Catchment natural driving factors and prediction of baseflow index for Continental United States based on Random forest technique

Shanshan Huang (690240354@qq.com)
Wuhan University https://orcid.org/0000-0002-4865-6872
Qianjin Dong (dqjin@whu.edu.cn)
Wuhan University
Xu Zhang
Hong Kong University: University of Hong Kong
Weishan Deng
Wuhan University

Research Article

Keywords: baseflow, driving factors, Random Forest, long-term prediction

Posted Date: June 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-171758/v1

Version of Record: A version of this preprint was published at Stochastic Environmental Research and Risk Assessment on July 15th, 2021. See the published version at https://doi.org/10.1007/s00477-021-02057-2.

Abstract

Baseflow plays a critical role in maintaining the aquatic environmental health. However, the driving factors and predictions of baseflow have not been rigorously investigated on a large scale, partly preventing hydrologist from deeply understanding runoff generation. To this end, the Lyne–Hollick (LH) digital filter method and the automatic baseflow identification technique (ABIT) were used to estimate the long-term and seasonal baseflow index (BFI) of 619 catchments across Continental United States (CONUS) from 1981 to 2014. Six natural driving factors are selected from the 31 catchment attributes about topography and location, soil, geology, land cover, and climate characteristics. The Random Forest (RF) technique was used to predict the BFI with the selected six driving factors as predictors. Results show that the long-term average BFI was 0.49, and the BFI value was different in four seasons, with the highest value of 0.55 in winter and the lowest value of 0.46 in autumn. The forest fraction, clay proportion and snow fraction were the most powerful factors affecting the long-term average BFI. The RF technique predicts the BFI across the 619 sites in CONUS with a R^2 of 0.59 after Leave-One-Location cross-validation, which was more satisfactory than the multiple linear regression method. This study can provide a deep insight into the generation and variation of baseflow and guide the annual baseflow prediction for water resources management.

1 Introduction

Baseflow, including the groundwater, the slow soil runoff, and the water resource replenishment from lakes, reservoirs, and glaciers, greatly helps in determining the available water resources and environmental protection during dry periods (Hall 1968). Given that baseflow is released slowly when it flows through the soil matrix and maintains a relatively stable flow volume for a long time (Beck et al. 2013), it has important ecological functions, such as regulating the seasonal distribution of river flow (Gan et al. 2015), maintaining the aquatic habitat under extreme weather conditions (Fan et al. 2013), and transferring physical and chemical substances (Bosch et al. 2017). Sufficient river discharge must be ensured based on the efficient quantification or prediction of baseflow source (Goncalves et al. 2020). However, the spatial variation and generation of baseflow remain unclear due to the small number of gauging stations and the complex aquifer condition. Thus, the in-depth analysis of the driving factors of the baseflow is essential to understanding the hydrological cycle and sustainable water resources management (Araza et al. 2020).

BFI is the ratio of the mean baseflow to the mean streamflow in a certain region in a long-term period and can be calculated after the baseflow is separated from the daily streamflow. Many methods, which can be divided into tracer and non-tracer methods, have been developed for separating the baseflow to calculate the BFI values as a hydrological characteristic (Brunner et al. 2018; Georgek et al. 2018; Jones et al. 2006). Compared with tracer methods, which require high labor and operation costs, non-tracer methods have been widely used by hydrologists (Lott and Stewart 2016). Non-tracer methods include graphic methods (McNamara et al. 1997), digital simulation methods (Chapman 1991; Eckhardt 2005; Furey and Gupta 2001; Lyne and Hollick 1979), and physical chemistry methods (Cook et al. 2006). Many studies have discussed and compared these methods in terms of baseflow separation performance. For example, Xie et al. (2020) applied hydrologic model and digital filtering methods to separate the daily baseflow in CONUS and attempted to explore the separation performance. They found that the BFI estimated by Eckhardt digital filtering has the largest Nash Sutcliffe efficiency (NSE) median value among the 1145 study sites. Zhang et al. (2017a) evaluated the relative advantages of three types of non-tracer separation methods by using the tracing method as the criterion, and their results demonstrate that the use of automatic baseflow identification technology (ABIT) in estimating the recession constant produces smaller errors than the

The BFI can reflect the baseflow and streamflow status of the catchment, which is closely related to the catchment characteristics, such as the soil characteristics, the geological conditions, and the topographical factors (Henderson and Wooding 1964; Zhang et al. 2019b). According to previous studies, the BFI is often used as a hydrological factor to reflect the low flow conditions of catchments (Loveridge and Rahman 2014; Seo et al. 2018; Yang et al. 2018). Generally, regions with a high BFI have fewer extreme floods and more stable streamflow (Bastola et al. 2018). Many studies on the BFI have been conducted for water resource management and planning. For example, Sapač et al. (2020) and Zhang et al. (2017b) defined the BFI as the most common indicator in low flow studies and used the BFI as part of the hydrograph recession analysis to provide an intuitive overview of the low-flow environment; Yang et al. (2020) calculated the BFIs of the middle reaches of Yellow River to evaluate the ecological construction of the Loess Plateau in support of the environmental management; Schiling and Zhang (2004) and Zhu et al. (2019) relied on the BFI to estimate and control pollutants transmission.

Although the BFI in a certain site could be calculated easily using the aforementioned non-tracer methods, exploring the distribution and generation mechanism of baseflow and BFI in ungauged sites to support river flow management remains a challenge. Several studies have attempted to apply the knowledge on gauged catchments to the baseflow or BFI prediction in ungauged catchments (Zhang et al. 2016). For example, Molla and Tegaye (2019) established the stepwise regression equation between the BFI, the slope, and the drainage density, providing a base tool for exploring the baseflow in ungauged catchments. Zhang et al. (2020) conducted a large-scale comparison of the BFI prediction performance for 596 catchments in Australia, and their results show that the multilevel regression method with a NSE of 0.75 and a bias of 19% outperformed the linear regression methods and the hydrological models (i.e., SIMHYD and Xinanjiang model) in terms of prediction. Thus, Zhang et al. (2020) concluded that multilevel regression could improve BFI prediction and be applied to predict large-scale hydrological characteristics. Singh et al. (2019) applied the RF technique in estimating the baseflow of all river reaches across New Zealand based on some catchment attributes and then obtained overall satisfactory predictions of the long-term average BFI and seasonal BFIs spatial distribution.

Despite many efforts on baseflow separation and simulation, the driving factors of baseflow have not been rigorously and systematically identified in a large area. Accurately identifying the driving factors helps in understanding the baseflow generation and the prediction of BFI for ungauged stations. Therefore, this study has the following objectives: 1) to investigate the spatiotemporal pattern of baseflow in 619 catchments of CONUS, 2) to intensively analyze the driving factors of baseflow for a better understanding of the baseflow generation, and 3) to provide a reliable approach to the BFI prediction of ungauged sites using the RF technique to guide the development of water resources management.

2 Materials And Methods

2.1 Study area and data

A total of 619 catchments in CONUS were used in this study for analyzing the BFI. The streamflow, the precipitation, the potential evapotranspiration, and 31 catchment attributes were collected from the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) datasets (Addor et al. 2017; Newman et al. 2015). Before application, 619 gauged sites with a length of more than 30 years (1981–2014) were selected to ensure the reliability of the long-term average calculation, and the streamflow change trend during the study period was analyzed as shown in Fig. 1.

The average daily streamflow datasets can be obtained from the USGS National Water Information System server (http://waterdata.usgs.gov/usa/nwis/sw). 31 attributes from five categories (i.e., topography and location, climatic indices, soil signatures, land cover characteristics, and geological characteristics) were applied to describe the generation of baseflow, and these factors were derived from the CAMELS data sets. Among the 31 attributes, 5 are topography and location attributes, 6 are climatic indices, 6 are land cover characteristics, 11 are soil signatures, and 3 are geological characteristics types (Table 1).

Table 1
Description of 31 catchment attributes

	Name	Unit	Description		Name	Unit	Description
Topography and location	LA	° north	gauge latitude	Soil signatures	SC	cm/hr	soil conductivity
	EL	m	catchment mean elevation		MWC	m	maximum water content
	SM	m/km	catchment mean slope		SDS	m	soil depth to water and bedrock layers
	SIZ	km^2	catchment size		SIF	-	silt fraction
	LO	° east	gauged longitude		SP	-	volume porosity of soil
climate indices	SF	-	fraction of precipitation falling as snow		WF	-	water fraction in soil
	PET	mm/day	mean daily PET	Land cover characteristics	FF	-	forest fraction
	PM	mm/day	mean daily precipitation		LM	-	maximum leaf area index
	AR	-	aridity index		DL	-	dominant land cover type
	PS	-	seasonality and timing of precipitation		GVF	-	green vegetation fraction
	TEM	°C	mean temperature		RD50	m	root depth of 50% land cover vegetation
Soil signatures	CF	-	clay fraction		RD99	m	root depth of 99% land cover vegetation
	SD	m	soil depth to bedrock	Geological characteristics	GSP	-	geology subsurface porosity
	OF	%	organic fraction in soil		CR	-	carbonate rocks fraction
	OF1	%	other fraction in soil		GP	m^2	geology permeability (log10)

	Name	Unit	Description	Name	Unit	Description
	SA	-	sand fraction			

2.2 Methods

2.2.1 LH Baseflow separation

The Lyne and Hollick recursive digital filter method (the LH method), a widely used digital signal method, was used to separate the streamflow into baseflow and quick flow (Lyne and Hollick 1979). LH is easy to apply on large contiguous areas and can obtain the same results from different people through its objective and automaticity (Chapman 1991). The model formula is as follows (Nathan and McMahon 1990):

$$q_q(i) = \begin{cases} \alpha q_q(i-1) + \frac{(1+\alpha)}{2}[q(i) - q(i-1)], if q_q(i) \geq 0 \\ 0, otherwise \end{cases}$$

1

$$q_{b(i)} = q(i) - q_q(i)$$

2

where i $[d]$ is the time step; $q(i)$ [mm/d] is the measured original streamflow; $q_q(i)$ [mm/d] and $q_b(i)$ [mm/d] are the fast flow and the separated baseflow after filtering in period i, respectively; and α [1/d] is a filter parameter with a decisive role in baseflow separation. Many studies have used this method by setting α to 0.925 (Nathan and McMahon 1990; Xie et al. 2020). However, in a large-scale study with different catchment features, such as soil and terrain, a constant parameter cannot provide a satisfactory baseflow estimate for all regions (Zhang et al. 2017a). Before separating the baseflow, the ABIT was used to assess the recession parameter of each catchment. The ABIT was proposed by Cheng et al. (2016) based on the BN77 method, which eliminates the impact of evapotranspiration and avoids the uncertainty of the recession time period after the rainfall, to obtain the baseflow objectively and quickly (Brutsaert and Nieber 1977). The formulation to describing the baseflow is express as:

$$\frac{dq}{dt} = -\frac{1}{K}q$$

3

Where t is daily step, and K is the catchment characteristic drainage time $[d]$. By integrating Equation 1, the baseflow decay expression can be obtained:

$$q_b = q_0 \exp(-\frac{t}{K})$$

4

Where q_0 is the baseflow in the beginning of the study period [mm/d], and the recession constant of each catchment can be expressed as:

$$\alpha = \exp(-\frac{1}{K})$$

5

All the recession points are selected as the component of the baseflow record (Cheng et al.2016). This method aims to plot the points of logarithmic (-dq/dt\) against drought flow q and obtain the lower envelope, keeping 5% points below it (Fig. S1). As a result, the combination of the LH method and the ABIT is hydrologically significant and can more reasonably estimate the baseflow generation process.

The daily baseflow was separated from the streamflow records, and the long-term BFI was calculated by averaging the baseflow to streamflow ratio over the entire period. The baseflow and the streamflow were split into four seasons: spring (March, April, May), summer (June, July, August), autumn (September, October, November), and winter (December, January, and February of the following year). Then, the seasonal BFIs were averaged the ratio of baseflow and streamflow in each season.

The coefficient of variation (CV) was applied to assess the variability of flow on an annual scale:

$$CV = \frac{\sigma}{\mu}$$

6

where σ is the standard deviation of annual flow, and μ is the average annual flow.

2.2.2 Random Forest (RF)

The RF method is an intelligent artificial algorithm that can be used for classification and regression and has been applied in predicting the BFI by regarding the BFIs as a function of a series of selected driving factors. The RF method was proposed by Breiman (2001) and Cutler et al. (2011) based on bootstrap sampling methods and

According to the continuously growing decision trees, the non-linear relationship is integrated to connect the response with a set of predictors (Ellis et al. 2012; Singh et al. 2019). Each branch in the decision tree represents the current prediction results, and the final output of the model is determined by each decision tree in the forest. There are two aspects of randomness in the regression process that have ensured the robustness and stability of the prediction results, which are the sampling of sub-training sets and the selection of attributes for each decision tree during construction (Breiman 2001; Cutler et al. 2011).

The simulation results were qualitatively assessed with Out-Of-Bag (OOB) prediction. In OOB, the study sites were randomly divided into two groups, namely, the “gauged sites” for model training and the “ungauged sites” of model assessment. The “gauged sites” were involved in the construction of RF regression models through repeated training. The “ungauged sites” were assumed to be unobserved, and the long-term and seasonal BFIs were predicted by a series of predictors through constructed models (Booker and Snelder 2012). The prediction performance and reliability were validated qualitatively by plotting the predicted–measured values.

The RF method has three important parameters. The number of variables randomly sampled as candidates at each split is equal to 4 after successive selection. The number of decision trees in the forest is equal to 600 to stabilize error fluctuations. Importance values, which represent the increase in model prediction error when the variable is subtracted, are a variable importance matrix and measured by the residual sum of squares.

To improve the robustness of the prediction model, we developed two criteria for selecting the driving factors from the catchment attributes of CAMELS data set: 1) the driving factors have a good correlation with the BFI, and the significance level is less than 0.05; 2) no obvious collinearity exists between the selected driving factors.

2.2.3 Evaluation metrics

The Leave-Location-Out (LLO) cross-validation approach was used to test the BFI prediction performance. Unlike the traditional cross-validation method, LLO packages neighboring sites together to eliminate spatial autocorrelation and avoid the optimistic prediction caused by over-fitting (Brenning 2005; Brenning and Lausen 2010). According to the longitude and latitude of the sites on the map, the K-means clustering method was used to cluster the neighboring sites in space without considering any catchment attribute mapping of the sites. During the LLO cross-validation, each cluster was assumed to be ungauged in turns, and the remaining clusters were used as the training set to establish the prediction model. Finally, the error result of the prediction model is the mean of the deviation of each prediction on the test set (Hüllermeier et al. 2010).

Furthermore, three metrics were applied to quantitatively assess the model performance, including the NSE coefficient (Nash and Sutcliffe 1970), the bias (Legates and McCabe 1999), and the root mean squared error (RMSE) (Gupta et al. 1999). BFI_M is the “measured” BFI, BFI_P is the predicted BFI by RF simulation, and n is the number of catchments.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(BFI_M - BFI_P\right)^2}{\sum_{i=1}^{n}\left(BFI_M - \overline{BFI}_M\right)^2}$$

7

$$\text{bias} = \left| \frac{\sum_{i=1}^{n} \left(BFI_P - BFI_M \right)}{\sum_{i=1}^{n} BFI_M} \times 100\% \right|$$

8

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(BFI_M - BFI_P \right)^2}{n}}$$

9

3 Results

3.1 Spatial pattern of estimated baseflow in CONUS

The baseflow of 619 sites across CONUS was estimated based on LH method, and the spatial distributions of the long-term and seasonal BFI are shown in Fig. 2. In general, the long-term average BFI of all sites range from 0.02 to 0.98, with an average value of 0.49. The BFIs were markedly higher in the northern-central part of CONUS (e.g., Michigan along with the shores of Lake Michigan and Southwestern Wisconsin) than those in the other parts, that is, approximately between 0.7 and 1. The BFIs were high mainly because these regions are located along the coast of Lake Michigan, which has sufficient water supply and high soil moisture, which are conducive to the generation of baseflow (Wolock 2003; Zhang et al. 2013). Meanwhile, the BFIs in the eastern and northwestern regions of CONUS ranged from 0.46 to 0.75, which is higher than the national average level. However, the BFIs of the central region were generally below the national average level, especially in the middle Mississippi Valley regions and the Western Gulf Coast, where the contribution of desert rock landscape and the lack of vegetation coverage led to weak soil storage capacity and most of the rainfall flows away in the form of quick flow.

In seasonal BFIs, the average BFI in winter was the highest at 0.55; the average BFIs in spring and summer were lower at 0.49 and 0.53, respectively; and the lowest BFI in autumn was 0.48. However, anomalies occurred in the Midwest CONUS (e.g., Rocky Mountains), which exhibited the highest BFIs in autumn and winter and the lowest range in spring. This phenomenon occurred because these areas are dominated by snowfall, and the freezing period generally begins early in October. Therefore, when the season shifts to autumn, streamflow is mostly used to recharge the subsurface and leads to sufficient baseflow, resulting in high BFIs in the four seasons. On the contrary, as the temperature gradually rises in spring, most of the ice melts and forms quick flow, but the baseflow responds with latency and presents the lowest BFI pattern in the season (Bryant et al. 2020; Karlstrom and Houston 1984; Liefert et al. 2018).

Furthermore, the coefficient of variation (CV) was used to measure the variation of baseflow and streamflow for each site during the study period, as shown in Fig. 3. The comparison of the long–term BFI distribution patterns (Fig. 2e) indicates that the BFI is negatively correlated with the CV, which is consistent with the results of the Longobardi and Villani (2020). The station with a high BFI is usually accompanied by a low CV, indicating that in the high BFI regions, both baseflow and streamflow will generally become more stable than those in the low BFI

regions. For example, the BFIs were generally higher (0.5–0.75) in the Pacific Northwest regions of CONUS, but the CVs of baseflow (0–0.3) and streamflow (0.4–1) were lower than the mean national levels.

3.2 Systematic analysis on driving factors of BFI

3.2.1 Selection of driving factors

The driving factors were selected from the 31 catchment attributes are shown in Fig. 4, and the results of significance analysis and collinearity calculation among selected factors are described in Table S1, S2. Six attributes (e.g., clay fraction (CF), soil depth to bedrock (SD), mean of catchment elevation (EL), forest fraction (FF), subsurface porosity (GSP), snow fraction (SF)) were selected as dominant factors to predict the long-term and seasonal BFIs. Although the mean of catchment slope (SM) factor has a significant positive correlation with BFI, it has a serious collinearity with other factors, so it was deleted. CF and SD represent the soil characteristics and could reflect the local water retention capacity (Bloomfield et al. 2009b). EL is the topography and location features and reflect the influence of terrain conditions on hydrological processes (Santhi et al. 2008). FF is a landcover factor that reflects vegetation coverage. GSP is the feature of geology, which indirectly reflects the aquifer condition of different regions. SF refers to the fraction of precipitation falling as snow when the temperature is below 0°C and is included in the climate characteristics that affect the baseflow generation under the influence of low temperature (Xie et al. 2020).

3.2.2 Importance of driving factors

The generation of baseflow is always connected to local catchment attributes (Xie et al. 2020; Zhang et al. 2020). To understand the baseflow generation mechanism under the influence of the catchment attributes, the RF method was used to evaluate their importance because the method can assign an importance value to each attribute, and the results are shown in Fig. 5. Importance is equivalent to the error that results from deleting the attribute from the linear model, which represents the accuracy contribution of the attribute when it is independently predicted (Booker and Snelder 2012). The forest fraction (FF) is the most important predictor for the long-term average BFI with an importance value of 3.9, followed by the clay fraction (CF) predictor according to their importance values in the RF method. Forest fraction determines the proportion of baseflow in most catchments during the long period. Given that lush forests can affect the formation of soil texture and vegetation landscape and improve the soil storage capacity, they provide a suitable environment for the formation of baseflow (Zhang et al. 2019a). Furthermore, the generation mechanism of baseflow is varies throughout the seasons. In spring, the generation mechanism of baseflow is similar to the long-term BFI, with the forest proportion having the greatest influence, followed by the clay proportion. The generation mechanism of baseflow in autumn and winter is similar. The most influential driving factor is the snow fraction (SF) and then the average elevation (EL) factor has flow. This phenomenon is mainly due to the special climate in the high-altitude regions of Midwestern CONUS, with an early and long winter (usually beginning in October and ending in June) and an annual 1,000-mm precipitation dominated by snow (Musselman et al. 1994).

3.3 Prediction performance of long-term and seasonal BFIs

The LLO cross-validation approach was used to evaluate the performance of the RF technique in terms of predicting long-term and seasonal BFIs. Meanwhile, for comparative analysis, we applied the widely used multiple linear regression (MLR) method to predict the BFIs (including long-term and seasonal BFIs) and adopted the LLO

cross-validation approach to evaluate the prediction performance as well. The validation results are shown in Fig. 6. The points represent the predicted-measured BFI values of all the sites, and the scatter points generally follow the 1:1 linear distribution. The RF models exhibited better performance than the MLR method in predicting long-term and seasonal BFIs in terms of R^2, indicating that the RF prediction results are reliable. The quantitative validation of the predicted performance was assessed by the NSE, the RE, and the bias, as shown in Table 2. The RF technique is better than MLR for all metrics. Thus, the six selected driving factors are appropriate for BFI prediction based on the RF technique for the 619 catchments in CONUS.

Table 2
The performance of models for predicting long-term and seasonal BFI

Methods	**Random Forest technique**					**Multiple Linear Regression**				
BFIs	long-term	Spring	Summer	Autumn	Winter	long-term	Spring	Summer	Autumn	Winter
NSE	0.59	0.55	0.52	0.60	0.62	0.28	0.30	0.27	0.40	0.43
Bias (%)	16	15	19	15	12	29	28	35	30	22
RMSE	0.11	0.10	0.15	0.12	0.09	0.24	0.22	0.25	0.26	0.18

To illustrate the performance of spatial prediction, we interpolated the long-term and seasonal BFI predictions and plotted the error of each site in Fig. 7. Compared with Fig. 2, the predicted BFI spatial patterns in Fig. 7 show good agreement with the estimated values. In the long-term BFI spatial prediction, 489 of 619 "ungauged" sites have a reasonable prediction with an error of less than 0.3. Overall, the prediction performance of the model is satisfactory, especially in the area with high BFI value, such as the northwestern regions of CONUS and the Atlantic coast.

4 Discussion

To better understand the existence of baseflow on a large scale, the LH method was used with the ABIT to estimate the baseflow of 619 sites across CONUS, and the RF algorithm was applied for an in-depth and systematic analysis of the driving factors of the BFI for exploring the generation mechanism of baseflow. Consequently, the FF and CF played an important role in the long-term baseflow generation, and the SF is dominant in the generation of seasonal baseflow (especially in autumn and winter). Additionally, the large-scale long-term and seasonal BFIs prediction based on the RF technique exhibited promising results.

4.1 Variations of baseflow, and BFI

The spatiotemporal patterns of streamflow, baseflow, and BFI across CONUS were analyzed in this study, and the pattern of BFI is consistent with the previous study by the U.S. Geological Survey (Wolock 2003). Many studies have explored the baseflow/BFI variation in a large-scale area and analyzed the impacts under the catchment attributes. Santhi et al. (2008) applied the USGS data set to explore the BFI spatial variation in CONUS, but there were differences. Their study is to explore the correlation between baseflow and hydro-geologic variables in different hydrological landscape regions, and the purpose of this work is to analyze the influence of natural driving factors on the baseflow generation and make accurate prediction. Gudmundsson et al. (2019) studied the historical trend of global river flow and assesses the availability of global water resources by focusing on changes in low flows, average flows, and high flows. They have done meaningful work on an overview of global flows and provide support for the analysis of the baseflow mechanisms. Furthermore, the changes in baseflow and streamflow at each site during the study period were also considered and measured as CVs. The comparison of the long-term BFI

distribution (Fig. 2e) and the CV distribution (Fig. 3) across CONUS indicates that the CVs of baseflow and streamflow present a significant negative correlation with the BFI of -0.54 and − 0.59, respectively, as shown in Fig. 8. This result indicates that regions with high BFIs are usually accompanied by stable baseflow and streamflow. A similar study was conducted by Bastola et al. (2018) who tried to analyze the contribution of the monthly and annual baseflow changes to streamflow, and their results show that a region with a high BFI usually has a low CV value. Thus, attention must be given to low BFI regions whose baseflow and streamflow are vulnerable, which may threaten the sustainable development of water resources.

4.2 Mechanism of baseflow generation under various catchments attributes

The FF exhibit good performance for the long-term BFI prediction across CONUS. In general, FF is the most important indicator for evaluating baseflow, that is, the higher forest coverage is, the more permeable the soil and the greater the water retention capacity are (Lacey and Grayson 1988). Although some studies have suggested that higher watershed forest cover increases evapotranspiration rates and thus leads to lower baseflow, many other studies have shown that higher catchment forest cover increases baseflow due to higher soil permeability providing adequate groundwater recharge (Bruijnz Ee L 2004; Bruijnz Ee L and Unesco 1991). Ma et al. (2009) conducted a study of the watershed in southwestern China, and the results showed that the reforestation increased baseflow and reduced streamflow, and attributed this phenomenon to the increase in soil infiltration. Price et al. (2011) showed that there was a significant positive correlation between forest cover and baseflow in the southern Appalachian Highlands. The CF is the second important factor, which is conducive to the infiltration of surface water and provides a favorable environment for the baseflow generation (Bloomfield et al. 2009a). Xu et al. (2013) discussed the baseflow generation by using the climate elasticity approach to assess the sensitivity of hydrological changes to climate and land surface changes in Midwest CONUS and found that the influence of land surface changes on the baseflow and the BFI are significantly greater than the effects of climate change, and this study proves this.

Besides, snow fraction is an important driving factor on baseflow generation as well. When the temperature reaches a certain point, seasonal snowmelt will lead to the formation of peak discharge in spring and summer (Zhang et al. 2014). Most of the water melt from the snowpack will be supplied to the river in the form of quick flow. Besides, during the snowmelt period, the presence of frozen soil hinders the infiltration of water and accelerates the generation of quick flow (Shanley and Chalmers 1999). Therefore, in the high-altitude frozen catchment of CONUS, the BFI is the lowest in the spring and summer, which is consistent with the seasonal changes we have analyzed. The snow fraction factor has the most obvious impact on the generation of BFI in autumn and winter, especially reflected in the Midwestern CONUS, as shown in Fig. 2 (c, d). Catchments in the Midwestern CONUS have the high-altitude, low-temperature and perennial snow cover basin characteristics, which have an insignificant increase of BFI in the autumn and winter. Because these catchments have fewer precipitation in autumn and winter, lead to fewer replenishment for streamflow, and the average temperature in winter is below 0 degrees Celsius for most sites, thus the streamflow is significantly less, but the decline in the baseflow is relatively stable (Woods 2009). Considering the BFI is in fact relative relationship between baseflow and streamflow, the ratio of base flow to streamflow (BFI) is large, and BFI has a clear increasing trend during these seasons.

However, the snowmelt process can also bring uncertain effects on the generation of baseflow. In large catchments, the snowmelt rate varies with the geographical conditions, the physical environment, and the elevation of the catchments (Zhang et al. 2016). In addition, most of the snowmelt water does not directly supplement the river

discharge but enters the snowdrift for the refreezing process (Marsh and Woo 1984). These uncertain processes pose a challenge to analyzing the generation of the baseflow.

4.3 BFI prediction for water resource planning and management

The RF technique was used in this study to predict long-term and seasonal BFIs, and the R^2, NSE, bias, and RMSE indexes are better than those in the MLR method, as shown in Table 2. The results indicate that the RF technique, which trains many decision trees with the randomness of attributes and sub-training sets, has good applicability in large-scale learning prediction. Besides, there are many studies using RF technology to predict hydrology and achieved pretty results. Olson and Hawkins (2013) established an RF model to predict the continuous spatial variation of total P and total N in rivers of the western United States and predicted better than the previous physical model. Fouad and Loaiciga (2020) evaluated the regression models of the percentile flow in 918 basins of CONUS and believed that the prediction effect of the RF technique is superior to the baseline regression procedure. Although RF is an empirical technique that cannot reflect any physical mechanism, the six predictors selected cover the characteristics of basin topography, soil, geography, land cover, and climate. The comparison of different combinations of catchment attributes and the application of MLR methods indicate that the current prediction model showed a satisfactory performance, especially in areas with high BFI values. Therefore, integrating digital filtering analysis and the RF technique into a framework for large-scale baseflow separation and prediction is a promising approach, which can provide an effective method for low-flow prediction and water ecological management.

Baseflow is always connected to complex hydrogeology conditions and affected by human activities and climate change (Rodiger et al. 2020; Tan et al. 2020). The catchments of the CAMELS data set used in this study are minimally affected by human activities, which reduces the impact of human activities on the generation of baseflow. In the following research, it is meaningful to explore the driving mechanism of human impact (such as reservoir construction, water resource regulation, vegetation protection, etc.), as well as the driving mechanism under the combined influence of catchment attributes (River and Richardson (2018), Zhu et al. (2019)). It is worth noting that, Cheng et al. (2016) proposed an automatic technique based on BN77 theory (Brutsaert and Nieber 1977) to obtain the recession constant more objectively and quickly. Singh et al. (2019); Zhang et al. (2020) applied this method to the baseflow analysis in the Australian and New Zealand basins and obtained satisfactory performances.

5 Conclusions

This study aims to systematically analyze the spatial variation of baseflow and the driving factors of baseflow and provide a reliable approach to BFI prediction. On the basis of the study of CONUS from years 1981 to 2014, the preliminary conclusions were as followed:

1) In the spatial pattern, the Great Lakes region and the Rocky Mountains have the highest BFIs across CONUS, with enough water supply. However, the BFI is lower in the central region, especially in the middle of the Mississippi Valley and along the Western Gulf Coast, where the existing desert rock landscape and the sparse vegetation cover lead to weak soil storage capacity, resulting in the conversion of rainfall into quick flow.

2) In the long-term baseflow generation, FF and CF are the most important natural factors to regulate the baseflow situation in CONUS. Among the six selected driving factors, SF plays the most important role in seasonal baseflow

3) The results of long-term and seasonal BFIs prediction based on the RF technique are satisfactory, especially in regions with high BFI values and the technique can be applied to similar research.

Declarations

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China (Grant No. 51979198). Greatly thank Newman et al. (2015) and Addor et al. (2017) for providing hydrometeorological time series dataset and catchment attributes in the Contiguous United States. The data are available at https://ral.ucar.edu/solutions/products/camels.

Funding: This work is financially supported by the National Natural Science Foundation of China (Grant No. 51979198), and chaired by Qianjin Dong.

Conflicts of interest/Competing interests: We declare there is no conflict of interest, and no conflict of interest exists in the submission of this manuscript.

Availability of data and material: The data used in this work are from the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) datasets, available at https://ral.ucar.edu/solutions/products/camels.

Code availability: The R language was used to finish this work.

Authors' contributions:

Shanshan Huang: Conceptualization, Methodology, Writing - original draft preparation; Qianjin Dong: Funding acquisition, Writing - review and editing, Supervision; Xu Zhang: Writing - review and editing; Weishan Deng: Writing - review and editing.

References

1. Addor N, Newman AJ, Mizukami N, Clark MP (2017) The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol Earth Syst Sci 21:5293–5313. doi:10.5194/hess-21-5293-2017
2. Araza A, Perez M, Cruz RV, Aggabao LF, Soyosa E (2020) Probable streamflow changes and its associated risk to the water resources of Abuan watershed, Philippines caused by climate change and land use changes Stochastic Environmental Research and Risk Assessment doi:10.1007/s00477-020-01953-3
3. Baak M, Koopman R, Snoek H, Klous S (2020) A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics Comput Stat Data Anal 152 doi:10.1016/j.csda.2020.107043
4. Bastola S et al (2018) Contribution of Baseflow to River Streamflow: Study on Nepal's Bagmati and Koshi Basins. KSCE J Civ Eng 22:4710–4718. doi:10.1007/s12205-018-0149-9
5. Beck HE, Van Dijk AIJM, Miralles DG, De Jeu RAM, Bruijnzeel LA, McVicar TR, Schellekens J (2013) Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resour Res 49:7843–7863. doi:10.1002/2013WR013918
6. Bloomfield JP, Allen DJ, Griffiths KJ (2009a) Examining geological controls on Baseflow Index (BFI) using regression. J Hydrol 373:164–176

7. Bloomfield JP, Allen DJ, Griffiths KJ (2009b) Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin. UK Journal of Hydrology 373:164–176. doi:10.1016/j.jhydrol.2009.04.025
8. Booker DJ, Snelder TH (2012) Comparing methods for estimating flow duration curves at ungauged sites. J Hydrol 434–435:78–94. doi:https://doi.org/10.1016/j.jhydrol.2012.02.031
9. Bosch DD, Arnold JG, Allen PG, Lim KJ, Park YS (2017) Temporal variations in baseflow for the Little River experimental watershed in South Georgia. USA Journal of Hydrology: Regional Studies 10:110–121. doi:10.1016/j.ejrh.2017.02.002
10. Breiman L (2001) Random forests Machine Learning 45:5–32. doi:10.1023/a:1010933404324
11. Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation Natural hazards and earth system sciences 5
12. Brenning A, Lausen B (2010) Estimating error rates in the classification of paired organs. Stat Med 27:4515–4531
13. Bruijnz Ee LLA (2004) Hydrological Functions of Tropical Forests: Not Seeing the Soil for the Trees? Agriculture Ecosystems Environment 10401:185–228
14. Bruijnz Ee LLA, Unesco P (1991) Hydrology of moist tropical forests and effects of conversion:. a state of knowledge review
15. Brunner MI, Furrer R, Sikorska AE, Viviroli D, Seibert J, Favre A-C (2018) Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods. Stoch Env Res Risk Assess 32:1993–2023. doi:10.1007/s00477-018-1523-3
16. Brutsaert W, Nieber JL (1977) Regionalized drought flow hydrographs from a mature glaciated plateau. Water Resour Res 13:637–643
17. Bryant SR et al (2020) Seasonal manganese transport in the hyporheic zone of a snowmelt-dominated river (East River, Colorado, USA). Hydrogeol J 28:1323–1341. doi:10.1007/s10040-020-02146-6
18. Cawley GC, Talbot NLC (2003) Efficient leave-one-out cross-validation of kernel Fisher discriminant. classifiers Pattern Recognit 36:2585–2592. doi:10.1016/s0031-3203(03)00136-5
19. Chapman TG (1991) Comment on “Evaluation of automated techniques for base flow and recession analyses” by Nathan RJ and T. A. McMahon Water Resources Research 27:1783–1784 doi:10.1029/91wr01007
20. Cheng L, Zhang L, Brutsaert W (2016) Automated Selection of Pure Base Flows from Regular Daily Streamflow Data: Objective Algorithm Journal of Hydrologic Engineering 21:06016008 doi:doi:10.1061/(ASCE)HE.1943-5584.0001427
21. Collins SL et al (2020) Groundwater connectivity of a sheared gneiss aquifer in the Cauvery River basin. India Hydrogeology Journal 28:1371–1388. doi:10.1007/s10040-020-02140-y
22. Cook PG, Lamontagne S, Berhane D, Clark JF (2006) Quantifying groundwater discharge to Cockburn River, southeastern Australia, using dissolved gas tracers 222Rn and SF6 Water Resour Res 42 doi:10.1029/2006wr004921
23. Cutler A, Cutler D, Stevens J (2011) Random Forests. In, vol 45. pp 157–176. doi:10.1007/978-1-4419-9326-7_5
24. Eckhardt K (2005) How to construct recursive digital filters for baseflow separation. Hydrol Process 19:507–515. doi:10.1002/hyp.5675
25. Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests: calculating importance gradients on. physical predictors

26. Fan Y, Li H, Miguez-Macho G (2013) Global Patterns of Groundwater. Table Depth Science 339:940–943. doi:10.1126/science.1229881
27. Fouad G, Loaiciga HA (2020) Independent variable selection for regression modeling of the flow duration curve for ungauged basins in the United States. J Hydrol 587:9. doi:10.1016/j.jhydrol.2020.124975
28. Furey PR, Gupta VK (2001) A physically based filter for separating base flow from streamflow time series. Water Resour Res 37:2709–2722. doi:10.1029/2001wr000243
29. Gan R, Sun L, Luo Y (2015) Baseflow characteristics in alpine rivers — a multi-catchment analysis in Northwest China. J Mt Sci 12:614–625. doi:10.1007/s11629-013-2959-z
30. Georgek JL, Solomon DK, Heilweil VM, Miller MP (2018) Using tracer-derived groundwater transit times to assess storage within a high-elevation watershed of the upper Colorado River Basin. USA Hydrogeology Journal 26:467–480. doi:10.1007/s10040-017-1655-4
31. Goncalves J, Mahamat Nour A, Bouchez C, Deschamps P, Vallet-Coulomb C (2020) Recharge and baseflow constrained by surface-water and groundwater chemistry: case study of the Chari River. Chad basin Hydrogeology Journal. doi:10.1007/s10040-020-02259-y
32. Gudmundsson L, Leonard M, Do HX, Westra S, Seneviratne SI (2019) Observed Trends in Global Indicators of Mean and Extreme Streamflow. Geophys Res Lett 46:756–766. doi:https://doi.org/10.1029/2018GL079725
33. Gupta HV, Sorooshian S, Yapo PO (1999) Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. J Hydrol Eng 4:135–143. doi:10.1061/(asce)1084-0699(1999)4:2(135)
34. Hall F (1968) Base-Flow Recessions-A Review. Water Resour Res 4:973–983
35. Henderson FM, Wooding RA (1964) Overland flow and groundwater flow from a steady rainfall of finite duration. J Geophys Res 69:1531–1540. doi:10.1029/JZ069i008p01531
36. Hüllermeier E, Kruse R, Hoffmann F (2010) Computational Intelligence for Knowledge-Based Systems Design. Springer, Berlin Heidelberg
37. Jones JP, Sudicky EA, Brookfield AE, Park Y-J (2006) An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow Water Resour Res 42 doi:10.1029/2005wr004130
38. Karlstrom KE, Houston RS (1984) The cheyenne belt: analysis of a proterozoic suture in Southern. Wyoming Precambrian Research 25:415–446. doi:https://doi.org/10.1016/0301-9268(84)90012-3
39. Lacey GC, Grayson RB (1988) Relating baseflow to catchment properties in south-eastern Australia. J Hydrol 204:231–250
40. Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241. doi:10.1029/1998wr900018
41. Liefert DT, Shuman BN, Parsekian AD, Mercer JJ (2018) Why Are Some Rocky Mountain Lakes Ephemeral? Water Resour Res 54:5245–5263. doi:10.1029/2017wr022261
42. Longobardi A, Villani P (2020) From at-site to regional assessment of environmental flows and environmental flows variability in a Mediterranean environment. Journal of Hydrology: Regional Studies 32:100764. doi:https://doi.org/10.1016/j.ejrh.2020.100764
43. Lott DA, Stewart MT (2016) Base flow separation: A comparison of analytical and mass balance methods. J Hydrol 535:525–533. doi:https://doi.org/10.1016/j.jhydrol.2016.01.063
44. Loveridge M, Rahman A (2014) Quantifying uncertainty in rainfall–runoff models due to design losses using Monte Carlo simulation: a case study in New South Wales. Australia Stochastic Environmental Research Risk 4-0862-y

45. Lyne VD, Hollick M Stochastic Time-Variable Rainfall-Runoff Modeling. In: Aust. Natl. conf. Publ. (pp. 89–93), 1979
46. Ma X, Xu JC, Luo Y, Aggarwal SP, Li JT (2009) Response of hydrological processes to land-cover and climate changes in Kejie watershed, south-west. China Hydrol Process 23:1179–1191. doi:10.1002/hyp.7233
47. Marsh P, Woo MK (1984) WETTING FRONT ADVANCE AND FREEZING OF MELTWATER WITHIN A SNOW COVER.1. OBSERVATIONS IN THE CANADIAN ARCTIC Water Resour Res 20:1853–1864 doi:10.1029/WR020i012p01853
48. McNamara JP, Kane DL, Hinzman LD (1997) Hydrograph separations in an Arctic watershed using mixing model and graphical techniques. Water Resour Res 33:1707–1719. doi:10.1029/97WR01033
49. Molla DD, Tegaye TA (2019) “Multivariate analysis of baseflow index in complex rift margin catchments: The case of Abaya-Chamo lakes basin, southern Ethiopia” Groundwater for Sustainable Development 9 doi:10.1016/j.gsd.2019.100236
50. Musselman RC, Fox DG, Schoettle AW, Regan CM (1994) Introduction: The glacier lakes ecosystem experiments site General technical report RM - Rocky Mountain Forest and Range Experiment Station, US Department of Agriculture, Forest Se:1–10
51. Nash J, Sutcliffe I (1970) River flow forecasting through conceptual models A Discussion of Principles Journal of Hydrology 10
52. Nathan RJ, McMahon TA (1990) Evaluation of automated techniques for base flow and recession analyses. Water Resour Res 26:1465–1473. doi:10.1029/WR026i007p01465
53. Newman AJ 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 19:209–223 doi:10.5194/hess-19-209-2015
54. Olson JR, Hawkins CP (2013) Developing site-specific nutrient criteria from empirical models Freshwater Science 32:719–740 doi:10.1899/12-113.1
55. Price K, Jackson CR, Parker AJ, Reitan T, Dowd J, Cyterski M (2011) Effects of watershed land use and geomorphology on stream low flows during severe drought conditions in the southern Blue Ridge Mountains, Georgia and North Carolina, United States Water Resources Research 47:p.W02516.02511-W02516.02519
56. River M, Richardson CJ (2018) Stream transport of iron and phosphorus by authigenic nanoparticles in the Southern Piedmont. of the US Water Res 130:312–321. doi:10.1016/j.watres.2017.12.004
57. Rodiger T, Magri F, Geyer S, Mallast U, Odeh T, Siebert C (2020) Calculating man-made depletion of a stressed multiple aquifer resource on a national scale. Sci Total Environ 725:16. doi:10.1016/j.scitotenv.2020.138478
58. Santhi C, Allen PM, Muttiah RS, Arnold JG, Tuppad P (2008) Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. J Hydrol 351:139–153. doi:10.1016/j.jhydrol.2007.12.018
59. Sapač K, Rusjan S, Šraj M (2020) Assessment of consistency of low-flow indices of a hydrogeologically non-homogeneous catchment: A case study of the Ljubljanica river catchment. Slovenia Journal of Hydrology 583:124621. doi:https://doi.org/10.1016/j.jhydrol.2020.124621
60. Seo SB, Mahinthakumar G, Sankarasubramanian A, Kumar M (2018) Assessing the restoration time of surface water and groundwater systems under groundwater pumping. Stoch Env Res Risk Assess 32:2741–2759. doi:10.1007/s00477-018-1570-9

61. Shanley JB, Chalmers A (1999) The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont Hydrol Process 13:1843–1857 doi:10.1002/(sici)1099-1085(199909)13:12/13<1843::Aid-hyp879>3.0.Co;2-g
62. Singh SK, Pahlow M, Booker DJ, Shankar U, Chamorro A (2019) Towards baseflow index characterisation at national scale in New Zealand. J Hydrol 568:646–657. doi:10.1016/j.jhydrol.2018.11.025
63. Tan X, Liu B, Tan X (2020) Global Changes in Baseflow Under the Impacts of Changing Climate and Vegetation Water Resour Res 56 doi:10.1029/2020wr027349
64. Wolock DM (2003) Base-Flow Index Grid for the Conterminous United States Center for Integrated. Data Analytics Wisconsin Science Center
65. Woods RA (2009) Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks Advances. in Water Resources 32:1465–1481. doi:10.1016/j.advwatres.2009.06.011
66. Xie J, Liu X, Wang K, Yang T, Liang K, Liu C (2020) Evaluation of typical methods for baseflow separation in the contiguous United States J Hydrol 583 doi:10.1016/j.jhydrol.2020.124628
67. Xu X, Scanlon BR, Schilling K, Sun A (2013) Relative importance of climate and land surface changes on hydrologic changes in the US Midwest since the 1930s: Implications for biofuel production. J Hydrol 497:110–120. doi:10.1016/j.jhydrol.2013.05.041
68. Yang C, Zhang Y-K, Liang X (2018) Analysis of temporal variation and scaling of hydrological variables based on a numerical model of the Sagehen Creek watershed Stochastic Environmental. Research Risk Assessment 32:357–368. doi:10.1007/s00477-017-1421-0
69. Yang QN, Li ZB, Han Y, Gao HD (2020) Responses of Baseflow to Ecological Construction and Climate Change in Different Geomorphological Types in The Middle Yellow River. China Water 12:15. doi:10.3390/w12010304
70. Zhang F, Ahmad S, Zhang H, Zhao X, Feng X, Li L (2016) Simulating low and high streamflow driven by snowmelt in an insufficiently gauged alpine basin. Stoch Env Res Risk Assess 30:59–75. doi:10.1007/s00477-015-1028-2
71. Zhang F-Y, Li L-H, Ahmad S, Li X-M (2014) Using path analysis to identify the influence of climatic factors on spring peak flow dominated by snowmelt in an alpine watershed. J Mt Sci 11:990–1000. doi:10.1007/s11629-013-2789-z
72. Zhang J et al (2019a) Baseflow estimation for catchments in the Loess Plateau. China Journal of Environmental Management 233:264–270. doi:https://doi.org/10.1016/j.jenvman.2018.12.040
73. Zhang J, Zhang Y, Song J, Cheng L (2017a) Evaluating relative merits of four baseflow separation methods in Eastern Australia. J Hydrol 549:252–263. doi:https://doi.org/10.1016/j.jhydrol.2017.04.004
74. Zhang J et al (2020) Large-scale baseflow index prediction using hydrological modelling, linear and multilevel regression approaches J Hydrol 585 doi:10.1016/j.jhydrol.2020.124780
75. Zhang J, Zhang ZG, Lu M, Wang X, Shang X, Elias SB, Chopp M (2017b) MiR-146a promotes remyelination in a cuprizone model of. demyelinating injury Neuroscience 348:252–263. doi:10.1016/j.neuroscience.2017.02.029
76. Zhang Y, Liu S, Hou X, Cheng F, Shen Z (2019b) Landscape- and climate change-induced hydrological alterations in the typically urbanized Beiyun River basin, Beijing, China Stochastic Environmental Research. and Risk Assessment 33:149–168. doi:10.1007/s00477-018-1628-8
77. Zhang YQ, Ahiablame L, Engel B, Liu JM (2013) Regression Modeling of Baseflow and Baseflow Index for. Michigan USA Water 5:1797–1815. doi:10.3390/w5041797

Figures

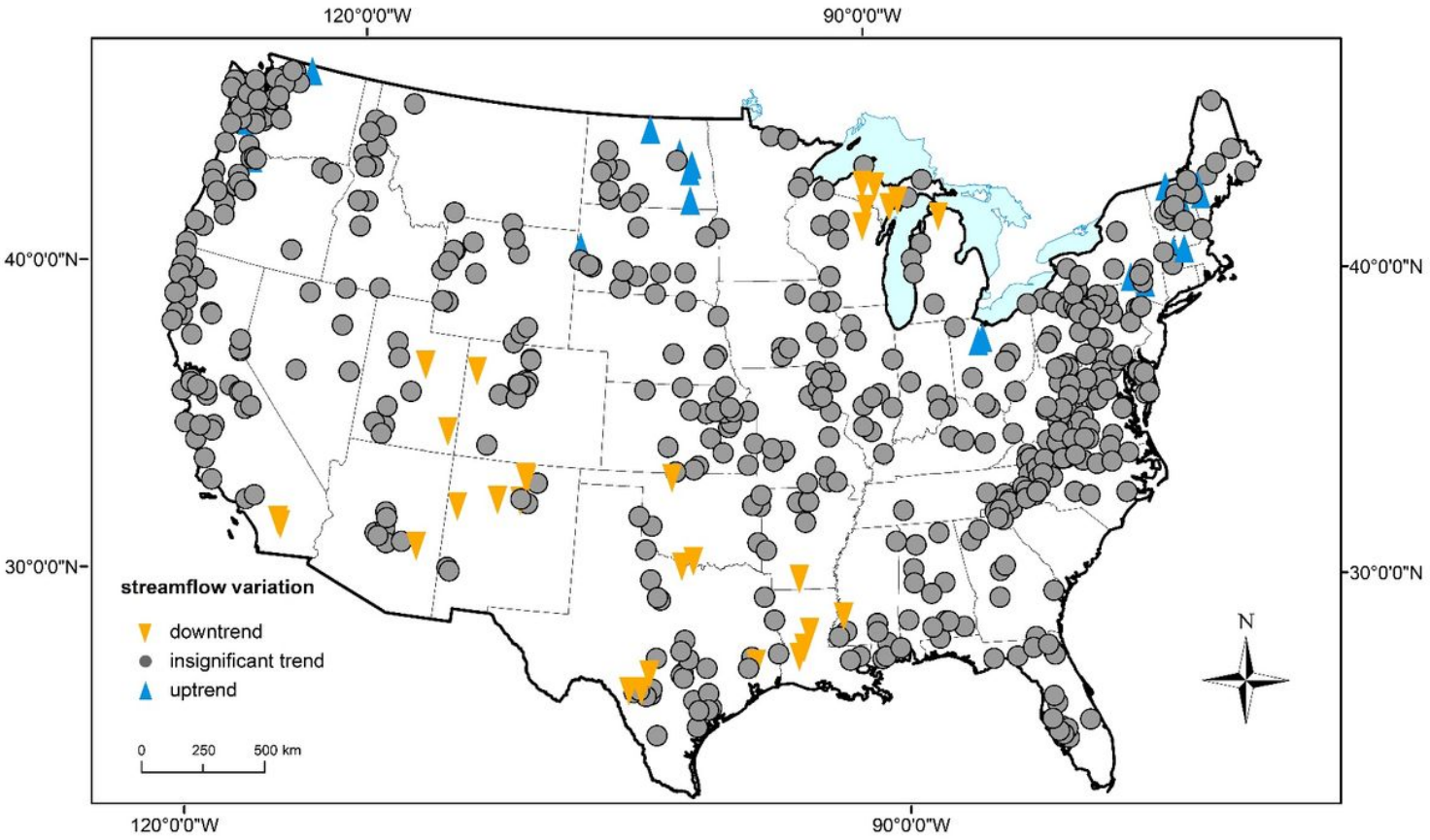

Figure 1

Outlet locations of the 619 catchments across the CONUS. The MK trend test was used to evaluate the streamflow variation in a 95% significance interval. The orange triangles, gray points, and blue triangles represent the different trends of streamflow during 1981-2004 period.

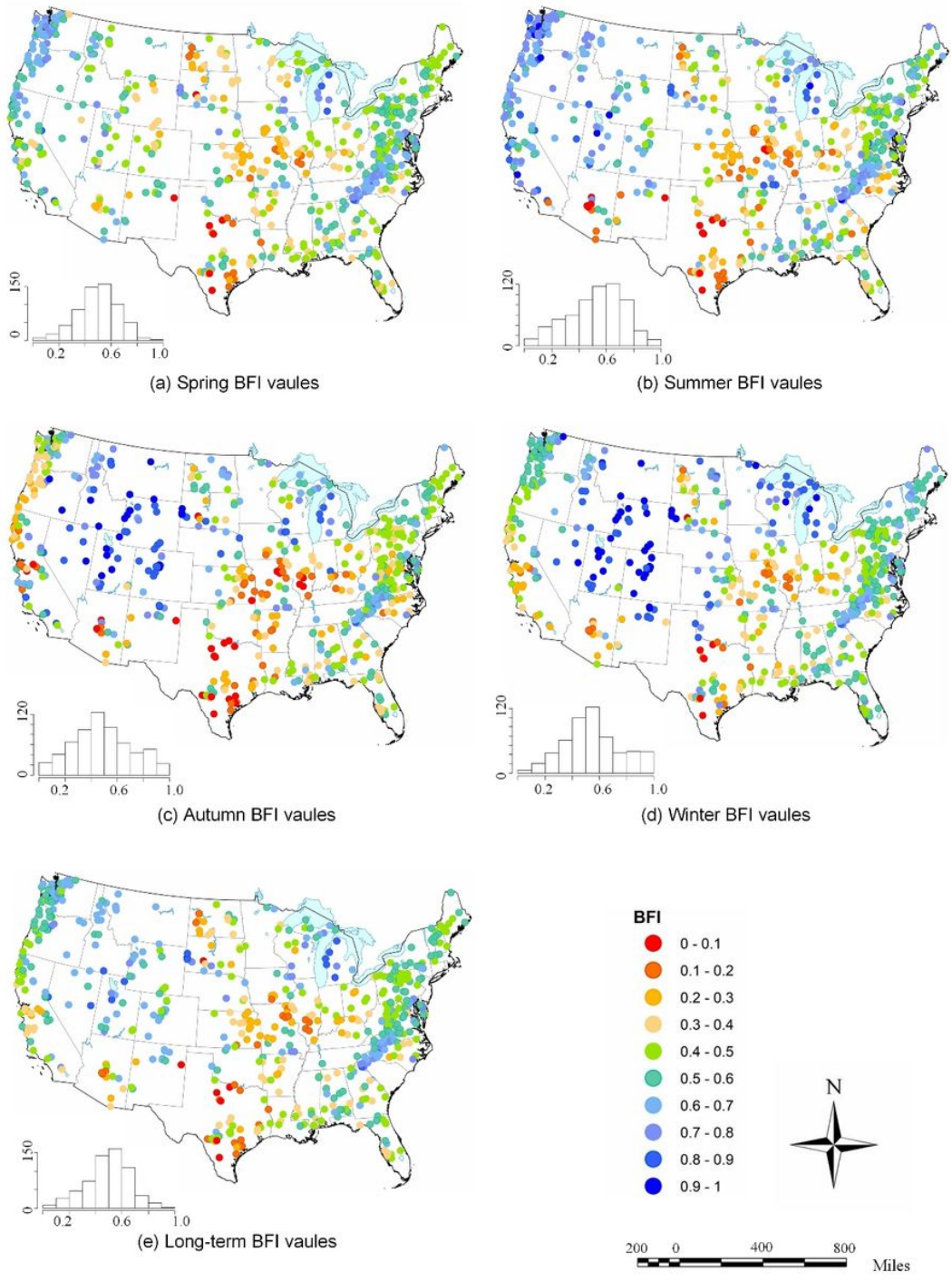

Figure 2

BFI distribution in long-term and different seasons for 619 sites across CONUS

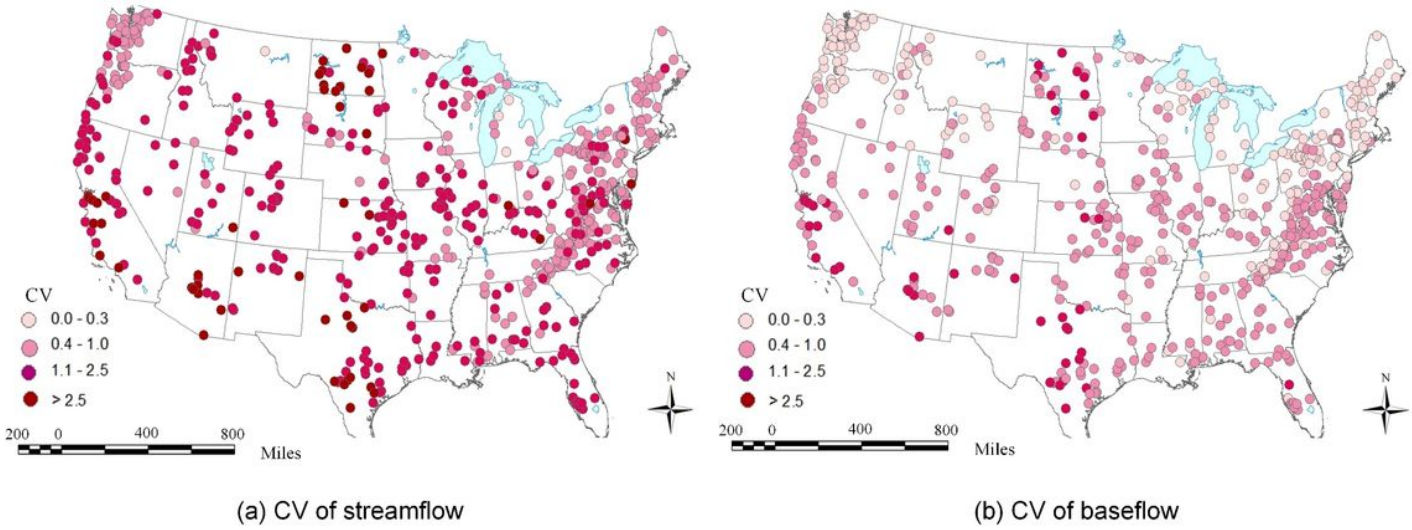

Figure 3

CV distribution of streamflow (a) and baseflow (b) during the study period

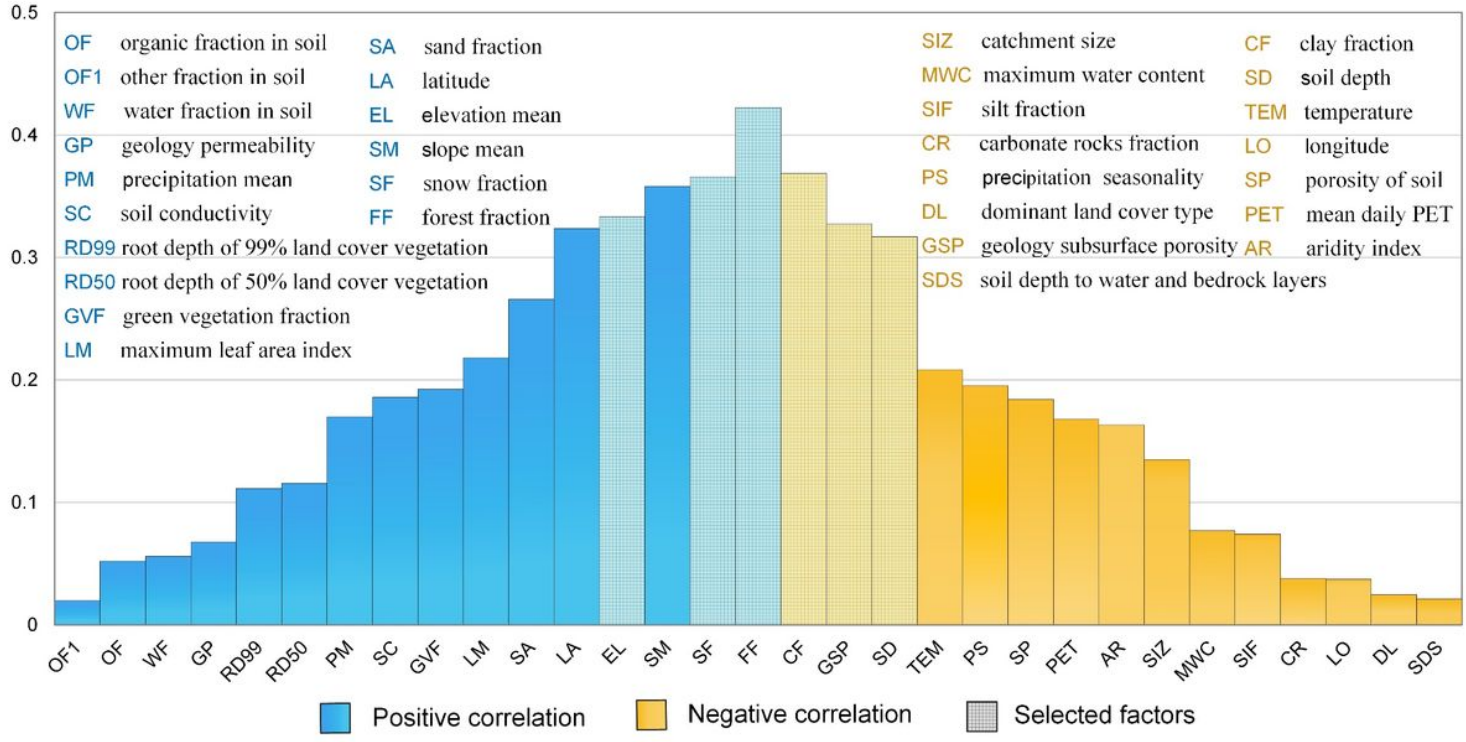

Figure 4

The correlations between catchment attributes and BFI by averaging over the 619 catchments

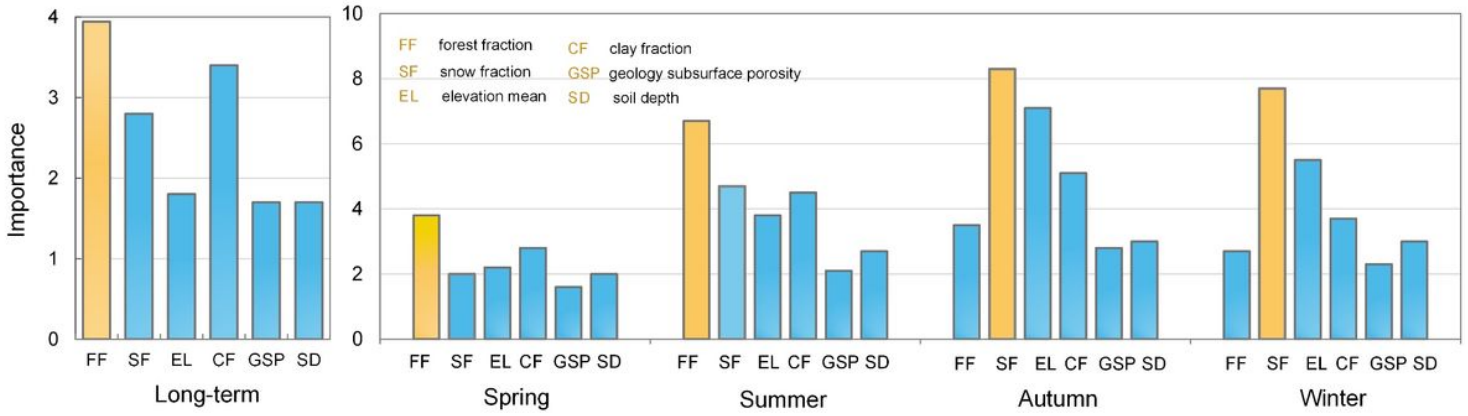

Figure 5

Importance values of predictors for determine long-term average BFI and seasonal BFIs, and the most influential factor is indicated in yellow

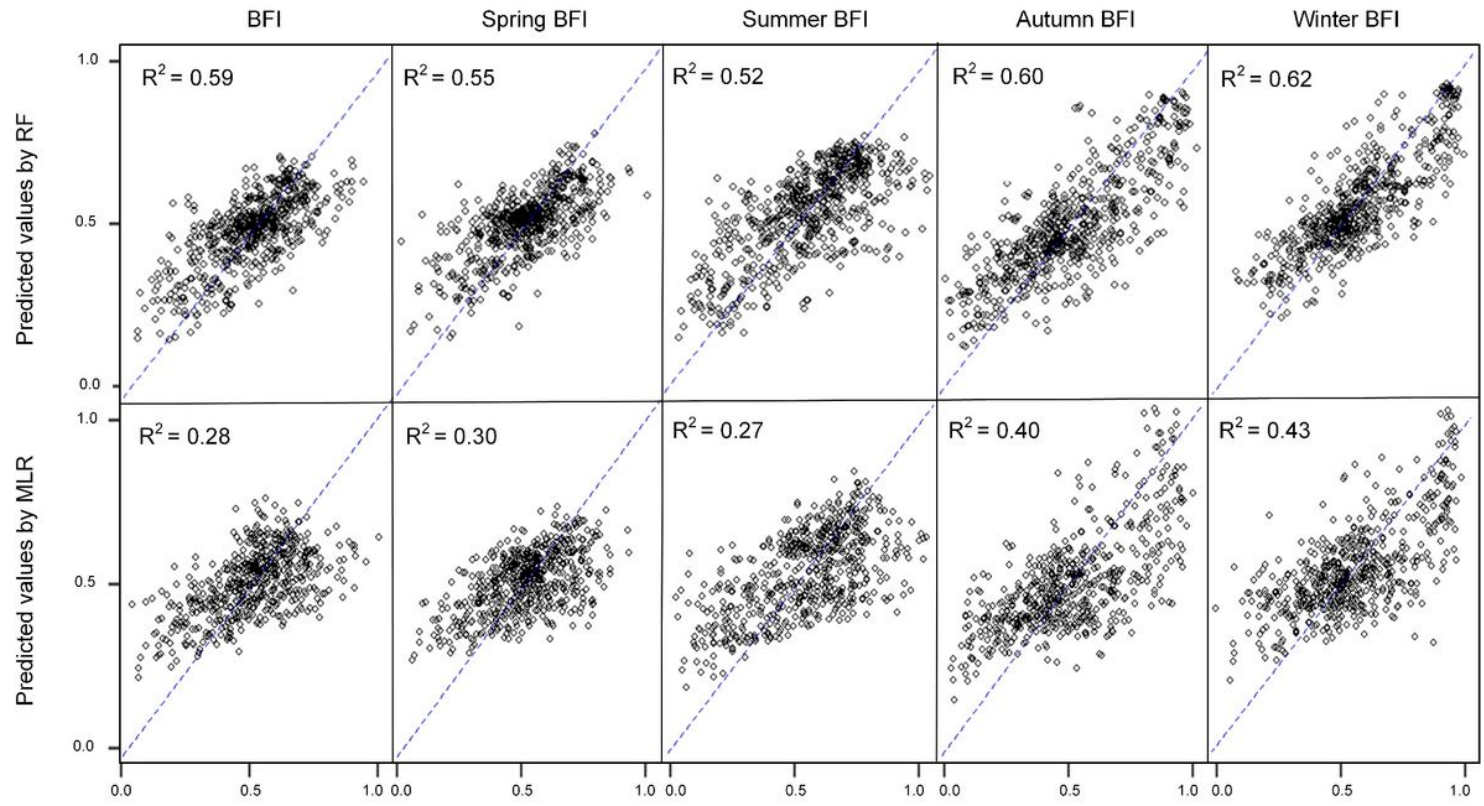

Figure 6

BFI prediction performance by RF (upper Fig.) and MLR (lower Fig.) in 619 catchments across CONUS

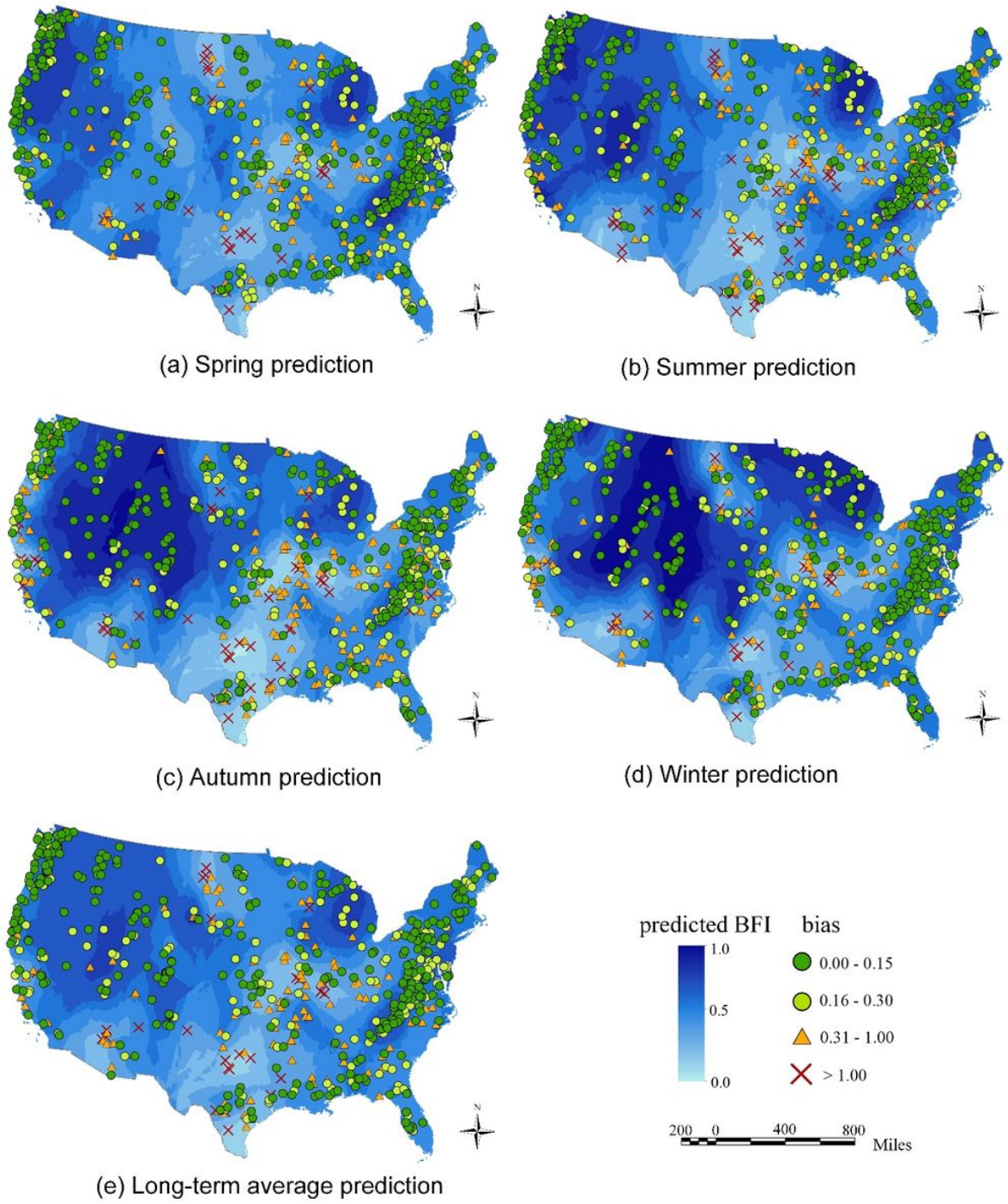

Figure 7

Predicted BFIs and spatial distributions of prediction error

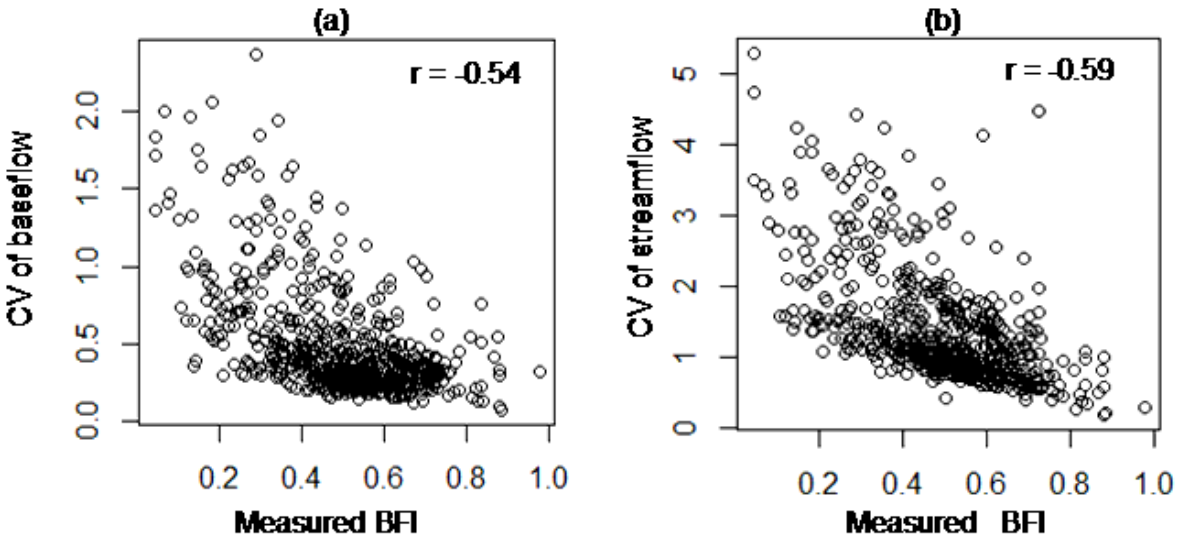

Figure 8

Scatter plot (a) between BFI and CV of baseflow, (b) between BFI and CV of streamflow

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterials.docx