namely: H1: $y = 0.5{*}x_1 + 0.2{*}x_2 + 2.5 + \mathrm{N}(0, 0.001)$, where N ($u$, sigma^2) is the Gaussian noise. To generate a data stream with gradual concept drift, w is gradually changed as follows. $w_1 = w_1 + 0.1{*}i/1000 + \mathrm{N}(0, 0.001)$, $w_2 = w_2 + 0.1{*}i/1000 + \mathrm{N}(0, 0.001)$, where $i = 1, \ldots, 1000$. In this way, the relationship between x and y is slowly changed over time. Finally, H1 is slowly changed to H2: $y = 0.6{*}x_1 + 0.3{*}x_2 + 2.5 + \mathrm{N}(0, 0.001)$. For sudden concept drift, w_1 and w_2 are changed ($w_1 = -0.5$, and $w_2 = 0.4$), making the relationship vary significantly (i.e., H3: $y = -0.5{*}x_1 + 0.4{*}x_2 + 2.5 + \mathrm{N}(0, 0.001)$). The corresponding synthetic data stream is plot in Fig. 4. Here 3000 examples are given, and the first 1000 example are in a stationary environment (H1), from the point of 1001–2000, the relationship gradually changed from H1 to H2; at the point of 2001, a sudden concept drift occurs, and the relationship is changed to H3.

From Fig. 4 (d), it can be observed that there is no abrupt change found for both stationary period and the period with gradual concept drift, and a sudden concept drift is detected at the 2008th point. For stationary period, it is relatively easy to model the relationship among y and x (x_1, x_2). For the period with gradual concept drift, according to the oldness of instances and the prediction performance, only important instances representing current concept are kept, making the data stream approach being able to model the evolving relationship effectively. For sudden concept drift, the proposed approach allows quickly detecting the change (with seven instances delay), and then the model is quickly

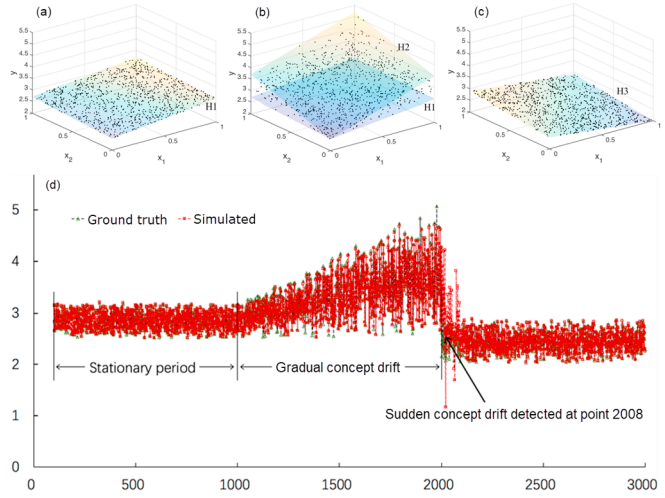

Fig. 4. Synthetic data stream and the simulation result.

updated to learn the new relationship and support high simulation performance later.

4.2. Real data

To verify the efficiency of the data stream model, we conduct experiments on real data (the Qingliu River catchment) to compare its performance with that of other data driven and hydrological models. The statistical information of the data set on monthly scale is illustrated in Table 1. Different models have different requirements for input data with respect of amount, temporal scale, spatial resolution, category, and format. Considering the data availability and to keep the consistency of time scale of all available input variables, the comparisons with the hydrologic models are on monthly scale. Finally, to make the comparison more distinct and straightforward, we provide Fig. 5 to illustrate the experimental setup.

The SWAT model requires a digital elevation model (DEM) with 30 m resolution from Geospatial Data Cloud of China, land use classification data (e.g., forest, farm land, water body, bare land and residential area) in 1989, derived from Landsat imagery. It also requires soil data from Harmonized World Soil Database (HWSD V1.1), daily meteorological data (rainfall, temperature, wind speed, relative humidity, and sun radiation) from the Chinese Meteorology Administration. Data is used to set up and run SWAT in ArcSWAT 2012. The FAO-Penman-Monteith method (Allen et al., 1998) is selected to calculate PET. Data from the period 1989–1999 is used for calibration (with one year for warm up), while data from 2000 to 2010 is selected for validation. The SUFI2 algorithm in SWAT-CUP is used for model calibration and NSE is taken as the objective function. The statistics of the input and output data over the two periods are presented in Table 1.

Five lumped models are implemented on the platform of the Rainfall Runoff Library (RRL). Model parameters are optimized by the Shuffled Complex Evolution (SCE-UA) algorithm against the objective function of NSE. Data segmentation used for calibration, warm up and validation is the same as that used in SWAT.

The input data for the data-driven models are monthly rainfall, temperature, pan evapotranspiration, and EVI. Data between January 1989 and December 1999 are used for training, and the rest of data (Jan. 2000–Dec. 2010) are used for testing. For the data stream method, the required input data and the training data are same with those in the data-driven models, while the rest of the monthly data enter the model one by one for runoff simulation in each month. The selected data driven models and the data stream model are implemented with Java on a PC with Windows 10 operating system.

To evaluate the runoff simulation under different conditions (e.g., climate change or climate and land cover changes), the data stream model is run with or without EVI data, respectively (as shown in Table 2). Considering that the important parameter K (i.e., the number of initial neighbors) in data stream model may affect the performance of the model, we run the model with different K values ranging from 30 to 60 with an interval of 5.

5. Results and discussion

5.1. Runoff simulation result with the data stream model

With different inputs data and K values in the data stream model, the corresponding runoff simulation results are summarized in Table 2. These suggest that the model shows a stable and good performance with high NSE and low errors. Specifically, NSE ranges from 0.85 to 0.88 or from 0.83 to 0.85 for the data stream model with or without EVI, respectively. The insensitivity of parameter K suggests the data stream model may suit many cases where only a limited length of data is available.

Noticeably, on average the results demonstrate that the data stream model with EVI achieved better runoff simulation results than that without EVI, with respect to all evaluation metrics. This finding implies EVI is an important contributor to runoff change and should be included during runoff simulation. This can be explained from two aspects: (1) EVI, representing for vegetation cover, may impact runoff via evapotranspiration and interception processes (Marques et al., 2007). (2) EVI variation, partially representing land use change (as mentioned in section 2.2.2), may influence runoff generation via changing soil storage capacity (Rogger et al., 2017). In addition, relative error (RE) of the data stream model with EVI is negative, while that without EVI is positive. This means that runoff in the former model is under estimated and runoff in the latter model is over-estimated.

Taking the best result of runoff simulation (i.e., NSE = 0.88, MAE = 4.03, RMSE = 6.7, RE = −0.04) as an example, where K = 30 and EVI is taken into consideration, Fig. 6 illustrates the modelled and observed monthly runoff at Chuzhou station during 2000 and 2010. It is intuitively noted that the data stream model simulates runoff response well, being able to capture the trend of runoff variation and evolution of the rainfall-runoff relationship.

5.2. Comparison with the existing models

To further test the data stream model, three data-driven models and six hydrological models are selected for comparison. Table 3 presents the best performances of different models for runoff simulation in terms

Table 1
Basic statistics of the data on monthly scale over the whole (1989–2010), training (1989–1999) and validation (2000–2010) periods, respectively, for the Qingliu River catchment.

Variable	Period	Unit	Min	Mean ± SD	Max
Temperature	Whole	℃	0.65	16.13 ± 8.83	30.58
	Training		0.65	15.75 ± 8.78	30.58
	Validation		1.23	16.51 ± 8.89	29.71
Precipitation	Whole	mm	0	83.93 ± 82.42	643.80
	Training		0	82.65 ± 80.63	495.00
	Validation		1.35	85.22 ± 84.45	643.80
Pan-Evaporation	Whole	mm	19.13	74.78 ± 34.55	157.94
	Training		19.59	74.45 ± 34.66	155.03
	Validation		19.13	75.12 ± 34.57	157.94
EVI	Whole	/	0.08	0.30 ± 0.14	0.57
	Training		0.08	0.28 ± 0.14	0.51
	Validation		0.10	0.31 ± 0.14	0.57
Runoff	Whole	m^3/s	0	9.73 ± 19.97	182.00
	Training		0	10.08 ± 20.74	176.00
	Validation		0.34	9.38 ± 19.25	182.00

Note: Sampling number: 264. EVI: enhanced vegetation index. No missing data.

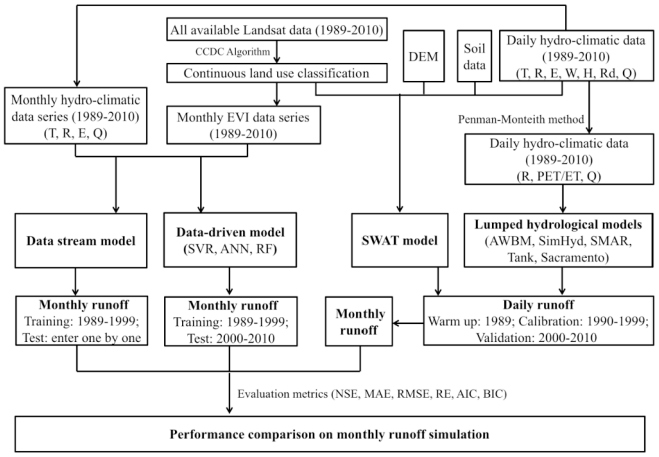

Fig. 5. Experimental setup for runoff response simulation comparison among different models.

Table 2
Performance of the data stream model with different values of parameter *K* for runoff simulation during 2000–2010 in the Qingliu River catchment, China.

	Without EVI				With EVI			
	NSE	MAE	RMSE	RE	NSE	MAE	RMSE	RE
K = 30	0.83	4.42	7.90	0.05	0.88	4.03	6.70	−0.04
K = 35	0.83	4.40	7.86	0.05	0.86	4.11	7.07	−0.04
K = 40	0.85	4.31	7.55	0.05	0.87	4.09	6.87	−0.04
K = 45	0.85	4.30	7.54	0.05	0.86	4.23	7.14	−0.04
K = 50	0.85	4.23	7.45	0.04	0.86	4.29	7.14	−0.04
K = 55	0.85	4.21	7.39	0.04	0.86	4.39	7.31	−0.05
K = 60	0.85	4.28	7.51	0.04	0.85	4.54	7.48	−0.05
Mean	0.84	4.31	7.60	0.05	0.86	4.24	7.10	−0.04

Note: EVI, enhanced vegetation index; NSE, Nash–Sutcliffe model efficiency coefficient; MAE, mean absolute error; RMSE, root mean square error; RE, Relative error.

of NSE, MAE, RMSE, RE, AIC, and BIC. In general, data stream model is superior compared to other comparable models with respect of all evaluation metrics except for MAE. Specifically, no matter whether EVI is included in the model, the data stream model gains the best results in term of NSE, RMSE, and RE compared with other comparable models. SWAT model also achieved relatively good results for runoff simulation, with NSE of 0.82 and with the smallest MAE value of 3.97. In contrast, the five lumped hydrological models and SVR produced relatively worse results with low NSE and/or high errors.

5.3. Advantages, limitations and future work

To present, hydrological models have been widely used in runoff simulation. However, more and more data-driven models have been proposed in recent years (Montanari et al., 2013; Gibert et al., 2018). To the best of our knowledge, both kinds of methods come with advantages and disadvantages. For hydrological models, since they are based on the physical mechanism, the pros are thus obvious: they are easy to understand and can be used to interpret hydrologic processes. However, the factors affecting runoff are diverse and the relationships between these factors are complicated. Hydrological models may not be easy to consider all affecting factors (direct or indirect) based on physical mechanisms. Thereby, the flexibility to incorporate new affecting factors may be limited and the performance of runoff simulation or prediction may suffer. Regarding data-driven model, its main desirable property is high simulation or prediction performance via mining all kinds of available data. However, for most existing data-driven models, the results are hard to interpret.

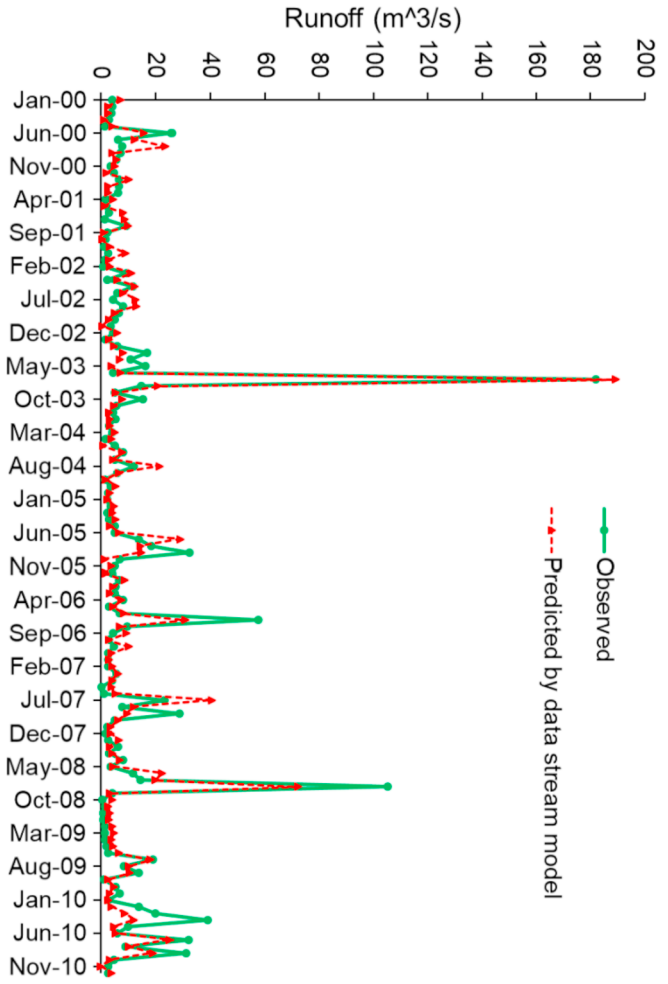

Fig. 6. Observed and simulated monthly runoff with data stream model at Chuzhou station during 2000 and 2010.

Traditional hydrological models are good at simulating the stationary relationship between runoff and its affecting factors (Xu et al., 2013), but fail to capture the dynamic relationship. The advantages of the data stream model are highlighted by its dynamics and flexibility. Regarding the dynamics, taking non-stationary rainfall-runoff relationships into account, the data stream model can dynamically simulate runoff response to environmental changes by detecting the relationship changes and updating the model. Although the data stream model is designed for non-stationary conditions, it also works for stationary scenarios. In terms of flexibility, when more impact factors related to runoff are included, namely land cover or population, they can be easily coupled into the data stream model once their data series are identified. As described in Equation (3), m represents the number of impact factors. The high flexibility allows for a wide range of applications of the data stream method. However, the data stream model lacks interpretation of physical mechanisms of driving changes in the hydrological process.

As a data-driven model, the data stream method provides a good supplementary tool for runoff simulation on monthly scale in a changing environment. Nevertheless, if the daily scaled data of runoff and its impact factors are available, the proposed approach can also be used for runoff simulation on daily scale via adjusting different K values. Therefore, the proposed data stream approach is useable for catchments in which long observational datasets exist.

Although the data stream can include more driving factors, in this study only EVI data is extended and tested. To better understand such interactions, quantitative attribution of factors leading to runoff change

Table 3
Performance comparison among different models for runoff simulation over the validation period (2000–2010).

Type of model	Model	NSE	MAE	RMSE	RE	AIC	BIC
Data stream model	With EVI	**0.88**	4.03	**6.70**	**−0.04**	**267.08**	**290.14**
	Without EVI	0.85	4.21	7.39	0.04	280.02	303.08
Data-driven model	SVR with EVI	0.44	5.15	14.34	−0.24	369.48	395.43
	SVR without EVI	0.44	5.14	14.38	−0.24	369.94	395.88
	ANN with EVI	0.48	5.62	13.84	0.30	364.80	390.75
	ANN without EVI	0.81	4.33	8.42	0.05	299.16	325.11
	RF with EVI	0.80	5.08	8.50	0.10	296.50	316.68
	RF without EVI	0.80	5.36	8.64	0.15	298.61	318.79
Semi-distributed hydrologic model	SWAT	0.82	**3.97**	8.12	0.07	312.45	364.34
Lumped hydrologic model	SimHyd	0.82	12.55	20.54	0.43	412.95	433.13
	Tank	0.75	13.64	20.05	0.44	431.77	483.66
	AWBM	0.77	11.33	24.32	0.26	437.25	460.31
	Sacramento	0.71	14.62	26.15	0.43	462.83	508.95
	SMAR	0.63	14.38	29.34	0.35	464.02	489.97

Note: NSE, Nash–Sutcliffe model efficiency coefficient; MAE, mean absolute error; RMSE, root mean square error; RE, relative error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; SVR, support vector regression; ANN, artificial neural network; RF, random forest.

needs to be determined. In the future, more impact factors such as population, water withdrawals (total and by sector), and reservoir capacities could be integrated into the model for more comprehensive study. In addition, the data stream model has the potential to be used for environment-related applications, such as water quality or air quality simulation.

6. Conclusions

In this study, a data stream method is introduced to simulate runoff response to environmental changes, where the data series of runoff and its impact factors (e.g., rainfall, temperature) are regarded as a data stream. Taking non-stationary rainfall-runoff relationships into account, the data stream method can dynamically simulate runoff response to historical climate and land cover changes by detecting relationship changes and updating the model. The Qingliu River catchment is used as a case study to verify the effectiveness of the data stream method. Model performance is compared with that of three data-driven models (SVR, ANN, Random Forest) and six internationally-used hydrological models (SWAT, AWBM, SimHyd, SMAR, Sacramento, and Tank). The results demonstrate that the data stream model achieves a stable and good performance with mean NSE over 0.84, being superior to all the comparable models. Additionally, the data stream model with EVI (NSE ranging from 0.85 to 0.88) outperforms that without EVI (NSE ranging from 0.83 to 0.85). The data stream approach provides a promising way for dynamic runoff simulation, in the context of a changing environment. Furthermore, the findings will be beneficial to local water resources management and planning.

Author contributions

Q. Y. and J. S. designed the research; G. W. provided hydrological data and made valuable suggestions and comments on the research design; Q.Y., and J. S. analysed the data; H. Z. processed the Landsat data; S. L. participated in SWAT model analysis; D. C. and W. P. did data pre-process; Q. Y. wrote the first manuscript draft; all authors read and commented on the manuscript.

Acknowledgments

This work has been financially supported by National Key Research and Development Program of China [grant number 2016YFA0601501, 2016YFB0502303], National Natural Science Foundation of China [grant numbers 41601025, 61403062, 41830863], State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering [grant number 2017490211], Science-Technology Foundation for Young Scientist of Sichuan Province [grant number 2016JQ0007], Sichuan Provincial Soft Science Research Program [grant number 2017ZR0208], and Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China [Grant number 161062]. We are grateful to Dr. Martin E. Parkes for constructive comments that improved the quality of the work. We are also grateful to the two editors and the three anonymous reviewers for their valuable advices on the earlier version of this manuscript.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19 (6), 716–723.
Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome 300 (9), D05109.
Almeida, E., Ferreira, C., Gama, J., 2013. Adaptive model rules from data streams. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, pp. 480–492.
Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. JAWRA Journal of the American Water Resources Association 34 (1), 73–89.
Bartoletti, N., Casagli, F., Marsili-Libelli, S., Nardi, A., Palandri, L., 2018. Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system. Environ. Model. Software 106, 35–47.
Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Pierce, S.A., 2013. Characterising performance of environmental models. Environ. Model. Software 40, 1–20.
Bifet, A., Gavalda, R., 2007. Learning from time-changing data with adaptive windowing. In: Proceedings of the 2007 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, pp. 443–448.
Boughton, W.C., 1995. An Australian water balance model for semiarid watersheds. J. Soil Water Conserv. 50 (5), 454–457.
Breiman, L., 2002. Manual on Setting up, Using, and Understanding Random Forests V3. 1, vol. 1 Statistics Department University of California Berkeley, CA, USA.
Cabrera, F.O., Sànchez-Marrè, M., 2018. Environmental data stream mining through a case-based stochastic learning approach. Environ. Model. Software 106, 22–34.
Chang, C.M., Yeh, H.D., 2018. Spectral analysis of temporal non-stationary rainfall-runoff processes. J. Hydrol. 559, 84–88.
Chang, J., Wang, Y., Istanbulluoglu, E., Bai, T., Huang, Q., Yang, D., Huang, S., 2015. Impact of climate change and human activities on runoff in the Weihe River Basin, China. Quat. Int. 380, 169–179.
Chiew, F.H.S., Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. Mathematical models of small watershed hydrology and applications 335–367.
Ejigu Eregno, F., Xu, C.Y., Kitterød, N.O., 2013. Modeling hydrological impacts of climate change in different climatic zones. International Journal of Climate Change Strategies and Management 5 (3), 344–365.
Evensen, G., 2006. Data Assimilation: the Ensemble Kalman Filter. Springer-Verlag, New York Inc., Secaucus, NJ.
Finnerty, B.D., Smith, M.B., Seo, D.J., Koren, V., Moglen, G.E., 1997. Space-time scale sensitivity of the Sacramento model to radar-gage precipitation inputs. J. Hydrol. 203 (1–4), 21–38.
Gaber, M.M., Zaslavsky, A., Krishnaswamy, S., 2009. Data stream mining. In: Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, pp. 759–787.
Gama, J., Medas, P., Castillo, G., Rodrigues, P., 2004. September). Learning with drift

detection. In: Brazilian Symposium on Artificial Intelligence. Springer, Berlin, Heidelberg, pp. 286–295.
Gan, T.Y., Burges, S.J., 2006. Assessment of soil-based and calibrated parameters of the Sacramento model and parameter transferability. J. Hydrol. 320 (1–2), 117–131.
Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Transactions of the ASABE 50 (4), 1211–1250.
Gibbs, M.S., McInerney, D., Greer, H., Thyer, M.A., Maier, H.R., Dandy, G.C., Kavetski, D., 2018. State updating and calibration period selection to improve dynamic monthly streamflow forecasts for an environmental flow management application. Hydrol. Earth Syst. Sci. 22 (1), 871.
Gibert K., Izquierdo, J., Sànchez-Marrè, M., Hamilton, S.H., Rodríguez-Roda, I., Holmes, G. (2018). Which method to use? An assessment of data mining methods in Environmental Data Science. Environ. Model. Software, (in press). https://doi.org/10.1016/j.envsoft.2018.09.021.
Gomes, H.M., Barddal, J.P., Enembreck, F., Bifet, A., 2017. A survey on ensemble learning for data stream classification. ACM Comput. Surv. 50 (2), 23.
Granata, F., Gargano, R., de Marinis, G., 2016. Support vector regression for rainfall-runoff modeling in urban drainage: a comparison with the EPA's storm water management model. Water 8 (3), 69.
Ikonomovska, E., Gama, J., Džeroski, S., 2011. Learning model trees from evolving data streams. Data Min. Knowl. Discov. 23 (1), 128–168.
Kuncheva, L.I., 2008, July. Classifier ensembles for detecting concept change in streaming data: overview and perspectives. In: 2nd Workshop SUEMA, vol. 2008. pp. 5–10.
Lawrence, J., 1994. Introduction to Neural Networks: Design, Theory, and Applications. California Scientific Software, Nevada City, CA, pp. 423.
Lee, K.T., Hung, W.C., Meng, C.C., 2008. Deterministic insight into ANN model performance for storm runoff simulation. Water Resour. Manag. 22 (1), 67–82.
Li, H., Zhang, Y., Zhou, X., 2015. Predicting surface runoff from catchment to large region. Adv. Meteorol., 720967 13 pages.
Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res.: Atmosphere 99 (D7), 14415–14428.
Liu, J., Zhang, Z., Xu, X., Kuang, W., Zhou, W., Zhang, S., Jiang, N., 2010. Spatial patterns and driving forces of land use change in China during the early 21st century. J. Geogr. Sci. 20 (4), 483–494.
Marques, M.J., Bienes, R., Jiménez, L., Pérez-Rodríguez, R., 2007. Effect of vegetal cover on runoff and soil erosion under light intensity events. Rainfall simulation over USLE plots. Sci. Total Environ. 378 (1–2), 161–165.
Marshall, L., Sharma, A., Nott, D., 2006. Modeling the catchment via mixtures: issues of model specification and validation. Water Resour. Res. 42 (11).
Mehr, A.D., Nourani, V., 2017. A Pareto-optimal moving average-multigene genetic programming model for rainfall-runoff modelling. Environ. Model. Software 92, 239–251.
Meng, S., Xie, X., Yu, X., 2016. Tracing temporal changes of model parameters in rainfall-runoff modeling via a real-time data assimilation. Water 8 (1), 19.
Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: implications for climate impact analyses. Water Resour. Res. 47 (2).
Milly, P.C., Dunne, K.A., Vecchia, A.V., 2005. Global pattern of trends in streamflow and water availability in a changing climate. Nature 438 (7066), 347.
Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., et al., 2013. "Panta Rhei—everything flows": change in hydrology and society—the IAHS scientific decade 2013–2022. Hydrol. Sci. J. 58 (6), 1256–1275.
Nourani, V., Baghanam, A.H., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J. Hydrol. 514, 358–377.
Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016a. Hydrologic modeling in dynamic catchments: a data assimilation approach. Water Resour. Res. 52 (5), 3350–3372.
Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016b. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. Adv. Water Resour. 94, 103–119.
Qi, W., Liu, J., 2018. A non-stationary cost-benefit based bivariate extreme flood estimation approach. J. Hydrol. 557, 589–599.
Refsgaard, J.C., Storm, B., Clausen, T., 2010. Système Hydrologique Europeén (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modelling. Nord. Hydrol 41 (5), 355–377.
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.
Ross, G.J., Adams, N.M., Tasoulis, D.K., Hand, D.J., 2012. Exponentially weighted moving average charts for detecting concept drift. Pattern Recogn. Lett. 33 (2), 191–198.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Shaker, A., Hüllermeier, E., 2012. IBLStreams: a system for instance-based classification and regression on data streams. Evolving Systems 3 (4), 235–249.
Shao, J., Ahmadi, Z., Kramer, S., 2014. Prototype-based learning on concept-drifting data streams. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 412–421.
Shao, J., Huang, F., Yang, Q., Luo, G., 2017a. Robust prototype-based learning on data streams. IEEE Trans. Knowl. Data Eng.
Shao, J., Wang, X., Yang, Q., Plant, C., Böhm, C., 2017b. Synchronization-based scalable subspace clustering of high-dimensional data. Knowl. Inf. Syst. 52 (1), 83–111.
Shao, J., Yang, Q., Dang, H.V., Schmidt, B., Kramer, S., 2016. Scalable clustering by iterative partitioning and point attractor representation. ACM Trans. Knowl. Discov. Data 11 (1), 5.
Smola, A.J., Schölkopf, B., 1998. A Tutorial on Support Vector Regression. NeuroCOLT, Technical Report NC-TR-98-030. Royal Holloway College, University of London, UK.
Sugawara, M., 1974. Tank Model with Snow Component. National Research Center for Disaster Prevention: Kyoto, Japan.
Tan, B.Q., O'Connor, K.M., 1996. Application of an empirical infiltration equation in the SMAR conceptual model. J. Hydrol. 185 (1–4), 275–295.
Tan, Q.F., Lei, X.H., Wang, X., Wang, H., Wen, X., Ji, Y., Kang, A.Q., 2018. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2018.01.015.
Vapnik, V., Golowich, S., & Smola, A. (1997). "Support vector method for function approximation, regression estimation, and signal processing", in M. Mozer, M. Jordan, and T. Petsche (eds.), Neural Information Processing Systems, vol. 9. MIT Press, Cambridge, MA.
Wang, G., Zhang, J., Yang, Q., 2016. Attribution of runoff change for the xinshui river catchment on the Loess Plateau of China in a changing environment. Water 8 (6), 267.
Widmer, G., Kubat, M., 1996. Learning in the presence of concept drift and hidden contexts. Mach. Learn. 23 (1), 69–101.
Xu, C., Chen, H., Guo, S., 2013. Hydrological modeling in a changing environment: issues and challenges. J. Water Resour. Res. 2, 85–95.
Yang, Q., Boehm, C., Scholz, M., Plant, C., Shao, J., 2015. Predicting multiple functions of sustainable flood retention basins under uncertainty via multi-instance multi-label learning. Water 7 (4), 1359–1377.
Yang, Q., Scholz, M., Shao, J., Wang, G., Liu, X., 2017. A generic framework to analyse the spatiotemporal variations of water quality data on a catchment scale. Environ. Model. Software. https://doi.org/10.1016/j.envsoft.2017.11.003.
Yang, Q., Shao, J., Scholz, M., Plant, C., 2011. Feature selection methods for characterizing and classifying adaptive sustainable flood retention basins. Water Res. 45 (3), 993–1004.
Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. J. Hydrol. 530, 829–844.
Yin, J., He, F., Xiong, Y.J., Qiu, G.Y., 2017. Effects of land use/land cover and climate changes on surface runoff in a semi-humid and semi-arid transition zone in northwest China. Hydrol. Earth Syst. Sci. 21 (1), 183.
Young, P.C., 2000. Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. Nonlinear and nonstationary signal processing 74–114.
Zhan, C., Zeng, S., Jiang, S., Wang, H., Ye, W., 2014. An integrated approach for partitioning the effect of climate change and human activities on surface runoff. Water Resour. Manag. 28 (11), 3843–3858.
Zhang, J., & Pu, L. J. (2008). On coordination between urbanization and farmland area of Chuzhou City in recent 30 years. Soils, 40, 523-528. (In mandarin).
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.