



Research papers

Sensitivity of drought resilience-vulnerability- exposure to hydrologic ratios in contiguous United States

Anoop Valiya Veettil^a, Goutam Konapala^a, Ashok K. Mishra^{a,*}, Hong-Yi Li^b^a Glenn Department of Civil Engineering, Clemson University, Clemson, SC, USA^b Department of Land Resources and Environmental Sciences and Institute on Ecosystems, Montana State University, Bozeman, MT, USA

ARTICLE INFO

Keywords:

Hydrologic ratio
Drought
Random forest
Sensitivity analysis

ABSTRACT

The atmospheric water supply and demand dynamics determine a region's potential water resources. The hydrologic ratios, such as, aridity index, evaporation ratio and runoff coefficients are useful indicators to quantify the atmospheric water dynamics at watershed to regional scales. In this study, we developed a modeling framework using a machine learning approach to predict hydrologic ratios for watersheds located in contiguous United States (CONUS) by utilizing a set of climate, soil, vegetation, and topographic variables. Overall, the proposed modeling framework is able to simulate the hydrologic ratios at watershed scale with a considerable accuracy. The concept of non-parametric elasticity was applied to study the potential influence of the estimated hydrologic ratios on various drought characteristics (resilience, vulnerability, and exposure) for river basins located in CONUS. Spatial sensitivity of drought indicators to hydrologic ratios suggests that an increase in hydrologic ratios may result in augmentation of magnitude of drought indicators in majority of the river basins. Aridity index seems to have higher influence on drought characteristics in comparison to other hydrologic ratios. It was observed that the machine learning approach based on random forests algorithm can efficiently estimate the spatial distribution of hydrologic ratios provided sufficient data is available. In addition to that, the non-parametric based elasticity approach can identify the potential influence of hydrologic ratios on spatial drought characteristics.

1. Introduction

Determining the long term natural water availability is extremely important for domestic, agricultural and industrial sectors to develop policy and decision makings at local to regional scale (Vorosmarty et al., 2000; Gleick, 2003; Biswas, 2004; Veettil and Mishra, 2016). This long term natural availability of water resources is mostly dependent on the regional atmospheric dynamics controlled by precipitation, streamflow and evapotranspiration (Oki and Kanae, 2006; Arnell, 1999; Huntington, 2006). Therefore, spatial variability in these hydro-climate variables might lead to regional alterations in the supply and demand of water resources availability. These changes may influence the resilience, vulnerability and exposure of hydrologic extreme events in any given river basin (Cook et al., 2004). In addition, the degree of influence with which these variables might affect the drought indicators can vary spatially within a river basin (Van loon et al., 2014; Mishra and Singh, 2011). Therefore, quantifying the potential influence of hydro-climatic variables on extreme events are crucial to improve water resources management in a river basin.

The atmospheric water balance can be quantified using several metrics. Among them, the aridity index, evaporation ratio, and runoff coefficient received a lot of interest due to their ability to capture the dynamics of the hydrologic cycle (Schaake et al., 2006). Aridity index (AI) defined as the ratio of precipitation to potential evaporation, which represents the atmospheric potential water availability over atmospheric water demand. This definition of AI is widely accepted for characterizing climate boundaries (Maestre et al., 2012) in addition to investigating the degree of aridity across the world (Nastos et al., 2013). Also, the aridity index was used to assess the effects of climate change on runoff, vegetation and desertification (Sawicz et al., 2011). Therefore, AI serves to identify and locate regions that suffer from available water deficit.

The runoff coefficient (RC) is defined based on the ratio between the annual runoff and the annual rainfall, and it can be a very good indicator to study the degree of moisture recycling as well as water holding capacity in a given area (Savenije, 1996; Sriwongsitanon and Taesombat, 2011). By using RC, it is possible to investigate the role of canopy architecture, leaf characteristics and biomass in controlling the

* Corresponding author.

E-mail address: ashokm@clemson.edu (A.K. Mishra).

runoff within the catchment (Muzylo et al., 2009; Rao et al., 2011; Ferreira et al., 2016). Similar to R-index (Yao, 1974), we calculated the relative evaporation ratio (ER) that is defined as the ratio between actual evapotranspiration (AET) to potential evapotranspiration (PET). In addition, these three indices can formulate useful metrics for hydrologic ratios (Schaake et al., 2006) and it forms the basis for budyko hypothesis, which is a widely used framework to examine energy and water fluxes at watersheds (Budyko, 1974).

In regions where high-quality long-term observations are scarce, these hydrologic ratios can be quantified using land surface models. However, these models, are subject to uncertainties from inputs, simplification, and parameterization (EK et al., 2003; Li et al., 2011). Further, these models can be computationally expensive, often requiring high-performance computing when running models over larger areas, and may limit a comprehensive assessment of uncertainty in forcing data or climate change scenarios (Bosshard et al., 2013; Elsner et al., 2014). In order to overcome such limitations, statistical models which require a limited number of input variables can be used as an alternative tool to investigate spatial characteristics of surface hydro-climate processes (Schneider, 1996; McCabe and Wolock, 2011; Abatzoglou and Ficklin, 2017). Previous studies have established that statistical models have potential to predict spatio-temporal hydrologic processes based on a set of controlling factors (Abatzoglou and Ficklin, 2017; Deshmukh and Singh, 2016; Van Loon and Laaha, 2015). Among them, Random Forest (RF) approach is one of the most popular and powerful machine learning algorithms that are built based on the principle of various combinations of classification and regressions trees (Breiman, 2001). The RF algorithm is a natural and non-linear modeling tool that provides estimates regarding the hierarchy of variables in the classification, and thus it can estimate individual variable's contribution to the spatial distribution of hydrologic ratios. In hydrology, the RF algorithm has been applied to eco-hydrological distribution modeling (Peters et al., 2007), prediction of natural flow regimes (Carlisle et al., 2010), and groundwater mapping (Naghibi et al., 2016). A great deal of theoretical and empirical studies have detailed the advantages of RF, which includes, high forecast accuracy, acceptable tolerance to outliers and noise, and easy avoidance of over-fitting problems. Therefore, RF model can be applied to spatial prediction and for rectifying multi-variable and non-linear issues.

Among the hydrologic extreme events, droughts are known to have impacts on multiple sectors, such as, domestic, agriculture, energy production and fishery (Mishra and Singh, 2010). Droughts usually span over large geographical areas and often last for months to years representing a dominant three-dimensional (latitude, longitude and time) space-time structure unlike other hydrologic extreme events (Lloyd-Hughes, 2012; Konapala and Mishra, 2017). In addition, climate models have projected an increase in future drought severity and duration over continental USA (Wehner et al., 2011; Konapala and Mishra, 2017). Therefore, investigating the potential influence of hydrologic ratios on drought characteristics can further improve our understanding on identifying critical (sensitive) areas leading to superior drought management strategies. As a result, we determine the various characteristics of droughts using the concepts of Resilience, Exposure, Vulnerability and duration. Estimation of these values at a catchment scale can help water resources managers to develop tools for studying catchment's vulnerability to droughts, ability to recover (resilience), frequency and its exposure to drought conditions.

The overall goal of this study is to: (i) develop a random forest model that can spatially predict the hydrologic ratios based on a set of geospatial climate, vegetation, and topographic variables, and (ii) to evaluate the potential influence of estimated hydrological ratios on spatial drought characteristics for river basins located in CONUS. Through this work, we also investigated the dominant factors responsible for the spatial distributions of hydrologic ratios. To achieve the objectives, the data sources and the adopted statistical methodology are explained in Section 2. In Section 3, the various features of the

Table 1

Pearson's Correlation Coefficients calculated between geospatial variables (watershed characteristics) and hydrologic ratios for the 438 MOPEX basins distributed over the CONUS.

Variable	Description	AI	RI	RC
Area	Watershed area	−0.26	−0.30	−0.30
Forest	Percentage of forest land	0.31	0.41	0.43
Shrub	Percentage of shrub land	0.28	0.29	0.33
Porosity	Soil porosity	−0.02	0.05	0.03
NDVI	Annual mean normalized difference vegetation index	0.44	0.60	0.55
A.PCP	Annual mean precipitation	0.80	0.77	0.81
A.PET	Annual mean potential evapotranspiration	−0.63	−0.73	−0.72
STD.DEV.PET	standard deviation of potential evapotranspiration	−0.54	−0.65	−0.66
STD.DEV.PCP	Standard deviation of precipitation	0.14	−0.06	−0.03
Slope	Watershed slope mean	0.41	0.26	0.34
AWC	Available water content	−0.30	−0.21	−0.27

calibrated random forest model and sensitivity of drought characteristics are discussed. Finally, the conclusions are summarized in Section 4.

2. Study area, data and methodology

2.1. Data description

To address our objectives, the datasets are collected from multiple sources. We first describe the data that used for building RF model. Majority of the data for building RF model was obtained from Model Parameter Estimation Experiment (MOPEX) basin. The watersheds considered in MOPEX are considered to have minimum human interference (Schaake et al., 2006). The dataset covers a wide range of climate, soil, and vegetation characteristics for CONUS. The watershed characteristics (Table 1) of 438 MOPEX basins distributed over the CONUS are acquired from the following sources. The mean annual precipitation and temperature data is processed from the National Weather Service River Forecast System (NWSRFS). The NWSRFS uses interpolation method based on inverse distance algorithm of PRISM (Schaake et al., 2006; Daly et al., 1994) data set using the climatology of 1961–1990. Within the MOPEX basins, watershed's physical characteristics is derived using the Digital Elevation Model (DEM) obtained from the National Operational Hydrologic Remote Sensing Center (NOHRSC); vegetation type was obtained from the University of Maryland Land cover database; porosity is derived from STATGO soils information database, and NDVI is derived from MODIS dataset (<https://lpdaac.usgs.gov/>). The spatial distribution of mean annual precipitation, the hydrological ratios, NDVI, and mean annual potential evapotranspiration across the MOPEX basins are illustrated in Fig. 1. More information about the data sources can be found in Schaake et al. (2006). The watershed slope and available water content (AWC) are derived from National Gap Analysis Program and STATGO database, respectively.

The selected variables (Table 1) for the watersheds across CONUS are collected and aggregated from various sources. The Pearson's Correlation Coefficients between geospatial variables (watershed characteristics) and hydrologic ratios (Table 1) for the 438 MOPEX basins distributed over the CONUS suggests possible relationship (positive/negative) between them. The USGS defined watersheds [Hydrologic Unit Code (HUC) 8] and major river basins (HUC 2) are obtained from USGS watershed boundary dataset (<https://nhd.usgs.gov/wbd.html>). The spatial distribution of major river basins is illustrated in Fig. 2. The mean annual precipitation and potential evapotranspiration are obtained and aggregated from PRISM dataset (<http://www.prism.oregonstate.edu/>). The NDVI over all the watersheds of CONUS are

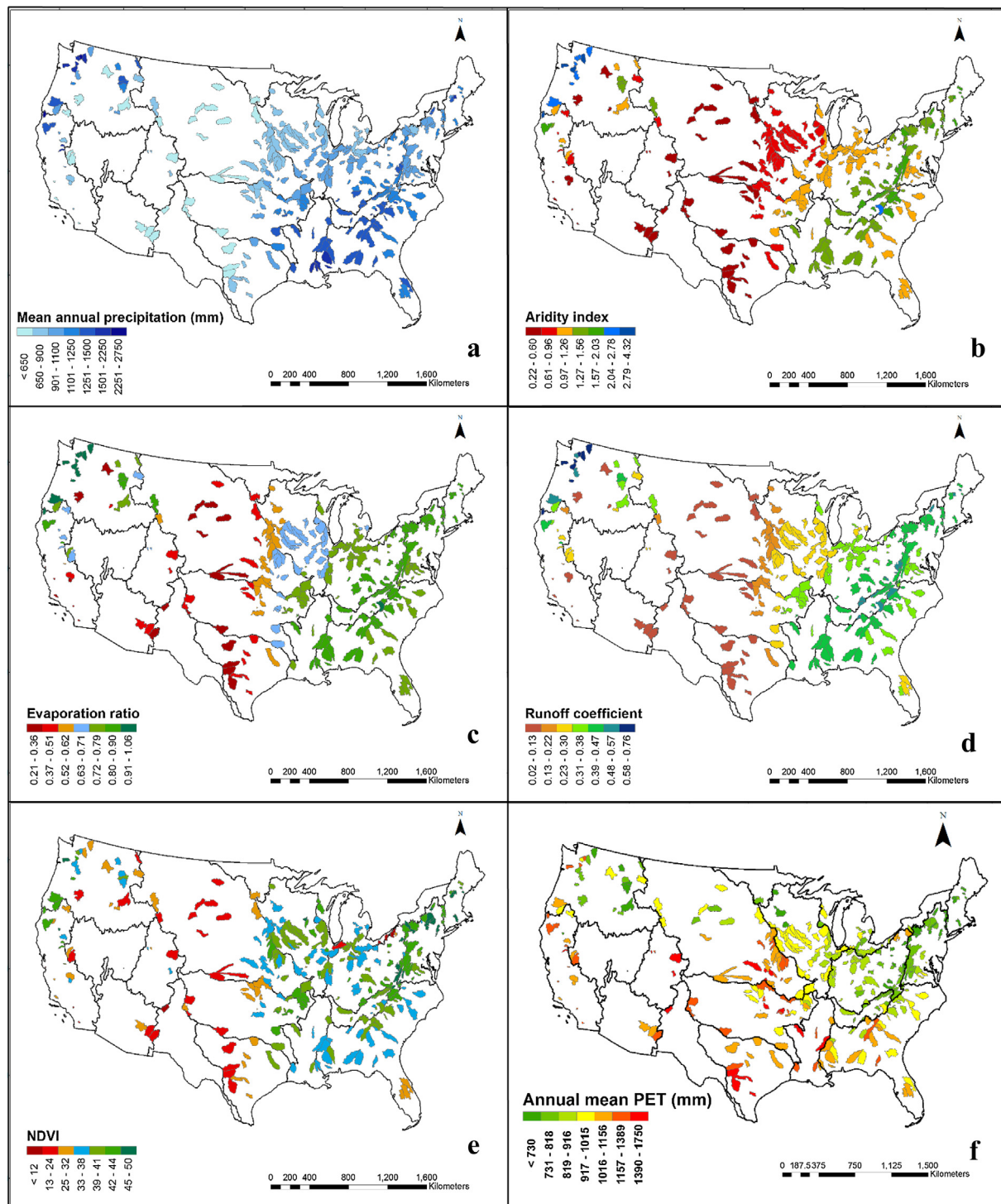


Fig. 1. The spatial distribution of (a) mean of annual precipitation, (b) aridity index, (c) evaporation ratio, and (d) runoff coefficient (e) NDVI, and (f) annual mean PET across the MOPEX basins.

obtained from MODIS NDVI dataset. The spatial distribution of the mean annual precipitation (potential evapotranspiration), and NDVI over the selected watersheds of CONUS are shown in Fig. 3.

For estimating drought characteristics, we used Standardized Precipitation Evapotranspiration Index (SPEI) dataset developed by Vicente-Serrano et al. (2010) and available at SPEIbase v.2.5 (<http://spei.csic.es/database.html>). SPEI is an extension to the widely used Standardized Precipitation Index (SPI). However, unlike, SPI, the SPEI takes into account both precipitation and potential evapotranspiration (PET) and thus captures the impact of increased temperature on water

demand (Vicente-Serrano et al., 2010). In addition to that, SPEI can also be computed at different time scales facilitating its usage for drought impacts on various water resources needs (Mishra and Singh, 2010, 2011). Therefore, in this study we utilize the 3-month SPEI for seasonal scale drought characteristics.

2.2. Random forest (RF) model

The RF modeling approach uses classification and regression trees as building blocks to derive effective prediction models (James et al.,

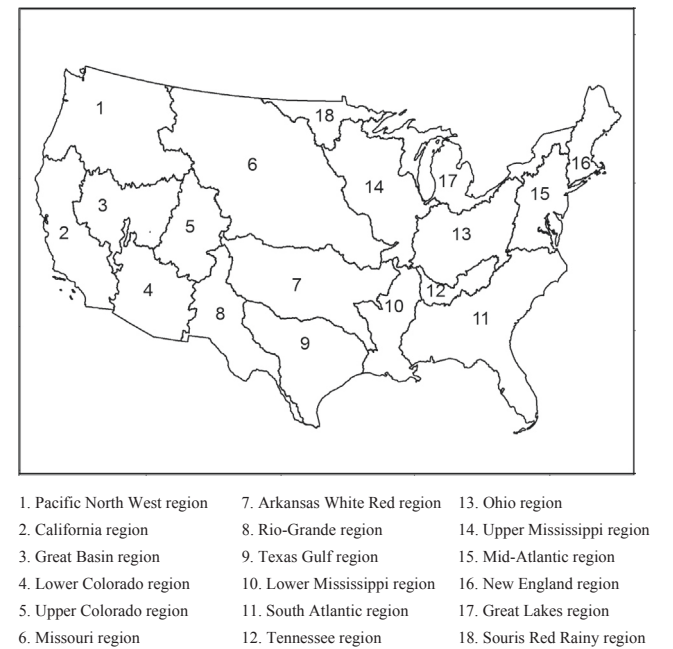


Fig. 2. The location of major river basins in the CONUS. The table provides the name of each major river basins corresponding to the number.

2013). Unlike the other tree-based methods, RF increases diversity among the classification trees by resampling data and randomly changing the predictive variable sets for different tree induction processes. Therefore, the key parameters for RF models are the number of trees and predictors used to determine the split at each node (Vorpahl et al.,

2012). Here the Random Forest (RF) (Breiman, 2001) model was first developed based on the MOPEX watersheds, then the model was applied to predict the hydrological ratios of HUC8 watersheds.

Fig. 4 illustrates the steps used to develop RF prediction model for hydrologic ratios for watersheds in CONUS. As shown in the flowchart both response and predictor variables of MOPEX dataset divided into 2 separate groups i.e., training and testing. The splitting of data set into training and testing data set is based on two conditions (a) does the dataset (training and testing) large enough to yield statistically meaningful results? And (b) does the dataset is a representative of whole? Therefore, split between training and testing may vary in each study. For example, Bachmair et al. (2016) applied RF modeling approach for predicting the drought impacts of each drought indicator by splitting the train and test data to 90:10. Whereas, Rahmati et al. (2016) split the train and test data to 70:30 for mapping the groundwater potential over the Mehran Region located in the northern part of Iran. Similarly, our study also showed an acceptable result for splitting the MPEX datasets to 70% testing data and 30% testing data in terms of correlation coefficient (r) and MSE. Here, the response variables are hydrologic ratios and the predictor variables are geospatial variables that are listed in Table 1. During the training phase, we utilized 500 trees derived from 500 bootstrapped data sets to build random forest model. Then, based on each bootstrapped sample, 500 regression trees are constructed. Split points of these regression trees were chosen from a random subset of all available predictor variables. Then, we averaged the resulting 500 predictions generated from individual regression tree. The individual tree may have higher variance, therefore by averaging these 500 trees likely to reduce the overall variance. Thus, RF modeling approach has potential to improve prediction (simulation) accuracy by combining hundreds of trees into a single procedure (James et al., 2013). Finally, we predicted the hydrologic ratios by applying the RF algorithm and

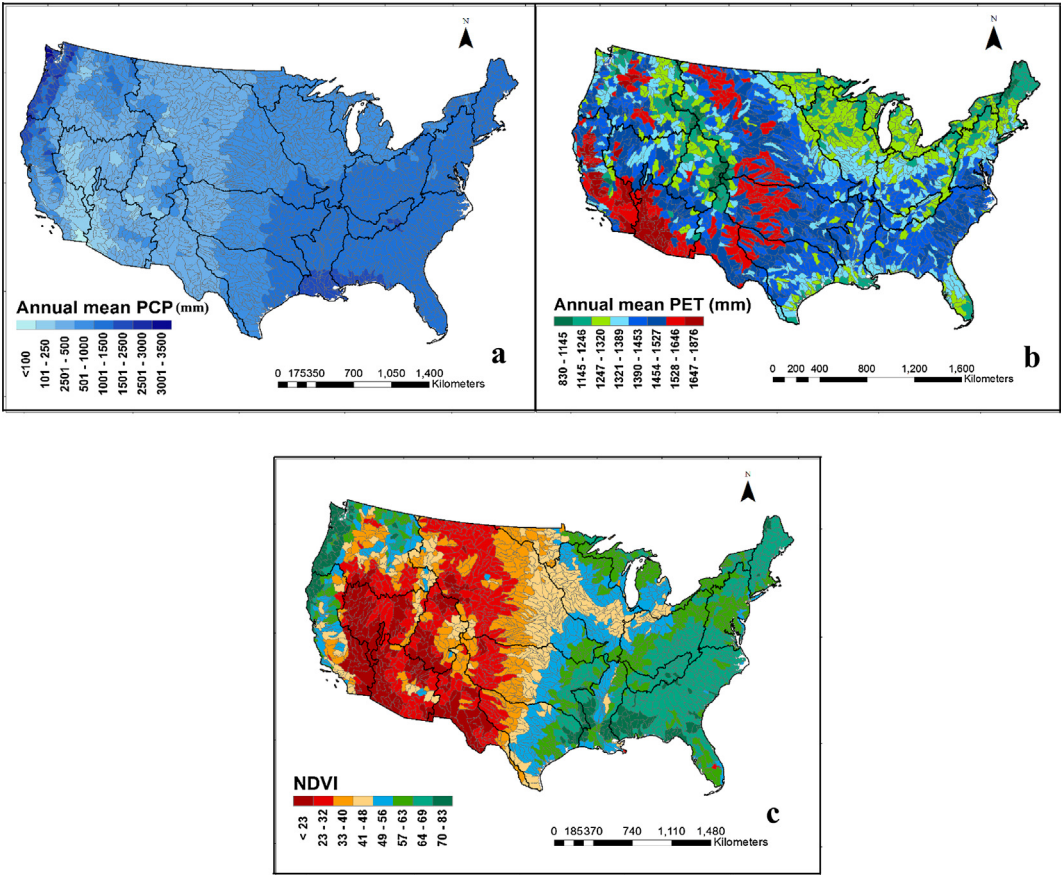


Fig. 3. The spatial distribution of (a) mean annual precipitation, (b) mean annual PET, and (c) NDVI across the HUC-8 watersheds of USA.

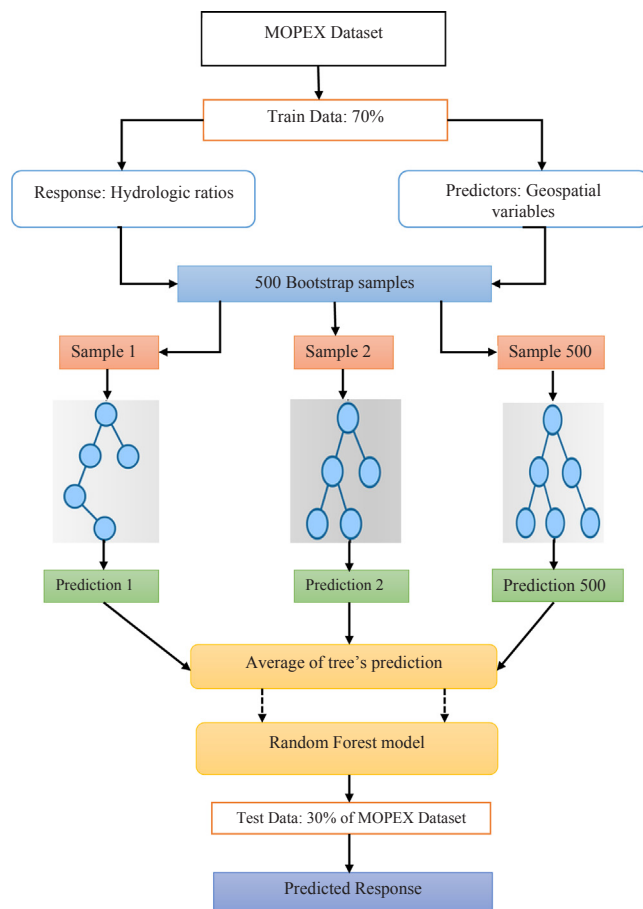


Fig. 4. The exploratory random forest modeling framework used to predict the hydrologic ratios of MOPEX River basin. [Note: 70% of the data is used for training the random forest model and 30% of data used for testing].

validated with 30% data, which was not included in the model development. The random forest model used in this study was built using the “randomForest” package (Liaw and Wiener, 2002) in the R software project (R Core Team, 2014). Subsequently, the model prediction accuracy was evaluated based on mean square error (MSE) (James et al., 2013) and strength of correlation (r) between the predicted and observed hydrological ratios. In the second phase of the analysis, we applied the calibrated RF model to predict hydrological ratios over the

CONUS by using the predictors aggregated at HUC-8 level watershed scale.

2.3. Drought resilience, vulnerability, exposure and frequency

We evaluated the watershed’s drought conditions based on the concepts of resilience, vulnerability, exposure and frequency. These concepts (Hashimoto et al., 1982; Loucks and Van Beek, 2017) are widely used in the water resources systems (Loucks and Van Beek, 2017; Asefa et al., 2014; Brown and Williams, 2015; Ayyub, 2014). In this study, we applied these concepts to characterize droughts in any given watershed. For illustration purpose, we develop a hypothetical example based on SPEI 03 as drought indicator for a period of 24 months (shown in Fig. 6). In this study, we consider that the drought condition likely to be prevalent in a watershed if the SPEI 03 value is below -1 for a period more than 2 months. However, selection of these thresholds can vary depending on the research interest and stakeholders need. Therefore, the hypothetical scenario exhibits two instances of the water shortage ranging for a period of 3 and 5 months. In the following text, we illustrate these drought indicators based on the example provided in Fig. 5.

2.3.1. Drought resilience

Drought Resilience (Re) can be defined as the ability of a watershed to recover from water shortage to water availability state (Hoque et al., 2016; Loucks and Van Beek, 2017; Sadeghi and Hazbavi, 2017). In our case, we defined Re based on a temporal scale as adopted previously by Maity et al. (2012) and Sadeghi and Hazbavi, (2017). As a result, we express drought resilience (Re) as the inverse of average drought duration given by Eq. (1)

$$Re = \frac{M}{\sum_{i=1}^M DD_i} \quad (1)$$

where M is the number of events and DD is the drought duration. In our hypothetical scenario (Fig. 5), we can observe that $M = 2$ and $\sum_{i=1}^M DD_i = 3 + 5$, indicating that Re is 0.4/months. As evident from the formulation, it can be seen that higher is the average drought duration, lesser will be resilient of a watershed to droughts. Therefore, high values of Re indicates that the watersheds are more resilient to drought.

2.3.2. Vulnerability

Drought vulnerability (Vu) can be defined as the average depletion of available water due to drought. In our case, the drought vulnerability is measured based on the magnitude of SPEI during the drought event

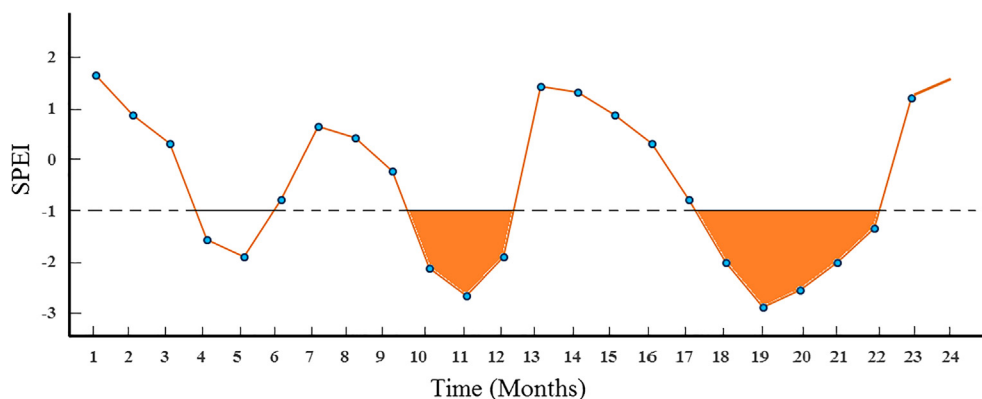


Fig. 5. Hypothetical schematic of SPEI time series used for drought analysis. The drought threshold is shown as dotted lines. Here the drought events are selected using SPEI value ≤ -1 , and for a duration ≥ 3 months.

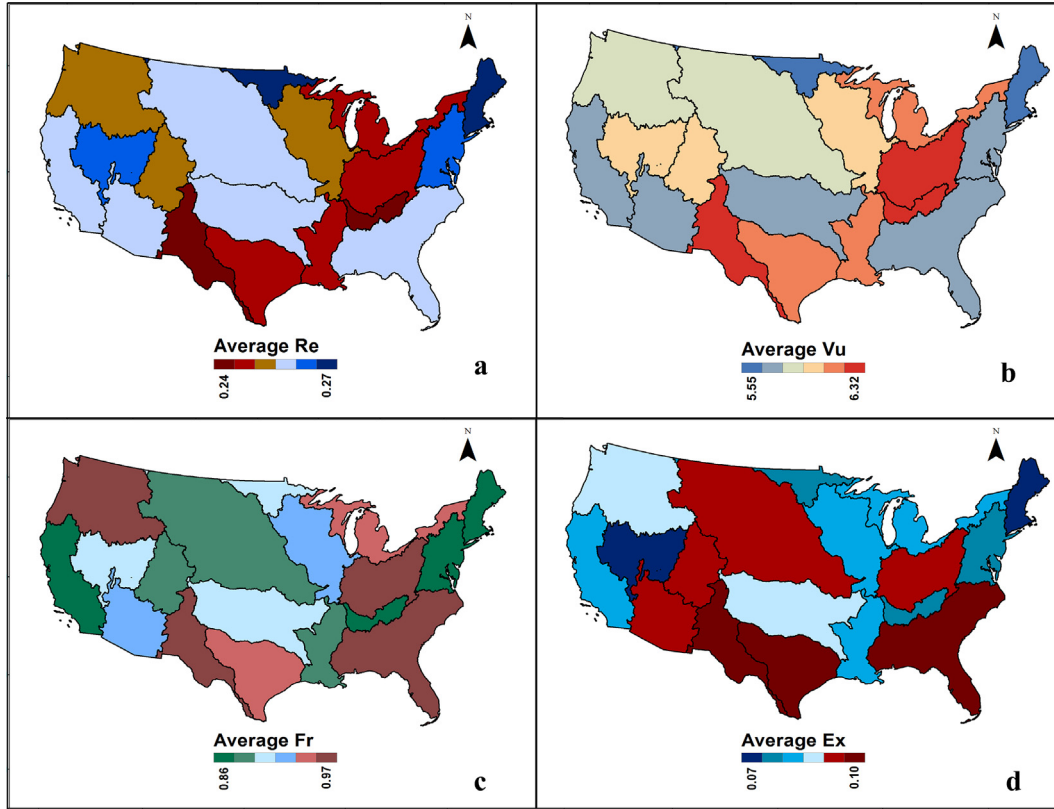


Fig. 6. Spatial distribution of average (a) drought resilience, (b) drought vulnerability, (c) drought frequency and (d) drought exposure across the major river basins of USA.

(Sadeghi and Hazbavi, 2017). Therefore, in the present study we estimated the drought vulnerability of each watershed as (Sadeghi and Hazbavi, 2017):

$$Vu = \frac{\sum_{SPEI < -1} SPEI}{M} \quad (2)$$

In the context of our hypothetical example (Fig. 5), drought vulnerability is given by $[(-2.2) + (-2.6) + (-1.9) + (-2.1) + (-2.9) + (-2.5) + (-2) + (-1.2)]/2$. Higher value of Vu indicates that the watershed exhibits higher degree of vulnerability to droughts.

2.3.3. Exposure

Drought exposure (Ex) can be defined as the relative amount of time the watershed is exposed to drought conditions (Liu et al., 2013). In this case we define this phenomenon based on temporal scale, and as a result, Ex is calculated as the ratio between total time periods (Liu et al., 2013) a watershed is undergoing drought to the total duration of the study period. It can be estimated as,

$$Ex = \frac{\sum_{i=1}^M DD_i}{TD} \quad (3)$$

where, TD represents the total number of months in the study period. Therefore, in context of Fig. 5, drought exposure can be estimated as $(3 + 5)/24$. As evident from the formulation, higher values of Ex represent high degree of drought exposure in the watersheds.

2.3.4. Drought frequency

The drought frequency (Fr) can be defined as the number of occurrences of drought event exceeding a certain threshold per unit time (Blenkinsop and Fowler, 2007; Spinoni et al., 2014). In our case, we expressed drought frequency on a yearly scale given by the total number of qualifying drought events ($SPEI < -1$ and $DD >$

2 months) divided by the total number years as

$$Fr = \frac{M}{n} \quad (4)$$

Where n is the number of years considered in this study and M is the number of drought events with $SPEI03$ value below -1 simultaneously for a period more than 2 months. Therefore, for the case of Fig. 5, drought frequency is estimated as $2/2$ indicating the frequency as a single drought event per year.

In this study, $SPEI 03$ was used as drought index to calculate drought resilience, vulnerability, exposure and frequency for CONUS at a spatial resolution of $0.5^\circ \times 0.5^\circ$ for the period of 1961–1990. We aggregated these gridded values to generate information for individual watersheds. The spatial distribution of Re , Vu , Ex , and Fr for the major river basins are shown in Fig. 6.

2.4. Sensitivity analysis between drought characteristics and hydrologic ratios

To quantify the spatial response of the estimated drought indicators to hydrologic ratios we utilized the concept of non-parametric sensitivity introduced by Sankarasubramanian et al. (2001). This approach is useful to quantify the relative change in one variable may affect the other variable. This sensitivity (Konapala and Mishra, 2016; Ahlblade et al., 2017) can be expressed as

$$\phi = median \left(\frac{y_i^f - \bar{y}^f}{x_i^f - \bar{x}^f} \times \frac{\bar{x}^f}{\bar{y}^f} \right) \quad (5)$$

where $x \in \{AI, ER, RC\}$ and $y \in \{Re, Ex, Vu, Fr\}$; f can be any one of the selected major river basins and i is the i^{th} watershed belonging to the river basin f . Whereas, $\bar{y}$ and $\bar{x}$ represents the spatial mean of x and y variables within the watersheds in a given river basin f . The value ϕ represents the spatial variation in drought characteristics with respect

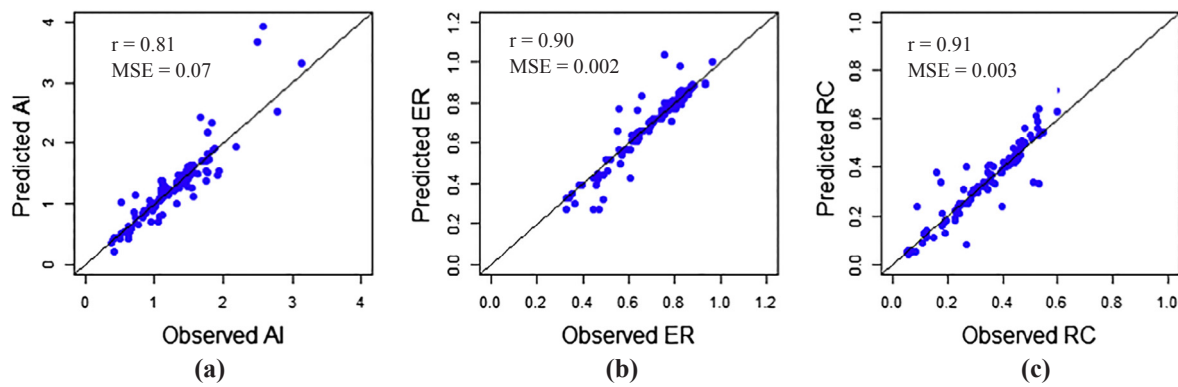


Fig. 7. Scatter plot between observed versus predicted (a) aridity index (AI), (b) evaporation ratio (ER), and (c) runoff coefficient (RC) of the MOPEX river basin.

to the change in hydrologic ratio of watersheds within a river basin. In addition, the ϕ value also indicates the degree of spatial influence of the hydrologic ratios on drought characteristics within a river basin. The higher value of ϕ for a river basin indicates that the drought characteristics of watersheds in the river basin are highly sensitive to the hydrologic ratios. Another advantage of this approach is that it distinguishes between positive and negative sensitivities. For instance, a positive ϕ value in case of aridity index and resilience, indicate that an increase in aridity index may result in increase in resilience over the river basin. Whereas, a negative ϕ value indicates that an increase in aridity index might result in decrease in drought resilience. Therefore, this approach is found to be suitable for assessing the spatial influence of hydrologic ratios on drought properties within a river basin. Then, the statistical significance of ϕ was evaluated with a bootstrap approach as mentioned in Konapala and Mishra (2016) by considering 999 samples. Only the ϕ values with p values less than 0.05 are considered in this analysis.

3. Results

3.1. Evaluation of random forest model

As a first step, we evaluated the performance of random forest (RF) model to predict hydrologic ratios (AI, ER, and RC) over the MOPEX basins. The scatter plot between observed and predicted hydrologic ratios are provided in Fig. 7. The random forest model for predicting AI (RF-AI) have a Pearson correlation value of 0.81 and an MSE of 0.07. Whereas, the RF model for ER (RF-ER) shows a Pearson correlation value of 0.9 and MSE of 0.002; and the RF model for RC (RF-RC) shows a Pearson correlation of 0.91 and MSE of 0.007. Even though all these models were able to replicate the index values with a considerable accuracy, the RF-RC model performs the best among the three models. In case of RF-AI and RF-ER models the annual mean precipitation, annual mean potential evapotranspiration, NDVI, and percentage of forest cover were the most important variables in predicting the AI and ER across the watersheds of USA. Whereas, in case of RF-RC model standard deviation of potential evapotranspiration also played an important role in predicting the RC. In RF-AI model, the AI values greater than 2 are less accurately predicted. Whereas, the AI values from 0.5 to 2 are more accurately predicted. As a result, the model might not accurately represent the humid conditions. In case of RF-ER model, the values which are at the lower end (i.e. 0.3 to 0.6) are less accurately predicted compared to the values in the range of 0.6 to 0.9. Finally, in case of RF-RC model, the predicted values greater than 0.5 witness higher uncertainty.

3.2. Spatial pattern of modeled aridity index across the major river basins

The calibrated RF models are applied to predict the hydrologic

ratios at HUC 8 level watersheds. In general HUC 8 maps the sub-basin level, similar to medium-sized river basins (about 2200 nationwide). Fig. 8(a) illustrates the spatial distribution of aridity index for the CONUS. The UNEP (1992) classified the aridity index climatic zone to hyper arid (< 0.05), arid (0.05–0.20), semi-arid (0.20–0.50), sub-humid (0.50–0.65), and humid (> 0.65) zones (Nastos et al., 2013). Based on our analysis, it was observed that most of the watershed were located in semi-arid, sub-humid or humid, and none of these watersheds depicted either hyper arid or arid climatic zone as classified by the UNEP (1992). The AI values for the eastern river basins were relatively higher and homogeneous in nature. For example, in South-Atlantic River basin the AI values range from 0.9 to 1.5 with a coefficient of variation (CV) of 11%. Where, the mean annual precipitation was also homogeneous throughout the watersheds. The Lower Mississippi River basin located in southeastern CONUS showed minimum CV (5%) in terms of predicted AI values with the values range from 1.05 to 1.6. Whereas, for the Great-Lake River basin located in the northeast of the CONUS exhibited maximum CV (14%). However, unlike the eastern river basins, the central river basins showed more diverse distribution of the AI values. For instance the Missouri River basin, Arkansas White Red region, and Texas-Gulf region showed an AI value range of 0.39–0.91, 0.43–1.29, and 0.53–1.32 respectively.

However, as we go from east to west the indices value further reduced and the spatial pattern of AI becomes more heterogeneous in nature (Sankarasubramanian and Vogel, 2003). Majority of the watersheds in the southwestern United States are classified as ‘dry’, hence their AI is less than 0.65, except few watersheds located in the California River Basin. Unlike the watersheds in the southwestern USA, the Pacific Northwest River basin showed high values of AI possibly due to the frontal weather systems arising from the Pacific Ocean and the Cascade Mountains (Schillinger et al., 2010). It was observed that, within the Pacific Northwest River basin the AI values range from 0.38 to 2.9, with a CV of 72%. This observed variability in AI may be associated with spatial variability in precipitation in the Pacific Northwest River basin (Bracken et al., 2015), where the annual precipitation ranges from 87 mm to 3300 mm.

The increase in temperature leads to increased evapotranspiration and lower AI values (Proedrou et al., 1997; Feidas et al., 2004; Philandras et al., 2008) in many watersheds. The maximum PET was observed in the watersheds located in the southwestern river basins. For instance, California and Lower Colorado River basins have number of watersheds with a higher PET, which may lead the watersheds of these river basins to lower aridity values. However, the spatial distribution of PET was more heterogeneous compared to the precipitation pattern. The mean annual normalized vegetation index (NDVI) is one of the important variables influencing the AI over the major river basins. Similar to the mean annual precipitation, the spatial distribution of NDVI was homogenous across the river basins of eastern USA. We also noticed that, the correlation between NDVI and AI was comparatively less in the

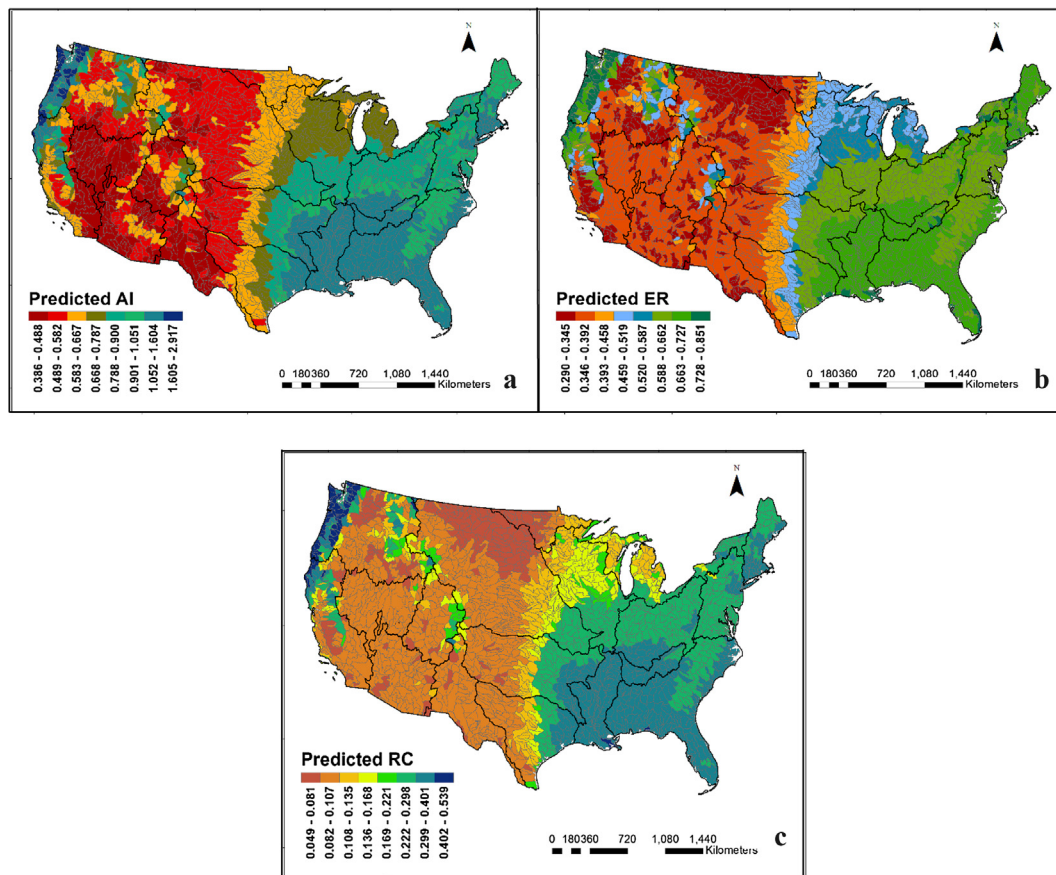


Fig. 8. Spatial pattern of simulated (a) aridity index, (b) evaporation ratio, and (c) runoff coefficient for the CONUS watersheds. Black lines show the boundaries of 18 major river basins.

eastern United States, where the majority of the land cover is dominated by forest cover (NLCD, 2011). Whereas, the NDVI for the western river basins showed high correlation with the aridity values (e.g. Pacific Northwest River basin, California region, Great-Basin region), which suggests that NDVI has major role in controlling the AI in most of the watersheds located in western USA. However, the result from the RF model suggests that the relationship between watershed slope and AWC are weak with AI, indicating that topography and soil characteristics have relatively less role in governing the AI across the watersheds.

3.3. Spatial pattern of modeled evaporation ratio across the major river basins

A higher value of relative evaporation ratio (ER) indicates that the region has optimum water that is necessary for the vegetation. The Fig. 8(b) illustrates the modeled ER based on RF modeling for the CONUS. The ER values vary from 0.33 to 0.86, with a CV of 28% over the watersheds located in CONUS. Similar to the AI, the ER values for the eastern river basins were relatively higher and homogeneous in nature. For example, the ER values of the New-England River basin, Mid and South-Atlantic River basins range from 0.63 to 0.77, 0.56–0.76, and 0.63–0.82 respectively. In addition to that, the ER values typically follows the spatial pattern of mean annual precipitation. Whereas, the river basins located in the central United States exhibits relatively high variability of ER. For instance, the Missouri River Basin, Arkansas white-red region, and Texas-Gulf region showed a CV of 24%, 27%, and 23% respectively. In contrast to AI, the spatial distribution of ER values were more homogenous throughout the southwestern river basins (e.g. Lower Colorado and Rio-Grande region). The southwestern river basins exhibited lower range of ER values. However, unlike the

Lower Colorado and Rio-Grande region, the ER value for California River basin range from 0.3 to 0.75. Similar to the AI, the Pacific Northwest River basin depicted high variation of ER values, with a CV of 33%.

The major contributing variables for estimating ER in the RF model were mean annual precipitation, PET, and vegetation cover. The percentage of forest cover also showed a better relationship with the ER in the random forest modeling. Overall, the combination of minimum precipitation, higher PET, and less vegetation growth may attribute to the lower ER in the southwestern parts of the CONUS. The land cover has a strong influence on ER (Liu et al., 2017), therefore the role of land use pattern in defining ER of a river basin is analyzed. For instance, Tennessee River basin and Ohio River basins located in the eastern USA has forest land cover of 58% and 48% showed relatively less variation of ER throughout the basins. Whereas, for the western USA, where the forest cover is comparatively less, exhibited higher variation in ER. Although, the ER of a region also depends upon crop type, stage of growth, soil moisture, health of plants and cultivation practices (Ayars and Hanson, 2014), we did not include all these variables in RF model due to limitation in available data. Overall, it was observed that RF modeling framework is appropriate for predicting the ER at a regional (or watershed) scale.

3.4. Spatial pattern of modeled runoff coefficient across the major river basins

The spatial distribution of runoff coefficient (RC) across the continental USA is illustrated in Fig. 8(c). Similar to AI and ER, the river basins located in eastern CONUS witnessed a homogenous distribution of RC (Sankarasubramanian and Vogel, 2003), comparatively with a

higher magnitude (Chang et al., 2014) varies from 0.1 to 0.54. However, unlike the other eastern river basins, the Great Lake River basin of northeastern USA exhibited high variation of RC. River basins located in central USA, i.e. Missouri River basin, Arkansas River basin, and Texas- Gulf region showed more diverse pattern of RC across the watersheds. In contrast to the AI values, the spatial pattern of RC was more homogeneous across the southwestern (e.g. Lower Colorado River basin) and Midwestern (e.g. Great-Basin region) river basins. For example, Lower Colorado River basin showed a CV of 8% and Great – Basin region showed a CV of 16%. Whereas, in California River basin the RC varies from 0.07 to 0.52, with a CV of 66%. Similar to the result presented by Chang et al. (2014), we also observed that Pacific Northwest River basin watersheds exhibited the high values of RC among the CONUS watersheds, where the RC was ranging from 0.04 to 0.54 possibly due to the spatial variability in precipitation.

Mean annual precipitation, mean annual PET, and NDVI were the important variables, which control the spatial pattern of runoff coefficient across the CONUS watersheds. As we mentioned, the range of NDVI was comparatively lesser in western United States, but it showed higher correlation with RC. For instance, the correlation between NDVI and runoff coefficient in Pacific Northwest basin was 0.81, which is comparatively higher than the river basins of eastern USA. Therefore, in the random forest modeling the NDVI may have a significant role in predicting the RC over western river basins. Overall, the spatial analysis of AI, ER, and RC showed distinct spatial pattern across the watersheds of CONUS (Chang et al., 2014; Sawicz et al., 2011; Sankarasubramanian and Vogel, 2003). We also noticed that, basins with low hydrologic ratios typically have higher CV (e.g. Pacific Northwest and California region). On the other hand, basins with higher hydrologic ratios depicted low CV (e.g. South-Atlantic Gulf region and Lower Mississippi region).

3.5. Sensitivity of drought indicators to hydrologic ratios

Fig. 9 illustrates the spatial pattern of sensitivity of various drought characteristics to AI in the major river basins of CONUS. The basins in grey color are not statistically significant in nature. Fig. 9(a) indicates sensitivity of drought resilience to AI value. The sensitivity of northeast (Great Lakes region) and north central (Souris Red Rainy region) regions showed a positive sensitivity with Great Lakes region and Souris Red Rainy region exhibiting around 111% and 16% of increase in drought resilience. It indicates that, within the river basin, an increase in AI of watersheds would result in increase of drought resilience of these watersheds. We also noticed that the Great Lakes region has the high positive sensitivity value to the aridity index. Whereas, the Tennessee, Lower and Upper Mississippi River basins showed a decrease in drought resilience with respect to aridity index. The Tennessee River basin showed a decrease of 53% and Lower Mississippi River basin showed a decrease of 34%. In the western part of USA, the only region that is significantly sensitive to AI is Upper Colorado River basin, where it showed a negative sensitivity value of 19%. This indicates that in those river basins, an increase in AI values in watersheds are accompanied by an increase of drought resilience. Also, it is interesting to note that the observed magnitude of positive sensitivity is more than the observed magnitude of negative sensitivity. Therefore, it may indicate that in the basins where humidity increases due to precipitation, drought resilience also increases (Sherwood and Fu, 2014). In case of sensitivity of drought vulnerability to AI (Fig. 9(b)), we can see that all the north eastern river basins exhibit a positive sensitivity except the new England region. Among them, the Great Lakes region has a high sensitivity of around 50% and the Mid-Atlantic region has a low sensitivity of around 11%. California region seems to be more sensitive than all the river basins in the CONUS region. Whereas, the Rio Grande and Souris Red Rainy region exhibit negative sensitivity values of 28%

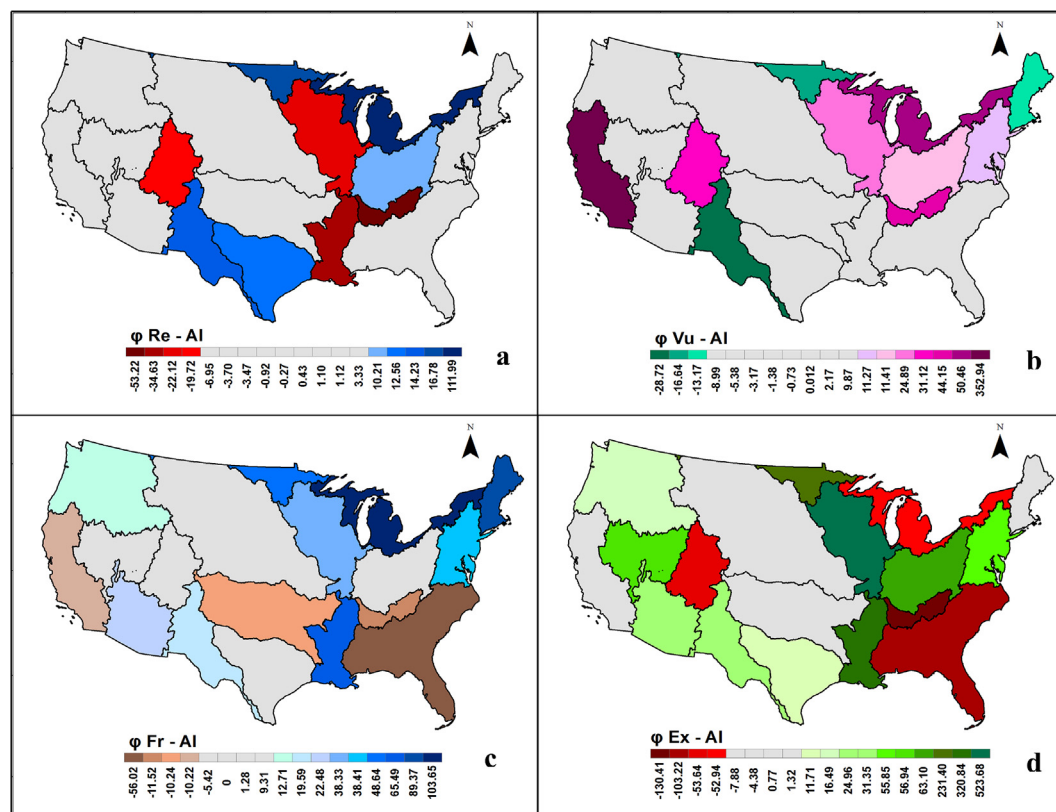


Fig. 9. Spatial distribution of sensitivity of (a) drought resilience, (b) drought vulnerability, (c) drought frequency, and (d) drought exposure to aridity index across the major river basins. The basins in grey are not statistically significant in nature.

and 16% respectively. In case of drought vulnerability too, the observed magnitude of positive sensitivity is more than the observed magnitude of negative sensitivity. Therefore, the similar conclusion of basins where AI value increases prone to increased drought vulnerability may be implied.

In case of sensitivity of drought frequency to AI (Fig. 9(c)), the northeastern river basins are considerably more sensitive than the other regions. The Great Lakes river basin and New-England river basin exhibits the high positive sensitivity, whereas the river basins in southeast showed comparatively low but negative and significant sensitivities. The California River basin (western USA) and Arkansas River basin of central USA also showed a negative sensitivity of almost 10%. The spatial sensitivity of drought frequency also showed that, more number of river basins (13 in total) are sensitive to aridity index. In this case also, we can see that the river basins exhibiting positive sensitivity values are more in number than the negative sensitivity values. Also, as in the case of other drought indicators, in the northeastern region the number of drought events is positively sensitive to aridity index. Among them, the Great Lakes river basin has the highest positive sensitivity of more than 100%. Whereas, the Pacific Northwest region has the least positive sensitivity of 12%. In addition, it is interesting to see that the drought frequency in the Rio Grande and Lower Colorado River basins were moderately sensitive (38% and 22% respectively) to AI. Overall, we can see that the drought indicators are more positively sensitive to AI value than negatively. Finally, in case of sensitivity of drought exposure to AI (Fig. 9(d)), all the eastern river basins except New England are sensitive. Among them, the South-Atlantic and Tennessee River basin exhibits the high negative sensitivity, whereas the river basins in western USA showed comparatively low but positive and significant sensitivities (e.g. Pacific North west, Upper and Lower Colorado River basins). As in the case of *Re*, *Vu*, and *Fr*, Drought Exposure (*Ex*) also showed more number of positively sensitive river basins across the CONUS.

Fig. 10 illustrates the spatial pattern of sensitivity of various drought indicators to ER of watersheds in major river basins of CONUS region. Compared to AI, we can see that in general drought characteristics in more number of river basins are sensitive to ER. Among the observed positive sensitivities, Great Basin region showed the highest sensitivity followed by Great lakes region. Whereas, in case of negative sensitivities, lower Mississippi and Ohio River region has high negative sensitivities and upper and lower Colorado exhibits moderate sensitivities. Overall, it can be observed that in these river basins, an increase in ER would lead to increase in drought resilience. In case of drought vulnerability (Fig. 10(b)), we can see that the Ohio and Lower Mississippi river basins has the higher positive sensitivity followed by Tennessee and Great basin region. Whereas, in case of negative sensitivities, Pacific Northwest river basin has the higher negative magnitude followed by north Atlantic river basin. In addition to that, the Rio Grande and Texas Gulf region has negative sensitivity values. Whereas, in case of frequency, 13 river basins found to be significantly sensitive to ER. However, in the case of sensitivity of drought frequency to ER (Fig. 10(c)), eastern river basins have shown statistically significant sensitivity values. Among them, great lakes river basin has the highest sensitivity value. Whereas, the least sensitivity is observed in pacific northwestern river basin. Finally, the sensitivity of ER (Fig. 10(d)) to drought exposure (*Ex*) has relatively higher number of river basins (16 out of 18 river basins). Among them, the north eastern river basins witness higher magnitudes of sensitivities (both positive and negative) in comparison to the western river basins. Among them, the Ohio River basin and upper Mississippi river basin exhibit higher positive sensitivity values followed by the lower Mississippi region. Whereas the lower Atlantic and Tennessee River basin exhibit higher negative sensitivity values.

Fig. 11 illustrates the spatial pattern of sensitivity of various drought indicators to RC of watersheds in major river basins of CONUS. In case of drought resilience (Fig. 11(a)), Great Lakes river basin is the

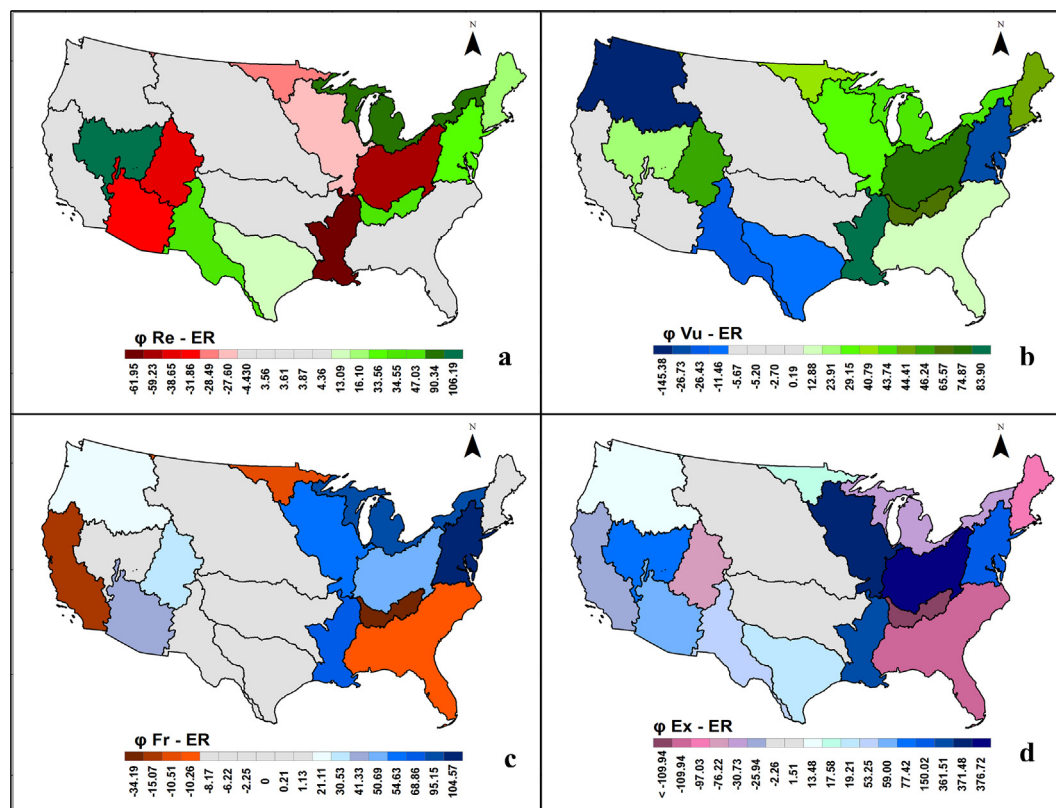


Fig. 10. Spatial distribution of sensitivity of (a) drought resilience, (b) drought vulnerability, (c) drought frequency, and (d) drought exposure to evaporation ratio across the major river basins. The basins in grey are not statistically significant in nature.

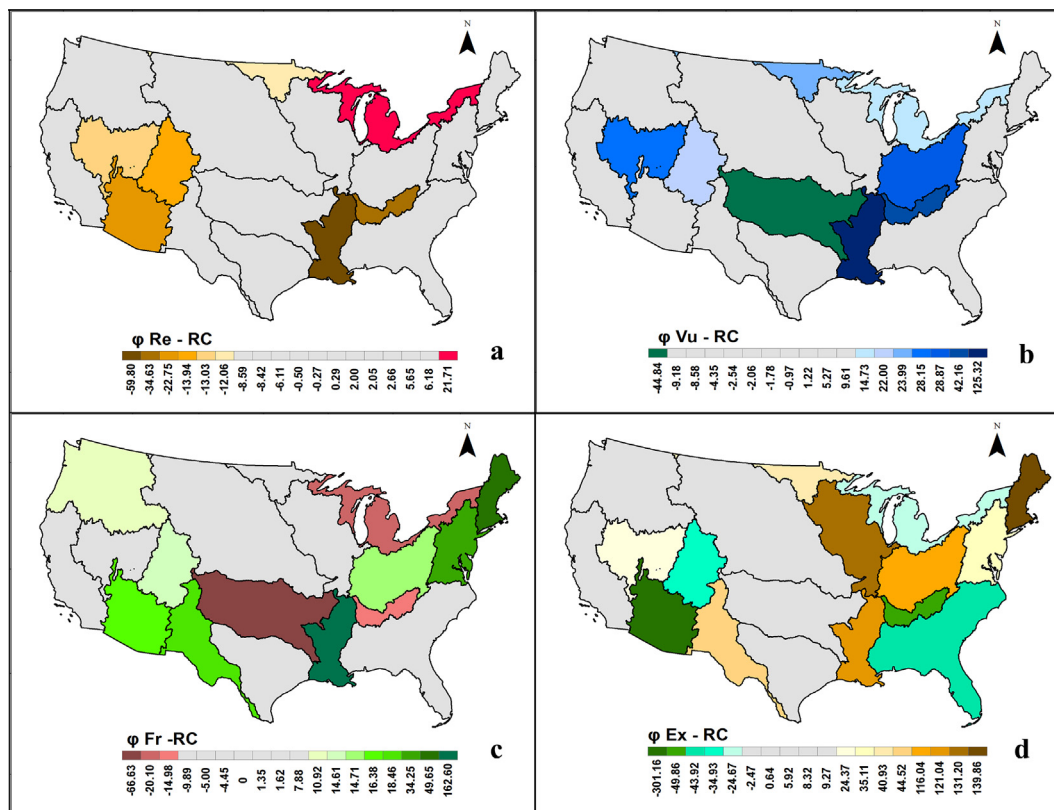


Fig. 11. Spatial distribution of sensitivity of (a) drought resilience, (b) drought vulnerability, (c) drought frequency, and (d) drought exposure to runoff coefficient across the major river basins. The basins in grey are not statistically significant in nature.

only basin exhibiting positive sensitivity to RC. Whereas, the lower Mississippi and Tennessee River basins are negatively sensitive. In the western CONUS region, the upper and lower Colorado region along with the great river basin also exhibit negative sensitivity. In case of vulnerability (Fig. 11(b)), the Arkansas River basin shows a negative sensitivity to RC. Whereas, the lower Mississippi, Ohio and Tennessee River basins exhibit positive sensitivity followed by upper Colorado and great basins. But, the great lakes region is shown to have less magnitude positive sensitivity. In case of frequency (Fig. 11(c)), overall 11 river basins have shown sensitivity to RC including Arkansas and Great lakes river basins exhibiting negative sensitive values. Whereas, the north eastern river basins along with lower Mississippi river basin have high positive sensitivity values. In case of drought exposure (Fig. 11(d)), the northeastern river basins witness positive sensitivity with New England River basin showing the higher sensitivity. Whereas, in southern Atlantic and Tennessee River basin have positive sensitivities. As a result, in general we can say that an increase in hydrologic ratios would cause an increase in the magnitude of drought indicators; however, the spatial sensitivities are not as prominent as in the case of temporal sensitivities.

In general aridity Index is calculated as a ratio of the long-term average annual precipitation to the long-term average annual evapotranspiration. Whereas, drought event is characterized by abnormally dry weather conditions, sufficiently long enough to cause a serious hydrological imbalance for a specific geographic location. A drought event may extend for a season, a year or several years (Mishra and Singh, 2010; Schneider, 1996). This indicates that, aridity is permanent, while drought is temporary. In other words, while the increase in aridity (less humid) may refer to the background climatology of the geographic location, however it may not necessarily signify drought (Sherwood and Fu, 2014; Li et al., 2017; Mukherjee et al., 2018). In addition, the less humid area may have less influence of aridity on water availability (Gudmundsson et al., 2016). Therefore, a change in

the aridity index may cause either positive or negative sensitivity over the river basins considered in this study. Also, Zarch et al. (2015) suggested that, for a long term record of datasets an increasing trend of precipitation and potential evapotranspiration may lead to a decrease in aridity index. In the present study we quantify drought across the continental USA by using SPEI 3 (accumulation period of 3 months). The temporal scale of aridity index is considered as the long term annual average, and temporal scale of drought is considered as the long term average of SPEI 3. This may lead the correlation between the SPEI drought and hydrologic ratios over the river basin to lesser values in our analysis. In addition, the snow cover of the watersheds are not considered in this study. The snow cover can form potential long-term moisture storage in a watershed in the form of snow/ice (Arora, 2002). This may influence the assumption that the hydrologic ratios (AI, ER, and RC) are prominently controlled by the evapotranspiration, precipitation, and surface runoff especially in the western and northern watersheds.

This analysis is based on the natural environment, therefore any discrepancy can be attributed to the neglecting human factors in the analysis. It was observed that the increase in hydrologic ratios may cause an increase in the severity of drought indicators in majority of the river basins. Among the considered hydrologic ratios, Relative evaporation ratio was found out to be more sensitive in influencing the drought indicators of most of the basin. Drought indicators of northeast river basins especially the Great-Lake region is more prone to sensitivity due to the spatial change in hydrologic ratios. Whereas, the Missouri River basin seems to be least sensitive drought region to all the hydrologic ratios. The magnitude of drought vulnerability, frequency, and exposure showed a positive relation with the runoff coefficient across the majority of the river basins, indicating that drought severity in watersheds of these river basins are increasing with runoff coefficient. Similarly, in case of aridity index, majority of the river basins showed positive relation with the drought indicators representing that

precipitation may have more control over drought indicators than potential evapotranspiration.

4. Conclusion

In this study, we utilize Random forest model to predict the spatial pattern of hydrologic ratios, which includes, Aridity Index (AI), Relative Evaporation Ratio (ER), and Runoff Coefficient (RC) for the watersheds located in Continental United States. The developed statistical modeling framework incorporates a set of geospatial climate, soil, vegetation, and topographic variables for predicting these hydrologic ratios. Conceptually, our results agree with the previous research that investigated the spatial distribution of long term natural water availability over the CONUS (Chang et al., 2014). Moreover, the variables used for modeling the hydrologic ratios are widely available, therefore the proposed statistical model can be expanded to any part of the world to investigate and improve the quantification of water availability and related water scarcity. However, there is room for further enhancement, for instance, (i) anthropogenic interventions (e.g. reservoir operation, irrigation water use) are excluded for explaining the hydrologic ratios. The addition of such anthropogenic variables can significantly improve prediction of runoff coefficient; (ii) quantification of groundwater contribution to the hydrologic ratio might improve the model prediction; and (iii) the proposed model may have limited potential to capture hydrologic ratios in regions with higher landscape disturbance such as agricultural land as well as a result of increase in urban sprawl (Abatzoglou and Ficklin, 2017; Hamel and Guswa, 2015). Overall, the proposed random forest prediction framework can be used for analyzing the distribution and variation of hydrologic ratios within the major river basins of CONUS. The following conclusions can be drawn from this study:

- (a) The watersheds located in the Pacific Northwest River basin showed maximum value of aridity index. However, within the Pacific Northwest River Basin the aridity index values can vary from 0.38 to 2.9, with a coefficient of variation of 72%.
- (b) Important variables used in RF model for quantifying the relative evaporation ratio includes, mean annual precipitation, mean annual PET, and vegetation cover. Additionally, the combination of minimum precipitation, higher PET, and less vegetation cover may attribute to low relative evaporation ratio. The spatial distribution of evaporation ratio was more homogenous throughout the southwestern river basins (e.g. Lower Colorado and Rio-Grande region).
- (c) The spatial pattern of runoff coefficient was homogeneous across the eastern (e.g. Mid Atlantic, Tennessee River basins), south-western (e.g. Lower Colorado River basin), and mid-western (e.g. Great Basin region) river basins. Moreover, the influence of vegetation (NDVI) on runoff coefficient was clearly visible in the western river basins of CONUS (e.g. Pacific Northwest River basin, California River basin).
- (d) Overall, the spatial analysis of hydrologic ratios showed distinct spatial pattern across the watersheds of the CONUS. Additionally, basins with low hydrologic ratios typically have high CV (e.g. Pacific Northwest and California region) and on the other hand, basins with high hydrologic ratios depicted low CV (e.g. South-Atlantic Gulf region and Lower Mississippi region).
- (e) The sensitivity of drought indicators to the hydrologic ratios for the 18 major river basins was also investigated. The sensitivity analysis can inform the change in drought indicators such as resilience, vulnerability, frequency, and exposure, with respect to the spatial change in hydrologic ratios. It was observed that the hydrologic ratios are considerably sensitive to drought characteristics in majority of the river basins. Among the considered hydrologic ratios, Relative evaporation ratio seems to be more sensitive to influence the drought indicators of most of the basin. Drought indicators of northeast river basins especially the Great-Lake region is more

sensitive to the spatial change in hydrologic ratios. Whereas, the Missouri River basin seems to be least sensitive to all the hydrologic ratios. The addition of anthropogenic factors (Wan et al., 2017, 2018) such as, land use change, water demand, and reservoir operation may improve the analysis, specifically for evaluating the response of drought to hydrologic ratios. For instance, changes in land use and land cover affect local and regional climate process and related hydrologic ratios. It may lead to variation of sensitivity of drought to hydrologic ratios within a river basin. Future investigation is necessary for the better understanding of the effect of climate change projection on hydrologic ratios and related sensitivity on drought indicators, and consideration of groundwater and snow related variables may improve the quantification of hydrologic ratios across the watersheds of continental USA.

Acknowledgement

This study was supported by the National Science Foundation (NSF) award 1653841. H. Li was supported by the United States Department of Energy, Office of Science, as part of research in Multi-Sector Dynamics, Earth and Environmental System Modeling Program. We very much appreciate editor and three reviewer's valuable comment that helped us to improve the manuscript.

References

- Abatzoglou, J.T., Ficklin, D.L., 2017. Climatic and physiographic controls of spatial variability in surface water balance over the contiguous United States using the Budyko relationship. *Water Resour. Res.* 53, 7630–7643.
- Ahiablame, L., Sheshukov, A.Y., Rahmani, V., Moriasi, D., 2017. Annual baseflow variations as influenced by climate variability and agricultural land use change in the Missouri River Basin. *J. Hydrol.* 551, 188–202.
- Arnell, N.W., 1999. Climate change and global water resources. *Global Environ. Change* 9, S31–S49.
- Arora, V.K., 2002. The use of the aridity index to assess climate change effect on annual runoff. *J. Hydrol.* 265, 164–177.
- Asefa, T., Clayton, J., Adams, A., Anderson, D., 2014. Performance evaluation of a water resources system under varying climatic conditions: reliability, resilience, vulnerability and beyond. *J. Hydrol.* 508, 53–65.
- Ayars, J.E., Hanson, B.R., 2014. Integrated Irrigation and Drainage Water Management. Salinity and Drainage in San Joaquin Valley, California, Springer, pp. 249–276.
- Ayyub, B.M., 2014. Systems resilience for multihazard environments: Definition, metrics, and valuation for decision making. *Risk Anal.* 34, 340–355.
- Bachmair, S., Svensson, C., Hannaford, J., Barker, L., Stahl, K., 2016. A quantitative analysis to objectively appraise drought indicators and model drought impacts. *Hydrol. Earth Syst. Sci.* 20, 2589.
- Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., et al., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resour. Res.* 49, 1523–1536.
- Bracken, C., Rajagopalan, B., Alexander, M., Gangopadhyay, S., 2015. Spatial variability of seasonal extreme precipitation in the western United States. *J. Geophys. Res.: Atmos.* 120, 4522–4533.
- Brown, E.D., Williams, B.K., 2015. Resilience and resource management. *Environ. Manage.* 56, 1416–1427.
- Biswas, A.K., 2004. Integrated water resources management: a reassessment: a water forum contribution. *Water Int.* 29 (2), 248–256.
- Blenkinsop, S., Fowler, H., 2007. Changes in drought frequency, severity and duration for the British Isles projected by the PRUDENCE regional climate models. *J. Hydrol.* 342, 50–71.
- Breiman, L., 2001. Random forests. *Mach. Learning.* 45, 5–32.
- Budyko, M., 1974. *Climate and Life*, 508 pp.
- Carlisle, D.M., Falcone, J., Wolock, D.M., Meador, M.R., Norris, R.H., 2010. Predicting the natural flow regime: models for assessing hydrological alteration in streams. *River Res. Appl.* 26 (2), 118–136.
- Chang, H., Johnson, G., Hinkley, T., Jung, I., 2014. Spatial analysis of annual runoff ratios and their variability across the contiguous US. *J. Hydrol.* 511, 387–402.
- Cook, E.R., Woodhouse, C.A., Eakin, C.M., Meko, D.M., Stahle, D.W., 2004. Long-term aridity changes in the western United States. *Science* 306, 1015–1018.
- Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.* 33, 140–158.
- Deshmukh, A., Singh, R., 2016. Physio-climatic controls on vulnerability of watersheds to climate and land use change across the US. *Water Resour. Res.* 52, 8775–8793.
- Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.: Atmos.* 108.
- Elsner, M.M., Gangopadhyay, S., Pruitt, T., Brekke, L.D., Mizukami, N., Clark, M.P., 2014.

- How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations? *J. Hydrometeorol.* 15, 1384–1403.
- Feidas, H., Makrogiannis, T., Bora-Senta, E., 2004. Trend analysis of air temperature time series in Greece and their relationship with circulation using surface and satellite data: 1955–2001. *Theor. Appl. Climatol.* 79, 185–208.
- Ferreira, C., Walsh, R., Shakesby, R., Keizer, J., Soares, D., González-Pelayo, O., et al., 2016. Differences in overland flow, hydrophobicity and soil moisture dynamics between Mediterranean woodland types in a peri-urban catchment in Portugal. *J. Hydrol.* 533, 473–485.
- Gleick, P.H., 2003. Water use. *Ann. Rev. Environ. Resour.* 28, 275–314.
- Gudmundsson, L., Greve, P., Seneviratne, S.I., 2016. The sensitivity of water availability to changes in the aridity index and other factors—A probabilistic analysis in the Budyko space. *Geophys. Res. Lett.* 43, 6985–6994.
- Hamel, P., Guswa, A.J., 2015. Uncertainty analysis of a spatially explicit annual water-balance model: case study of the Cape Fear basin, North Carolina. *Hydrol. Earth Syst. Sci.* 19, 839–853.
- Hashimoto, T., Stedinger, J.R., Loucks, D.P., 1982. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resour. Res.* 18, 14–20.
- Hoque, Y.M., Tripathi, S., Hantush, M.M., Govindaraju, R.S., 2016. Aggregate measures of watershed health from reconstructed water quality data with uncertainty. *J. Environ. Qual.* 45, 709–719.
- Huntington, T.G., 2006. Evidence for intensification of the global water cycle: review and synthesis. *J. Hydrol.* 319 (1), 83–95.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*. Springer.
- Konapala, G., Mishra, A.K., 2016. Three-parameter-based streamflow elasticity model: application to MOPEX basins in the USA at annual and seasonal scales. *Hydrol. Earth Syst. Sci.* 20, 2545–2556.
- Konapala, G., Mishra, A.K., 2017. Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *J. Hydrol.* 555, 600–620.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News.* 2, 18–22.
- Li, H., Huang, M., Wigmosta, M.S., Ke, Y., Coleman, A.M., Leung, L.R., et al., 2011. Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed. *J. Geophys. Res.: Atmos.* 116.
- Li, Y., Feng, A., Liu, W., Ma, X., Dong, G., 2017. Variation of aridity index and the role of climate variables in the Southwest China. *Water* 9, 743.
- Liu, C., Sun, G., McNulty, S.G., Noormets, A., Yuan, F., 2017. Environmental controls on seasonal ecosystem evapotranspiration/potential evapotranspiration ratio as determined by the global eddy flux measurements. *Hydrol. Earth Syst. Sci.* 21, 311.
- Liu, X., Wang, Y., Peng, J., Braimoh, A.K., Yin, H., 2013. Assessing vulnerability to drought based on exposure, sensitivity and adaptive capacity: a case study in middle Inner Mongolia of China. *Chin. Geogr. Sci.* 23, 13–25.
- Lloyd-Hughes, B., 2012. A spatio-temporal structure-based approach to drought characterisation. *Int. J. Climatol.* 32, 406–418.
- Loucks, D.P., Van Beek, E., 2017. *Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications*. Springer.
- Maestre, F.T., Salguero-Gomez, R., Quero, J.L., 2012. It is getting hotter in here: determining and projecting the impacts of global environmental change on drylands. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 3062–3075.
- Maity, R., Sharma, A., Nagesh Kumar, D., Chanda, K., 2012. Characterizing drought using the reliability-resilience-vulnerability concept. *J. Hydrol. Eng.* 18, 859–869.
- McCabe, G.J., Wolock, D.M., 2011. Independent effects of temperature and precipitation on modeled runoff in the conterminous United States. *Water Resour. Res.* 47.
- Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. *J. Hydrology.* 391, 202–216.
- Mishra, A.K., Singh, V.P., 2011. Drought modeling—a review. *J. Hydrol.* 403, 157–175.
- Mukherjee, S., Mishra, A., Trenberth, K.E., 2018. Climate change and drought: a perspective on drought indices. *Curr. Clim. Change Rep.* 1–19.
- Muzylo, A., Llorens, P., Valente, F., Keizer, J., Domingo, F., Gash, J., 2009. A review of rainfall interception modelling. *J. Hydrol.* 370, 191–206.
- Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 188 (1), 44.
- Nastos, P.T., Politi, N., Kapsomenakis, J., 2013. Spatial and temporal variability of the aridity index in Greece. *Atmos. Res.* 119, 140–152.
- National Land Cover Database 2011 (NLCD 2011) < <https://www.mrlc.gov/nlcd2011.php> > .
- Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. *Science* 313 (5790), 1068–1072.
- Peters, J., De Baets, B., Verhoest, N.E., Samson, R., Degroove, S., De Becker, P., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modelling. *Ecol. Modell.* 207 (2), 304–318.
- Philandras, C.M., Nastos, P.T., Repapis, C.C., 2008. Air temperature variability and trends over Greece. *Global Nest J.* 10 (2), 273–285.
- Proedrou, M., Theoharatos, G., Cartalis, C., 1997. Variations and trends in annual and seasonal air temperature in Greece determined from ground and satellite measurements. *Theor. Appl. Climatol.* 57, 65–78.
- Rahmati, O., Pourghasemi, H.R., Melesse, A.M., 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region. Iran. *Catena.* 137, 360–372.
- Rao, L., Sun, G., Ford, C., Vose, J., 2011. Modeling potential evapotranspiration of two forested watersheds in the southern Appalachians. *Trans. ASABE* 54, 2067–2078.
- Sadeghi, S.H., Hazbavi, Z., 2017. Spatiotemporal variation of watershed health propensity through reliability-resilience-vulnerability based drought index (case study: Shazand Watershed in Iran). *Sci. Total Environ.* 587, 168–176.
- Sankarasubramanian, A., Vogel, R.M., Limbrunner, J.F., 2001. Climate elasticity of streamflow in the United States. *Water Resour. Res.* 37, 1771–1781.
- Sankarasubramanian, A., Vogel, R.M., 2003. Hydroclimatology of the Continental United States. *Geophys. Res. Lett.* 30.
- Savenije, H.H., 1996. The runoff coefficient as the key to moisture recycling. *J. Hydrol.* 176, 219–225.
- Savicz, K., Wagener, T., Sivapalan, M., Troch, P., Carrillo, G., 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrol. Earth Syst. Sci.* 15, 2895–2911.
- Schaake, J., Cong, S., Duan, Q., 2006. US MOPEX data set < <https://e-reports-ext.llnl.gov/pdf/333681.pdf> > .
- Schneider, S.H. (Ed.), 1996. *Encyclopaedia of Climate and Weather*. Oxford University Press, New York.
- Schilling, W.F., Papendick, R.I., McCool, D.K., 2010. Soil and water challenges for Pacific Northwest agriculture. *Soil and Water Conservation Advances in the United States. SSSA Spec. Publ.* 60, 47–79.
- Sherwood, S., Fu, Q., 2014. Climate change. A drier future? *Science* 343, 737–739.
- Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., Vogt, J., 2014. World drought frequency, duration, and severity for 1951–2010. *Int. J. Climatol.* 34, 2792–2804.
- Sriwongsitanon, N., Taesombat, W., 2011. Effects of land cover on runoff coefficient. *J. Hydrol.* 410, 226–238.
- Team, R.C., 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- UNEP, 1992. *World Atlas of Desertification*. Edward Arnold, London.
- Van Loon, A., Laaha, G., 2015. Hydrological drought severity explained by climate and catchment characteristics. *J. Hydrol.* 526, 3–14.
- Van Loon, A., Tjeldeman, E., Wanders, N., Van Lanen, H., Teuling, A., Uijlenhoet, R., 2014. How climate seasonality modifies drought duration and deficit. *J. Geophys. Res.: Atmos.* 119, 4640–4656.
- Veetil, A.V., Mishra, A.K., 2016. Water security assessment using blue and green water footprint concepts. *J. Hydrol.* 542, 589–602.
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* 23, 1696–1718.
- Vorosmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. *Science* 289, 284–288.
- Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B., 2012. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* 239, 27–39.
- Wan, W., Zhao, J., Li, H., Mishra, A., Ruby Leung, L., Hejazi, M., et al., 2017. Hydrological drought in the Anthropocene: impacts of local water extraction and reservoir regulation in the US. *J. Geophys. Res.: Atmos.* 122.
- Wan, W., Zhao, J., Li, H., Mishra, A., Hejazi, M., et al., 2018. A holistic view of water management impacts on future droughts: a global multimodel analysis. *J. Geophys. Res.: Atmos.* <https://doi.org/10.1029/2017JD027825>.
- Wehner, M., Easterling, D.R., Lawrimore, J.H., Heim Jr, R.R., Vose, R.S., Santer, B.D., 2011. Projections of future drought in the continental United States and Mexico. *J. Hydrometeorol.* 12, 1359–1377.
- Yao, A.Y., 1974. Agricultural potential estimated from the ratio of actual to potential evapotranspiration. *Agric. Meteorol.* 13, 405–417.
- Zarch, M.A.A., Sivakumar, B., Sharma, A., 2015. Assessment of global aridity change. *J. Hydrol.* 520, 300–313.