https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Predicting streamflow drought in the conterminous United States using machine learning and a donor-gage approach, 1982-2020

Aaron Heldmyer[1], Roy Sando[2], Caelan Simeone[3], Michael Wieczorek[4], Scott Hamshaw[5], Phillip Goodling[5], Ryan McShane[1], Jeremy Diaz[5], David Watkins[5], Bryce Pulver[6], Apoorva Shastry[7], Konrad Hafen[8], John Hammond[4]

[1]U.S. Geological Survey, Wyoming-Montana Water Science Center, Cheyenne, WY, 82007, USA
[2]U.S. Geological Survey, Wyoming-Montana Water Science Center, Helena, MT, 59601, USA
[3]U.S. Geological Survey, Oregon Water Science Center, Portland, OR, 97204, USA
[4]U.S. Geological Survey, Maryland-Delaware-D.C. Water Science Center, Catonsville, MD, 21228, USA
[5]U.S. Geological Survey, Water Mission Area, Catonsville, MD, 21228, USA
[6]U.S. Geological Survey, Utah Water Science Center, Salt Lake City, UT, 84119, USA
[7]University of Waterloo, Waterloo, ON, N2L 3G1, Canada
[8]U.S. Geological Survey, Idaho Water Science Center, Boise, ID, 83702, USA

Correspondence to: Aaron Heldmyer (aheldmyer@usgs.gov)

Abstract. Drought is a highly consequential natural disaster that may likely increase in both severity and extent across the conterminous United States (CONUS). The mechanisms affecting the propagation of drought from the atmosphere to streamflow are complex and interactive, making the prediction of streamflow drought difficult with current modeling approaches. Machine learning is an emerging tool in the field of hydrology that may be well-suited to prediction of streamflow drought across large and topographically diverse areas. Here, we train and analyze 3,198 random forest models at U.S. Geological Survey streamgages to understand common meteorological drivers of streamflow drought and to define physiographic characteristics of basins sensitive to these drivers. We also develop a novel dynamic regionalization approach using donor gages to predict daily streamflow drought at pseudo-ungaged locations. For this study, CONUS was divided into nine regions for ease of reference in describing results. Our results show that teleconnections, temperature, evaporative demand, and snow-water equivalent are important drivers of streamflow drought in the West, Southwest, and Northern Rocky Mountains (Northern Rockies) regions of the United States, and precipitation and soil moisture are primary drivers of streamflow drought in the Northeast, Southeast, and the Northwest regions. Prediction using dynamic regionalization shows comparable performance to at-site models.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

1 Introduction

Droughts rank first among all natural disasters in terms of the number of people affected (Mishra and Singh, 2010). The vast extent and protracted nature of drought often result in severe and lasting consequences, both for ecosystem stability as well as numerous economic sectors including agriculture, recreation, tourism, and more (Basara et al., 2013; Huang et al., 2017; Manuel, 2008; Wlostowski et al., 2022; Zou et al., 2018). In the conterminous United States (CONUS), modern droughts such as the 2010-2013 drought in the southern United States, the subsequent 2012-2013 expansion into the North American drought that affected most of the United States, central and eastern parts of Canada, and parts of Mexico, and the 2012-2014 drought in California have all resulted in economic losses totalling billions of dollars (Griffin and Anchukaitis, 2014; Seneviratne, 2012; Williams et al., 2015). Moreover, while certain areas of the United States like the arid western part of the country are predisposed to drought development and persistence of droughts (Zhang et al., 2021), droughts remain a threat across all of the CONUS due to evolving climate conditions and land use.

Streamflow often provides the most easily monitored hydrologic drought signal in the CONUS owing to a robust network of U.S. Geological Survey (USGS) streamgages. This advantage provides important early warning signs of impending drought conditions in monitored areas, with two important caveats. The first caveat is that not all locations are monitored, and the spatial distribution of unmonitored locations is not random. For example, small watersheds (<10km2) and watersheds at higher elevations are underrepresented in the USGS streamgage network, particularly in arid regions (Deweber et al., 2014; Kiang et al., 2013; Krabbenholft et al., 2022). The second caveat is that the streamflow drought signal is derived from multiple complex and interconnected mechanisms that are introduced as the drought signal propagates from the atmosphere to the land surface. The interactions between mechanisms affecting the propagation of meteorologic drought to hydrologic drought result in complex, non-linear spatial and temporal effects that are still not fully understood. Beven (2001) notes the difficulty in identifying scale-invariant climatic controls on runoff generation due to spatial and seasonal variability in patterns of water storage and residence time.

Recent advancements in data mining and machine learning (ML) applications to the field of hydrology may provide new insights into the spatial and temporal patterns of streamflow drought and the associated drivers across the CONUS (Hamshaw et al., 2023; Kratzert et al., 2019; Shen et al., 2021). ML has a distinct advantage over more traditional approaches in its ability to utilize a wide variety of disparate datasets, potentially uncovering complex patterns in space and time (Jiang et al., 2022; Nunes Carvalho et al., 2022; Zhu et al., 2021). Additionally, the fewer assumptions made in ML-based modeling techniques offer advantages in both scalability to larger systems, as well as transferability to new systems altogether (Li et al., 2022; Tahmasebi et al., 2020).

Generally, there is a positive correlation between model performance and the amount of training data used (Goodfellow et al., 2016). However, workloads for artificial intelligence (AI) and ML potentially require extremely large volumes of data processing and storage capacity (Gbedawo et al., 2023). Further, increasing the interpretability of AI and ML methods for high-stakes decision making is crucial (Rudin et al., 2022). In hydrologic modeling, one approach to reducing the volume of

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

training data required while minimizing the amount of information loss is to use dynamic regionalization to select a subset of models trained at "donor" gages, similar in character but not necessarily proximal to the ungaged location of interest. This subset is then used for prediction at the ungaged location of interest (McIntyre et al., 2005). This method has the advantage of reducing the volume of training data and potentially avoiding the introduction of noise from including training data from dissimilar basins (Pagliero et al., 2019).

Through the advantages provided by ML and data mining techniques, we seek to further understand spatial and temporal streamflow drought mechanisms, as well as their interactions, within the CONUS. Summarily, we address the following research questions:

1. What are the common hydrometeorologic drivers of streamflow drought in the CONUS?
2. What are the defining physiographic characteristics of basins most sensitive to influential drivers of streamflow drought?
3. Can we use dynamic regionalization (i.e., donor gages) to predict daily streamflow drought at ungaged locations?

2 Methods

2.1 Study Area

Our study area consists of 3,198 watersheds across the CONUS monitored by USGS streamgages (Figure 1) from the dataset known as the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES) version 2 (Falcone, 2011), hereafter referred to as GAGES-II. Basins monitored by streamgages included in the GAGES-II dataset span the extent of the CONUS and capture much of the hydrologic complexity faced across the nation.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

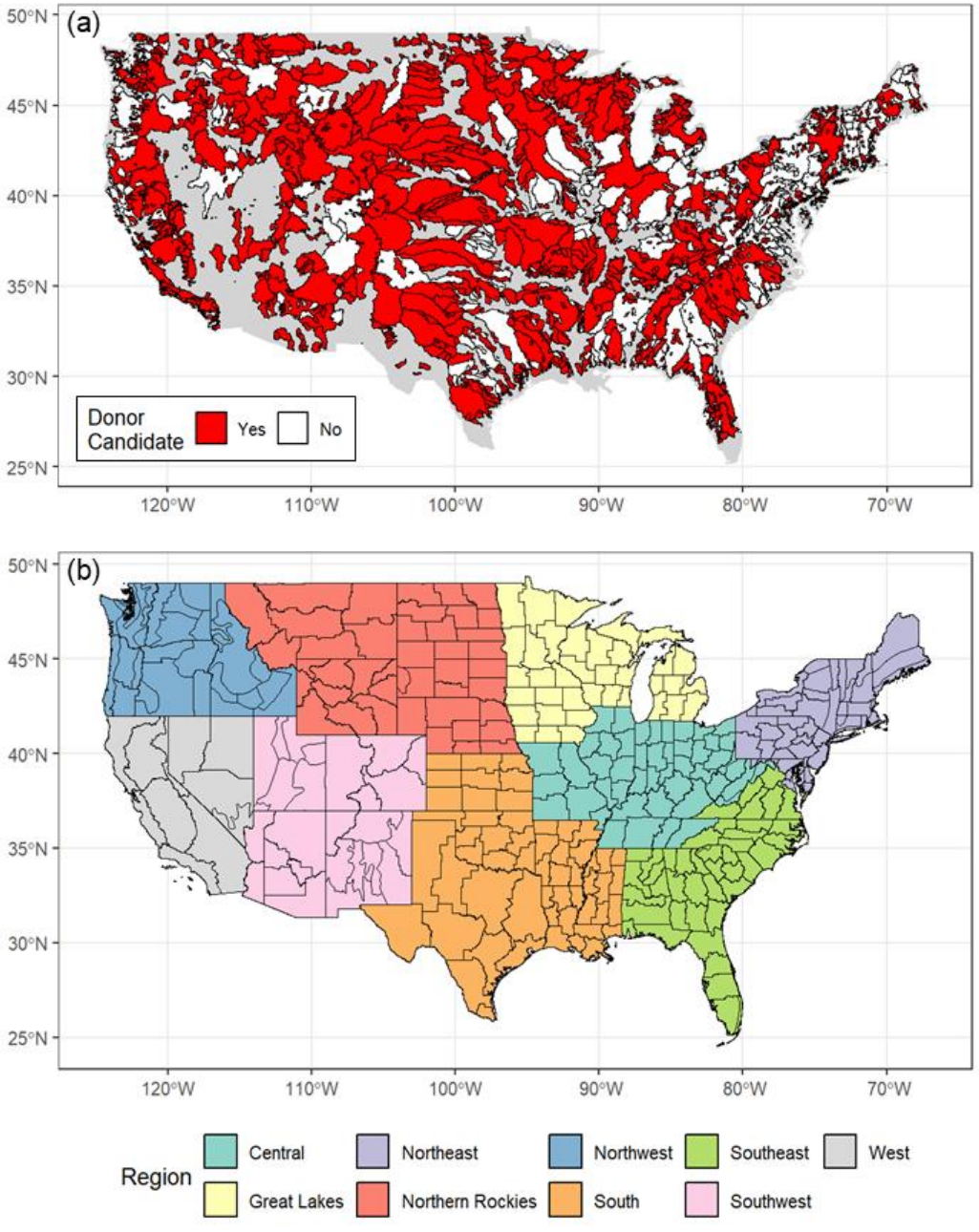


Figure 1. Panel a) shows the 3,198 basins in the conterminous United States which comprise the study area. Red basins indicate the 1,900 filtered basins with Kappa > 0.4 used as candidate donors. The filtering process is described in Section 2.4.1. Panel b) shows regional naming convention for reference.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Because of the non-random placement and objectives of USGS streamgages, there are well-documented biases associated with the GAGES-II dataset and the basins monitored (Deweber et al., 2014). These biases translate into nonnormal distributions of basin characteristics related to drainage area (i.e., larger rivers are disproportionately monitored), popularity (i.e., rivers considered to have higher recreational value are disproportionately monitored), water use (i.e., rivers that are primary sources of public use are disproportionately monitored), and others.

2.2 Data

2.2.1 Streamflow Data

We obtained daily streamflow data for 9,068 GAGES-II streamgages across the CONUS for climate years (April 1-March 31) from 1981-2020. To reduce inaccurate drought identification from sparse records, we selected streamgages where at least eight years per decade had streamflow observations for 95 percent of days in each climate year (Hammond et al., 2022; Simeone, 2022). We used climate years for the streamgage selection as they better include the full low-flow season (generally, late summer through fall, depending on gage location) in the CONUS (Carpenter and Hayes, 1996; Feaster and Lee, 2017). After processing and filtering based on daily streamflow records, our dataset consisted of 3,198 gages. A full description of the input data and processing methods are available in Simeone (2022). All data processing was done using R version 4.2.2 (R Core Team, 2021).

We identified streamflow droughts using daily streamflow values converted to percentiles using the Weibull plotting position (e.g., Schlögl and Laaha, 2017) following methods from Simeone (2022). Since streamflow often follows an annual seasonal cycle, we selected a definition of drought that represents a departure from typical annual cycles. To remove this seasonality, streamflow percentiles were based on the Weibull plotting positions computed from data for each day of the year using the values for a 30-day window centered on the given day from all years:

$$percentile = \frac{r}{n+1} \quad (1)$$

where r is rank and n is the number of data points. To account for the effect of zero-flow days in the daily streamflow data, we implemented the combined Threshold Level Method and Continuous Dry period method developed by van Huijgevoort et al. (2012). The method treats the number of continuous days with zero flow as a proxy for streamflow level, and breaks ties between zero-flow percentile rankings based on the number of preceding zero-flow days (Simeone et al., 2024). For this analysis, we conceptualize droughts as binary events. A streamflow percentile below a predefined severity threshold is considered a drought; above this severity threshold is a non-drought event. We selected the 20th percentile as the severity threshold for this analysis as a balance between a useable drought definition and an adequate number of occurrences to inform the statistical model, although we recognize that for specific needs other severity levels may be more related to risks in water resource scarcity. Our definition of streamflow drought means that every location in the CONUS will be in drought 20 percent

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

of the time, and that these droughts will be evenly distributed throughout the year. We did not de-trend the data, so long-term trends in streamflow could affect the temporal distribution of droughts due to our stationary definition of drought. A threshold-based drought identification method was chosen over other methods that incorporate drought longevity to avoid the additional complexity that arises from an added minimum duration threshold across a domain where typical drought longevity can vary substantially.

2.2.2 Climatic and physiographic drivers

Climatic and physiographic variables were selected as explanatory variables in the model based on previous literature, data availability, and exploratory data analysis (Table 1). Teleconnections data were included to provide additional insight into drought conditions, particularly at locations where local precipitation patterns do not often correspond strongly with streamflow such as in basins with high baseflow. Date information was included as a predictor to account for possible non-seasonal temporal patterns that may not be eliminated through Weibull transformation of the streamflow. Sunspot data were included as indicators of solar activity, which can influence atmospheric circulation, precipitation patterns, snowpack, and temperature (Yang and Xing, 2021). Antecedent streamflow data were not used as this would have prevented the utilization of the donor method (described in 2.4) at ungaged locations where streamflow is not available. For the individual gage models, only climatic (i.e., time-varying) data were included. This is because these models were only time-varying, and static variables would have provided no additional information. All geospatial data processing was done using the R package gdptools (McDonald, 2022) and followed the workflow described in Text S1.

All climatic variable data (with the exception of climate teleconnection variables) were transformed into percentiles using both the variable-threshold method (thresholds are calculated for each day of the year using only the values for that day and surrounding 30-day window from all years on record; identified by "weibull_jd" in the variable name) described in section 2.2.1 as well as a fixed-threshold method (all values in the period of record are used to calculate a single fixed threshold; identified by "site" in the variable name and described in Simeone, 2022) that results in a deviation from the long-term average. Transformed variables were included as predictors alongside their untransformed forms. In addition, we included values derived from rolling windows of 30 (monthly trends), 90 (seasonal trends), and 365 (annual trends) days to average or sum preceding daily values to provide the models with information on antecedent conditions. This resulted in 130 total predictors.

Table 1. Selected explanatory variables for the at-site models. With the exception of the teleconnections (AMO, PDO, ENSO, PNA), all variables were also transformed to percentiles and smoothed using rolling 30, 90, and 365-day rolling windows.

Variable	Units	Source	Reference
Minimum Temperature	°C	gridMET	Abatzoglou, 2013
Maximum Temperature	°C		
Precipitation	mm		
Evapotranspiration (Reference – grass)	mm		

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Standardized Precipitation Evapotranspiration Index (SPEI)	unitless		
Snow Water Equivalent (SWE)	mm	NASA NSIDC	Broxton et al., 2019
Soil Moisture (0-10 cm depth)	kg/m^2	NASA NLDAS2	Mitchell et al., 2004
Soil Moisture (10-40 cm depth)	kg/m^2		
Soil Moisture (40-100 cm depth)	kg/m^2		
Atlantic Multidecadal Oscillation (AMO)	unitless	National Oceanic and Atmospheric Administration	Enfield et al., 2001
Pacific Decadal Oscillation (PDO)	unitless	National Oceanic and Atmospheric Administration	Mantua, 1999
El Nino-Southern Oscillation (ENSO)	unitless	National Oceanic and Atmospheric Administration	Bjerknes, 1969
Pacific North America pattern (PNA)	unitless	National Oceanic and Atmospheric Administration	Barnston & Livezey, 1987
Sunspots	International sunspots	Royal Observatory of Belgium	Clette et al., 2015
Date	decimal date	N/A	N/A

2.3 At-site model development and evaluation

All model development was done using R version 4.2.2 (R Core Team, 2021) run on Amazon Web Services using Amazon SageMaker Studio. We chose to implement random forest classification models (Breiman, 2001) for the modeling framework because of the robustness of the method (Biau and Scornet, 2016; Tyralis et al., 2019), robustness to redundancies and correlations among predictor variables (post-hoc dimension reduction, described in 2.4.2, is utilized over pre-processing for this reason), and the ability to quantify the relative importance of individual predictor variables (Archer and Kimes, 2008). The modeling was done using the caret package in R (Kuhn et al., 2023).

2.3.1 Training and configuration

A random forest classification model was trained for each of the 3,198 streamgages included in the analysis. Because of the imbalance in the proportion of drought events (0.2), the training data were weighted inversely to the percentile threshold used to define a drought, or in this case, 0.2. That is, all observations of drought were weighted by 0.8 and all observations of non-drought were weighted by 0.2. This allows the random forest models to better represent minority class instances of drought by increasing the contribution of the drought instance to the data-splitting calculation at each node in the decision trees. In this

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

case, a weight of 0.8 for drought instances and 0.2 for non-drought instances means each instance of drought is treated as four instances for each instance of non-drought. Pre-processing on each model to reduce the number of predictor variables was avoided in favor of post-processing to facilitate direct comparison among the 3,198 models developed for the analysis. Because random forests are robust to correlated predictors and uninformative predictor variables, preserving the full input space for all models allows for interpretability and enhanced pattern detection, two major goals for this analysis.

Variable importance was computed using the Gini Index, or mean decrease in impurity (MDI), which sums the gain associated with all splits performed along a given variable. MDI was selected to compute variable importance because it has been shown that rankings based on MDI can be more robust to perturbations of the data compared with those obtained with permutation-based importance (Calle and Urrea, 2011). Minimum node size was set to 10 to avoid overfitting. Cohen's Kappa (Cohen, 1960) was used to select the optimal model because it is not prone to bias toward the majority class.

2.3.2 Model evaluation

To understand model performance at each streamgage, the models were trained on data from 1985-2015 and tested on data from 1982-1984 and 2016-2020. The early testing period (1982-1984) is shorter because the snow water equivalent data we used are not available prior to 1982. We included an early and late testing period to be able to account for and potentially analyze the presence of nonstationarity in the frequency of streamflow droughts at the streamgages of interest.

We selected evaluation metrics that provide a fair assessment of the prediction, accounting for the imbalance in the drought data. Selected metrics include:

Cohen's Kappa coefficient (Cohen, 1960): $\frac{p_o - p_e}{1 - p_e}$ (2)

Observed agreement, p_o : $\frac{Tpos+Tneg}{Tpos+Tneg+Fpos+Fneg}$ (3)

Expected agreement by chance, p_e : (4)

$$\frac{\boldsymbol{Tpos + Fneg}}{\boldsymbol{Tpos + Tneg + Fpos + Fneg}} \times \frac{\boldsymbol{Tpos + Fpos}}{\boldsymbol{Tpos + Tneg + Fpos + Fneg}} + \frac{\boldsymbol{Fpos + Tneg}}{\boldsymbol{Tpos + Tneg + Fpos + Fneg}} \times \frac{\boldsymbol{Fneg + Tneg}}{\boldsymbol{Tpos + Tneg + Fpos + Fneg}}$$

Model sensitivity: $\frac{Tpos}{Tpos + Fneg}$, (5)

Model specificity: $\frac{Tneg}{Tneg + Fpos}$, (6)

Balanced model accuracy: $\frac{Model\ sensitivy + Model\ specificity}{2}$, and (7)

Overall model accuracy: $\frac{Tpos+Tneg}{Tpos+Tneg+Fpos+Fneg}$ (8)

where Tpos is the number of true positives, Fneg is the number of false negatives, Tneg is the number of true negatives, and Fpos is the number of false positives.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Cohen's Kappa score (hereafter, "Kappa") is designed to evaluate the observed model accuracy compared to the expected model accuracy. By adjusting for agreement that might happen by random chance, it is a helpful metric for evaluating models trained on imbalanced data where a naïve model simply predicting the majority class might result in misleading metrics. Model sensitivity and specificity characterize the model performance specifically associated with events (droughts) and non-events (non-droughts), respectively. Balanced model accuracy represents the mean of model sensitivity and specificity and provides an unbiased evaluation of model performance as it relates to prediction of both drought and non-drought events.

2.4 Development and evaluation of dynamic regionalization method

Within the overall methodology (Figure 2), the analysis and application of the model output to a donor-based prediction method was conducted in four steps: 1) pre-analysis filtering of gages associated with poor model performance, 2) dimension reduction of important predictor variables using principal component analysis (PCA), 3) regression of principal components (PCs) using basin characteristics as predictors, and 4) development of a donor-gage drought prediction method at pseudo-ungaged basins using regressed principal components.

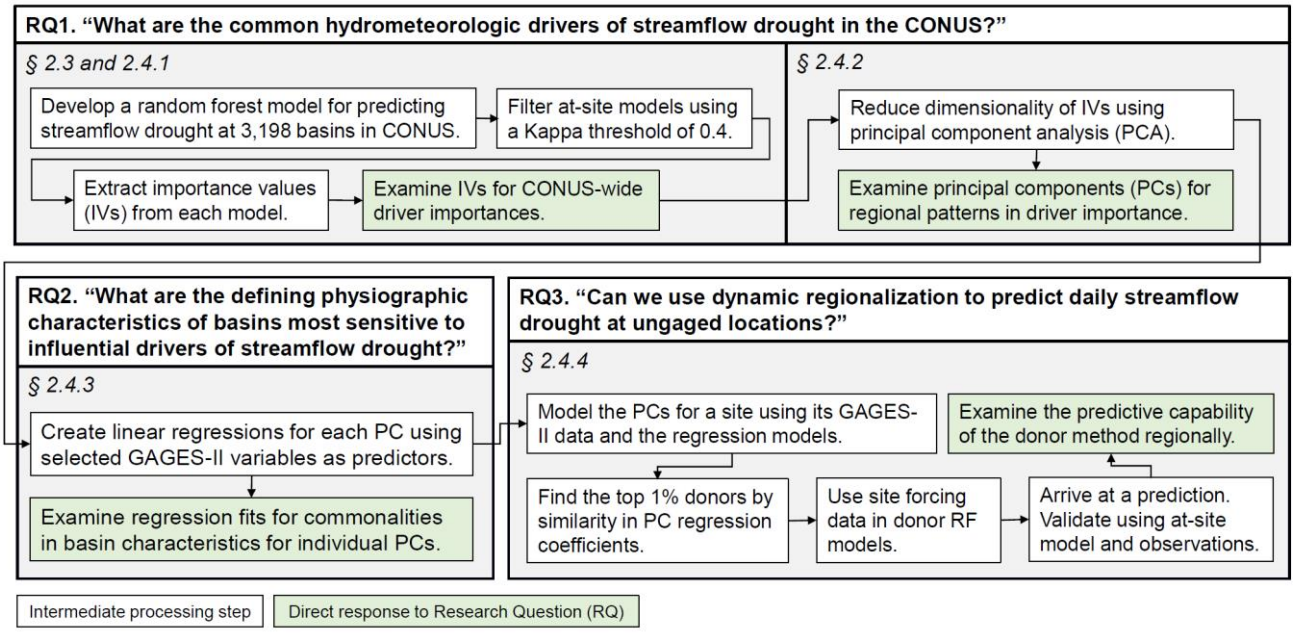


Figure 2. Methodology flowchart describing the intermediate processing steps in white boxes and steps that directly address research questions in green boxes. Boxes are organized by the three major research questions. Sections that elaborate on the steps are also listed in each box for reference. CONUS refers to the conterminous United States.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

2.4.1 Pre-analysis filtering of poor-performance gages

At-site models with poor performance (Kappa scores less than 0.4) were excluded from the potential donor candidate pool. We observed that poor performance was typically seen in the Northern Rocky Mountains (Northern Rockies), West, and Southwest regions of the United States, where in a greater prevalence of basins, water management practices can alter the observed streamflow (Falcone, 2011). Because these practices are not adequately represented with the available data, we are unable to model them accurately. Gages which are subject to high levels of disturbance are often located in the arid west where water management practices may have a larger proportional effect. Because of this, consideration was given to trade-offs between spatial representativeness and noise reduction in the filtering process.

Kappa was prioritized as the primary filtering mechanism above other metrics as it is a more robust measure of classification performance in imbalanced datasets. This decision ensured that the models used in the analysis not only demonstrated high predictive power, but also a robust ability to handle class imbalances and make informed classification decisions when applied to drought conditions. Ultimately, the threshold value of Kappa > 0.4 was set to include values of at least a "fair" to "moderate" rating (Fleiss et al., 2003; Landis and Koch, 1977). This reduced the dataset from 3,198 models to 1,900 filtered models. These filtered models were then used for the remainder of the analysis, as well as donor candidates in the method described in 2.4.4.

2.4.2 Dimension reduction of variable importance

The Gini Index variable importance measure described in 2.3.1 is a useful byproduct of the random forest (RF) models, as it provides insight into the discriminative ability of each predictor variable for streamflow drought classification (Archer and Kimes, 2008). However, the high dimensionality of the 130 predictor variable dataset for each RF model, as well as the influence of insignificant variables and correlative effects, necessitated transformation prior to analysis and interpretation (Jolliffe, 2002). Therefore, we applied PCA (Hotelling, 1933) to reduce the dimensional complexity of the predictor variable importance values generated for each of the filtered random forest models. The variable importances were scaled to have unit variance before analysis, and the top 27 PCs were selected accounting for 95% of the total variability in the dataset. PCA was conducted in R using the *stats* package (R Core Team, 2021). Correlation among variable importances were not explicitly calculated due to the scale of the predictor dataset. However, loadings with similar magnitude and direction within a principal component are generally proportional with the Pearson correlation coefficient (Frost et al., 2015). Therefore, loadings were examined here following the dimensional reduction as part of the overall analysis.

2.4.3 Regression of principal components using basin characteristics

To predict at locations outside the modeled dataset, PC scores were estimated using regressions built with 58 selected climatic and physiographic basin characteristics obtained from the GAGES-II dataset (Falcone, 2011). These variables were derived from a prior study (Konapala and Mishra, 2020), and selected based on their relatively high influence over streamflow drought (Addor et al., 2018; Rice et al., 2015; Stoelzle et al., 2014). Included are 10 climate variables, 15 hydrologic catchment

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

variables, 4 land cover variables, 23 soil characteristic variables, and 6 topographic variables. Table S2 lists these variables with a short description. A complete description of these variables can be found in the metadata associated with the GAGES-II dataset (Falcone, 2011). A linear model was deemed sufficient for modeling the principal components through inspection of each covariate relationship with the first 6 principal components, as well as ANOVA-based comparative model testing of alternatives to the Gaussian error distribution. Therefore, a linear regression for each of the 27 principal component scores was developed using the 58 GAGES-II variables at all donor-gage locations. The variables were each scaled such that they were centered with a mean of 0 and a standard deviation of 1 to facilitate comparison of the regression coefficients.

2.4.4 Development of donor-gage prediction method

To accomplish a donor-based prediction at a location of interest (LOI) outside the donor-gage network, we developed a method which selects characteristically similar donors for the LOI and combines them to generate a single prediction time series. First, a set of PC scores were estimated for the LOI using the regression models described in 2.4.3 and the LOI basin characteristics. To determine a set of suitable donors, a weighted sum of absolute differences between the estimated PC scores for the LOI and the PC scores for each candidate donor was computed. That is, the absolute value of the difference between each estimated PC for the LOI and each corresponding PC generated for each candidate donor by the random forest models was computed. Individual PC scores were weighted by the percent of total variance explained by the PC as described in 2.4.2. For example, if PC1 explains 25% of the total variance in the original dataset of variable importances, the calculated values for the absolute difference between PC scores for PC1 were weighted by 0.25. This ensured that high variance and therefore more determinative PCs for donor selection were given more weight for the donor selection process. The weighted differences for each PC score between the LOI estimate and each donor were then summed for each candidate donor. In an effort to dilute the effects of potential outlier donors due to possible low regression strength, we selected the top 1% (i.e., 19 of 1,900) of candidates that resulted in the minimal weighted absolute sum of differences to comprise the donors for the LOI. The random forest explanatory variables described in Table 1 for the LOI were then used to force each of the donor RF models and provide a set of drought probability time series. These were then averaged and converted to a binary series using a threshold of 0.35, which was determined based on maximization of Kappa between the donor predictions and observations through iterative testing at 0.05 intervals from 0.05 to 0.95.

2.4.5 Evaluation of donor-gage predictions

Donor-gage predictions were evaluated at all 3,198 sites (i.e., both filtered and unfiltered), as well as at the filtered subset of 1,900 sites, using the Kappa metric based on two targets. The first target was the relative performance of the donor-gage predictions when compared to the time series of streamflow droughts as predicted by the at-site model for the test period at each gage. This evaluation provided both an understanding of how well the donor-gage method performed compared to what can be expected from a model at that gage, as well as the validity of matching donors using patterns in variable importances.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

The second target we used was the timeseries of observed streamflow droughts at each gage. This target allowed us to assess the performance of the donor-gage prediction method with all types of error included in the analysis.

3 Results

3.1 At-site model results

3.1.1 Model performance

Performance for the at-site RF models evaluated with the metrics described in 2.3.2 are summarized below in Table 2. Performance across all metrics were higher at filtered sites than at all locations

Table 2. At-site model performance metrics for all sites (n=3,198) and filtered sites (n=1,900) based on comparison of at-site RF models to observed streamflow drought time series.

Metric	Subset	Mean	Median	Standard Deviation	Minimum	Maximum
Balanced Accuracy	All sites	0.75	0.77	0.12	0.35	1
	Filtered	0.82	0.82	0.06	0.63	1
Kappa	All sites	0.42	0.45	0.2	-0.16	0.88
	Filtered	0.56	0.55	0.09	0.4	0.88
Overall Accuracy	All sites	0.86	0.87	0.07	0.46	1
	Filtered	0.88	0.88	0.04	0.75	1
Sensitivity	All sites	0.59	0.65	0.26	0	1
	Filtered	0.73	0.75	0.14	0.27	1
Specificity	All sites	0.91	0.91	0.06	0.48	1
	Filtered	0.91	0.91	0.05	0.69	1

Spatial patterns in Kappa were evident, with greater Kappa scores primarily in the Central, Southeast, and Northeast regions, along the Pacific coast of the West and Northwest regions, and towards the center of the Southwest region (Figure 3).

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

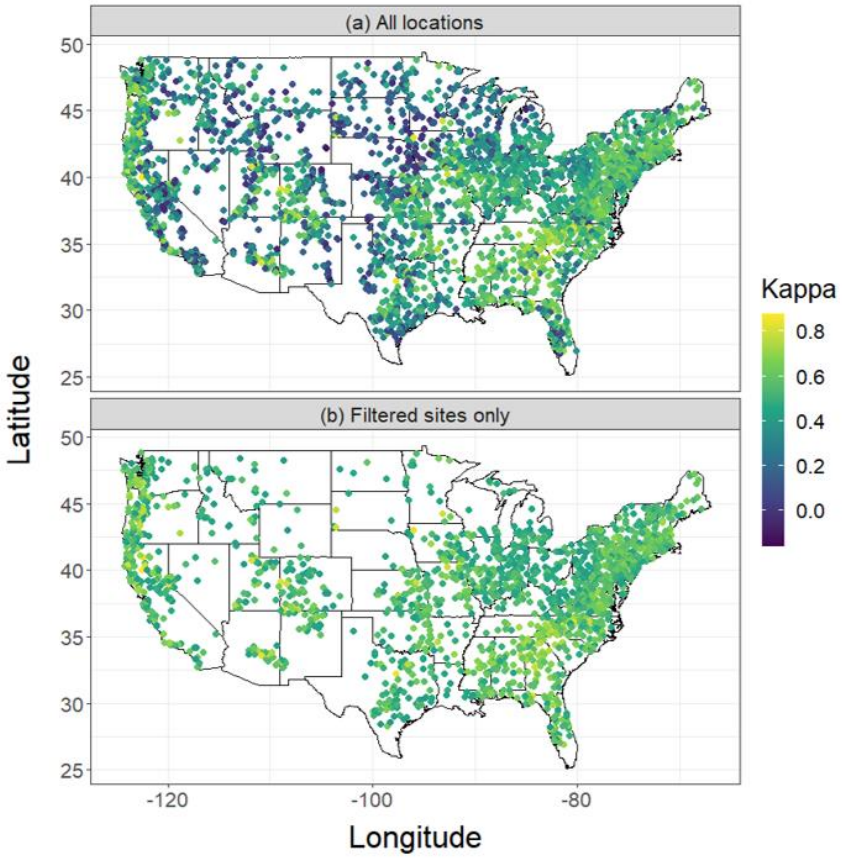
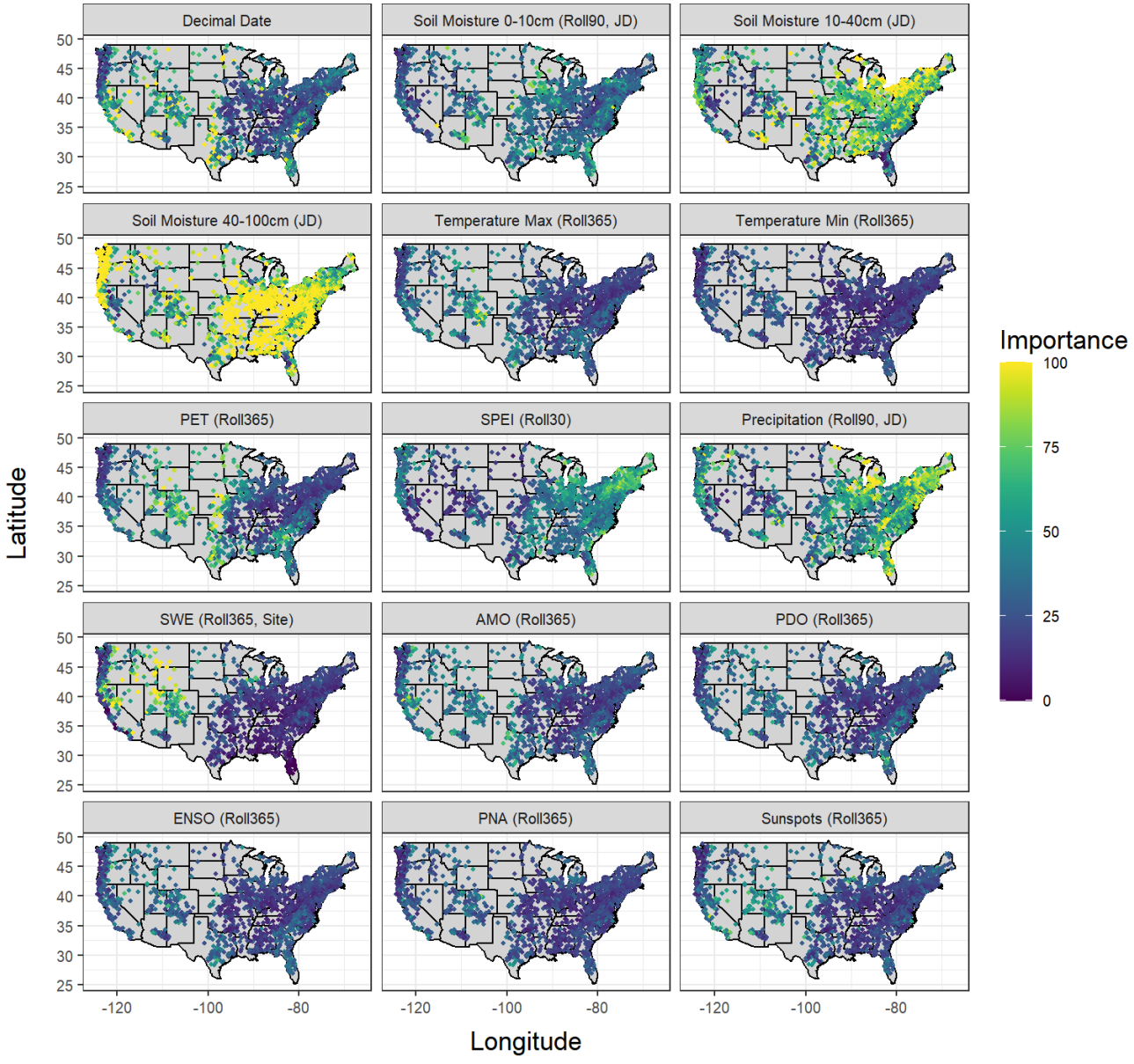


Figure 3. At-site model performance represented by the Kappa metric for a) all 3,198 sites, and b) 1,900 sites filtered by Kappa scores > 0.4.

3.1.2 Model Importance Values

Following the filtering described in 2.4.1, we examined the variable importance values produced for each of the 130 input variables included in the 1,900 models. Boxplots of values associated with each variable are provided in Figure S3. Across all models, the input variables with the highest median importances were soil moisture, precipitation, and Standardized Precipitation Evapotranspiration Index (SPEI) at various rolling window smoothing lengths, with some Weibull-transformed and others in raw form. However, spatial patterns across most variables were evident (Figure 4). For example, potential evapotranspiration (PET) was found to have higher importance in the South, Southwest, and Southeast regions. SPEI had higher importance across the CONUS except in the Southwest region, whereas snow water equivalent (SWE) importance was high in the Southwest, Northern Rockies, Northwest, and West regions. Teleconnections (AMO, PDO, ENSO, PNA), while low relative to other variables, were generally higher importance in the western CONUS (Northwest, West, Northern Rockies, Southwest regions). Decimal Date was found to be relatively low importance at most sites except in the West, Southwest, and South regions, where certain gages demonstrate an outsized value.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026



295

Figure 4. Spatial plots of the transformations (indicated in parentheses) with the highest median Gini Index scores (i.e., variable importance) for each variable at 1,900 at-site models across the CONUS. Roll30, Roll90, and Roll365 indicate rolling smoothing windows of 30 days, 90 days, and 365 days, respectively. JD and Site indicate Weibull transformation using the Julian Day (variable) or site-based (fixed) drought threshold definitions, respectively.

300 Overall, the variable with the greatest mean importance was Weibull-transformed soil moisture between 40- and 100-cm depth (mean Gini Index score of 82.5), followed by Weibull-transformed soil moisture between 10- and 40-cm depth (mean Gini Index score of 64.1). Following these in mean importance were three precipitation variables: 90-day rolling average Weibull-

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

transformed precipitation (mean Gini Index score of 57.2), 90-day rolling average precipitation (mean Gini Index score of 52.2), and 365-day rolling average Weibull-transformed precipitation (mean Gini Index score of 50.3). The 90-day rolling average Weibull-transformed SPEI was the next greatest (mean Gini Index score of 47.6). Variables besides soil moisture, precipitation, and SPEI did not have mean Gini Index scores greater than 37.1, which was the value found for the decimal date variable.

3.2 Principal Components Analysis

PCA results revealed that the first three principal components (PCs) explained about 60 percent of the variance in the variable importance data, and the first six PCs explained about 80 percent. While a total of 27 PCs were included in the analysis, only the top three will be discussed here in detail since regional patterns are only clearly apparent for these PCs. Interpretation of the PCs in terms of the original variables illuminates some high-level data structures present among the original RF model input variables. In general, loadings (the coefficients of the linear combination of the original variables from which the PCs are constructed) closer to -1 or 1 indicate that the variable strongly influences the component. However, due to the complexity and high dimensionality of the input data, loadings for the first three PCs did not exceed a magnitude of 0.25.

Relatively high magnitude loadings for PC1 include teleconnections such as PNA, ENSO, and PDO, as well as atmospheric energy and evaporative demand metrics such as temperature (both minimum and maximum daily) and PET (Figure 5). However, these variables were less influential in the Central and coastal Northwest regions of the CONUS. These loadings were between -0.11 and -0.13. PC2 is dominated by moisture-related metrics: precipitation, SPEI, and soil moisture. These loadings were between 0.11 and 0.18. Notably, a clear split between 365D soil moisture and 30D precipitation is evident in the loadings between the eastern (Northeast and Central regions) and the western (West, Southwest, Northwest, and Northern Rockies regions) parts of the CONUS. PC3 is largely characterized by soil moisture (between -0.13 and -0.21) and SWE (between 0.10 and 0.19). Soil moisture is dominant in the Southeast and South regions, whereas SWE is more influential in the Northeast, Northwest, Northern Rockies, Southwest, and interior parts of the West along the Sierra Nevada mountain range .

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

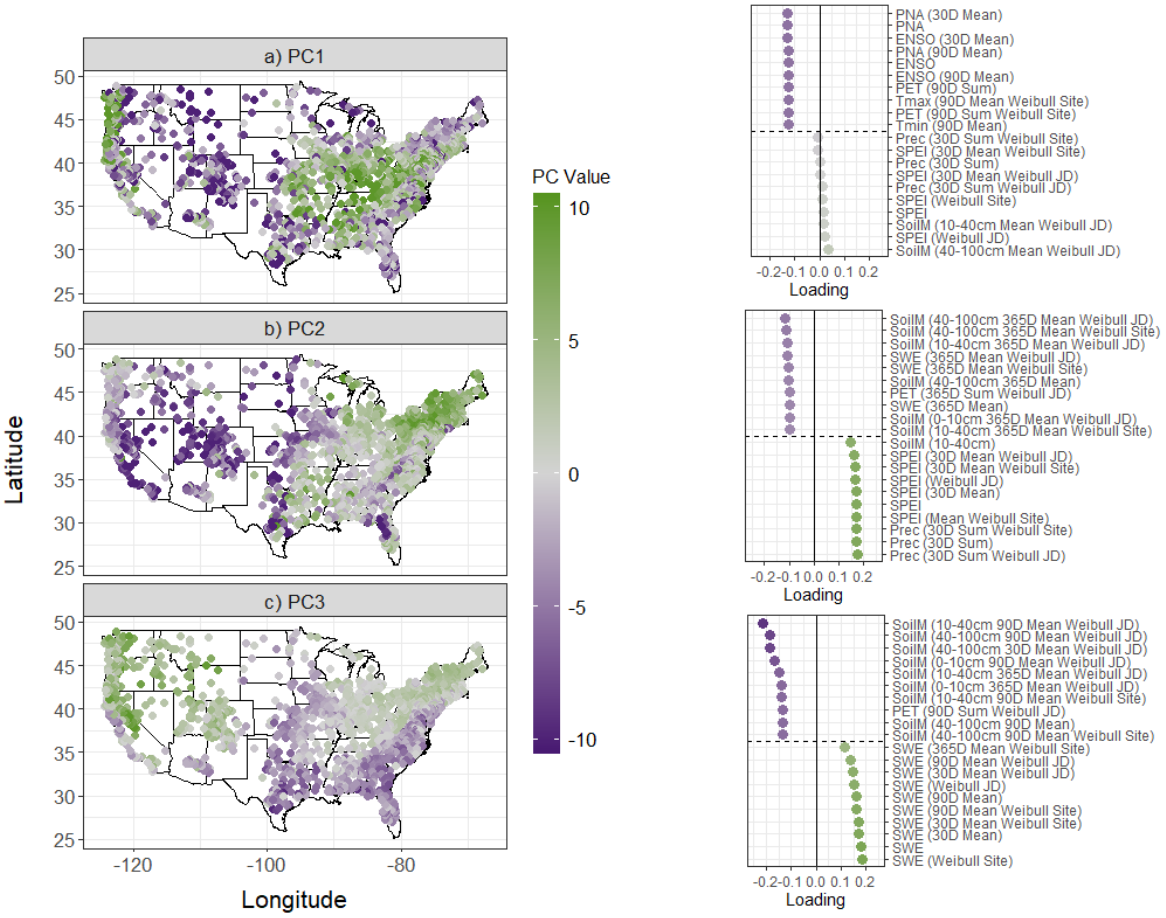


Figure 5. The top three principal components (PCs) for variable threshold random forest drought model importance values at each donor candidate site (n=1900) across the contiguous United States. Generally, PC scores indicate covariation among the predictors listed for each subplot below their associated spatial maps. Here, the 10 largest positive and negative loadings are listed in order, with a horizontal dashed line indicating where the dataset was truncated. The color of these variable points corresponds with the spatial maps and indicates a relatively strong influence of the variable at the sites of the same color.

3.3 Regression of PCs with donor basin characteristics

Regression statistics for the first three PCs are provided below in Table 3. Adjusted R2 for the first three principal components were found to be between 0.41 and 0.74. Statistics for all 27 principal components are also provided in Table S4.

Table 3. Regression statistics for the first 3 principal components modeled using GAGES-II variables. For each PC, the R2, adjusted R2, residual standard error (SE), F-statistic, F-critical, degrees of freedom (df), p-value, and Akaike Information Criterion (AIC) score are provided.

PC	R2	Adjusted R2	Residual SE	F-statistic	F-critical	df	p-value	AIC
1	0.42	0.41	5.26	36.84	37	1862	< 0.01	6348.68
2	0.67	0.67	3.05	101.2	38	1861	< 0.01	4270.88
3	0.74	0.74	1.73	143.5	37	1862	< 0.01	2112.95

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

For PC1, the F-statistic was below the F-critical value. However, all p-values were found to be below 0.01, indicating statistical significance for alpha = 0.05. An abbreviated table for the most significant predictor variables (p-value < 0.001) with coefficients above 1 or below -1 for the first three PCs are shown below in Table 4 along with their coefficients. The coefficient threshold was implemented to focus the interpretation on significant variables that also generate the greatest response in the PC value.

Table 4. Linear regression coefficients for highly significant (p < 0.001) predictor variables with coefficient absolute values > 1 for the first 3 principal components (PCs).

GAGES-II Variable	Short description	Category	PC1	PC2	PC3
BFI_AVE	Base Flow Index	Hydro		-1.68	
DEVNLCD06	Watershed percent “developed”	LC06_Basin			
ELEV_MEAN_M_BASIN	Mean watershed elevation	Topo	-1.38	-2.66	
FORESTNLCD06	Watershed percent “forest”	LC06_Basin	1.28		
FST32F_BASIN	Mean day of year of first freeze	Climate	-2.09	3.11	1.78
LAT_GAGE	Gage latitude	BasinID			1.04
LNG_GAGE	Gage longitude	BasinID			
LST32F_BASIN	Mean DOY of last freeze	Climate		3.35	2.88
NO10AVE	Average percent of soil by weight passing through No. 10 sieve	Soils			
NO200AVE	Average percent of soil by weight passing through No. 200 sieve	Soils			
PERMAVE	Average permeability	Soils	2.44		
PET	Mean annual potential evapotranspiration	Climate	-10.93	-1.38	-1.89
PLANTNLCD06	Watershed percent “planted/cultivated”	LC06_Basin	1.69		
PPTAVG_BASIN	Mean annual basin precipitation	Climate	3.42	-1.01	
PRECIP_SEAS_IND	Precipitation seasonality index	Climate	-1.22		
RFACT	Rainfall and runoff factor	Soils		1.11	
SNOW_PCT_PRECIP	Snow percent of total precipitation	Climate			2.02
T_AVG_BASIN	Average annual basin air temperature	Climate	10.65		2.74
WDMIN_BASIN	Average annual basin minimum days of precipitation	Climate	-1.27	1.34	

For the PC1 regression, PET and T_AVG_BASIN were found to have very high magnitude coefficients (>10), but with opposite sign.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

3.4 Donor-gage prediction results

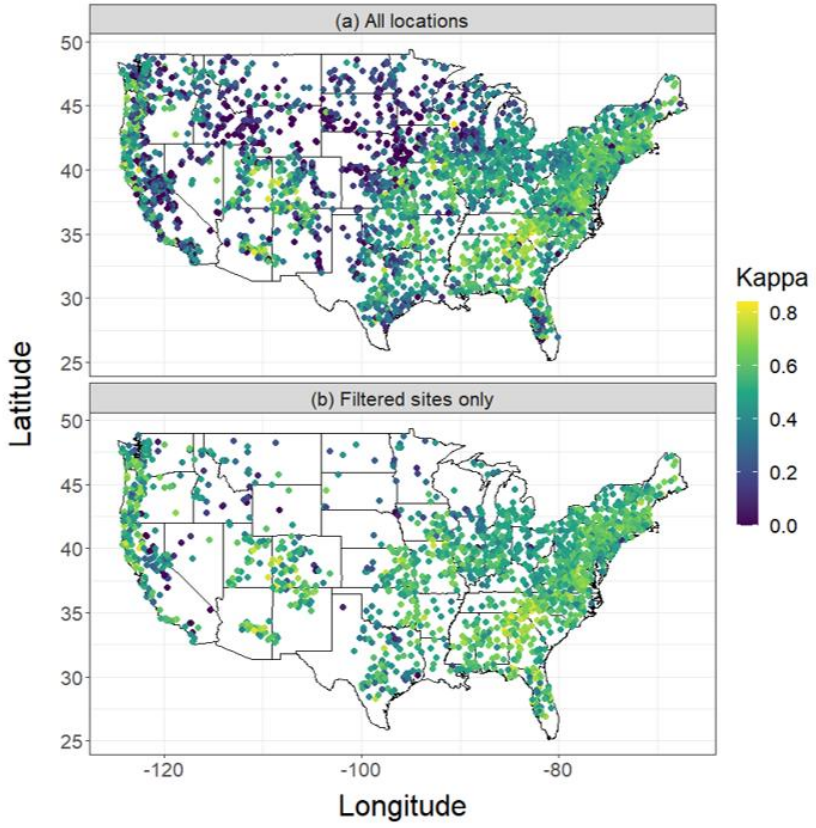
Mean Kappa scores for donor-gage predictions for the test period (Water Year [WY; Oct 1-Sept 30] 1982-1984 starting April 1982, WY 2016-2020) at all sites was 0.40 with a standard deviation (σ) of 0.22). Other metrics are summarized below in Table 5.

Table 5. Donor-gage model performance metrics for all sites (n=3,198) and filtered sites only (n=1,900) based on comparison of donor-gage model to observed streamflow drought time series.

Metric	Subset	Mean	Median	Standard Deviation, σ	Min	Max
Balanced Accuracy	All sites	0.73	0.75	0.13	0.3	1
	Filtered	0.79	0.81	0.09	0.37	0.98
Kappa	All sites	0.4	0.44	0.22	-0.26	0.84
	Filtered	0.52	0.53	0.14	-0.23	0.84
Overall Accuracy	All sites	0.86	0.87	0.07	0.38	1
	Filtered	0.88	0.88	0.04	0.62	0.99
Sensitivity	All sites	0.54	0.59	0.26	0	1
	Filtered	0.67	0.71	0.19	0	1
Specificity	All sites	0.91	0.91	0.05	0.5	1
	Filtered	0.91	0.91	0.04	0.64	1

Geographically, higher Kappa scores (0.4 and above) were distributed similarly to the at-site model Kappas (Figure 3), primarily in the East and Southeast regions (Figure 6). Higher Kappa scores were also found in the Southwest and coastal parts of the West and Northwest. Low Kappa scores (below 0.4) were found primarily in the Northern Rockies, Southwest, and Great Lakes regions.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

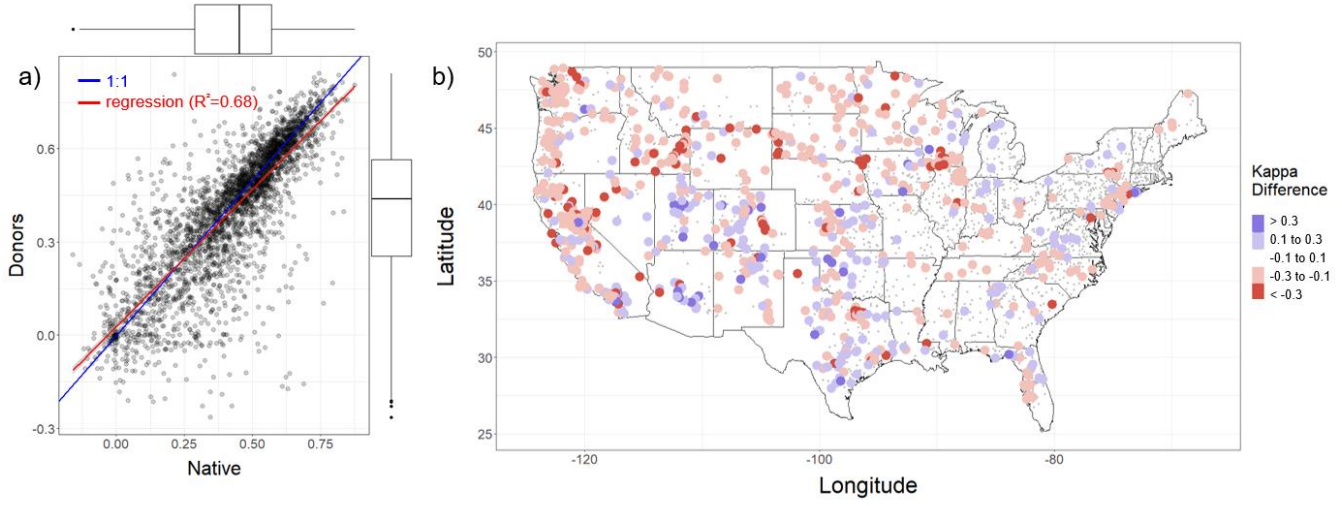


360 **Figure 6. Model performances of donor-based predictions represented by Kappa scores for the test period (1982-1985, 2015-2020) at a) all 3,198 gage locations, and b) 1,900 sites filtered by Kappa scores > 0.4.**

3.4.1 Comparisons to at-site model performance

Donor-based predictions had a mean Kappa score across all sites 0.02 lower than the at-site model predictions (0.03 for the filtered subset). Generally, the donor-based predictions outperform low-Kappa at-site models and underperform high-Kappa
365 at-site models, indicating an overall smaller distribution of scores (Figure 7). Furthermore, the geographic distribution of donor-based model performance generally followed the distribution of at-site model performance. The spatial structure of the comparative model performance was generally undefined. However, the donor-based method underperformed the at-site models in the coastal part of the West and Northwest regions, in particular.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026



370 **Figure 7. A comparison between donor-based model accuracy and at-site model accuracy at 3,198 sites in the CONUS. Panel (a) is a scatterplot showing the relationship between at-site and donor-based models and corresponding boxplots showing model distributions, and panel (b) is a map showing the differences between donor-based and at-site model Kappa values. Sites with less than a 0.1 absolute difference are represented with small grey points.**

3.4.2 Selected example gages

375 Two sites were selected to demonstrate the donor-gage method in the Northeast and Southwest regions of the United States. (Figure 8). These sites include basins represented by USGS streamgage 09152500 in the west (U.S. Geological Survey, 2024b) (Gunnison R. Near Grand Junction, CO; U.S. Geological Survey, 2024b) and USGS streamgage 01440400 in the east (Brodhead Creek Near Analomink, PA; U.S. Geological Survey, 2024a).

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

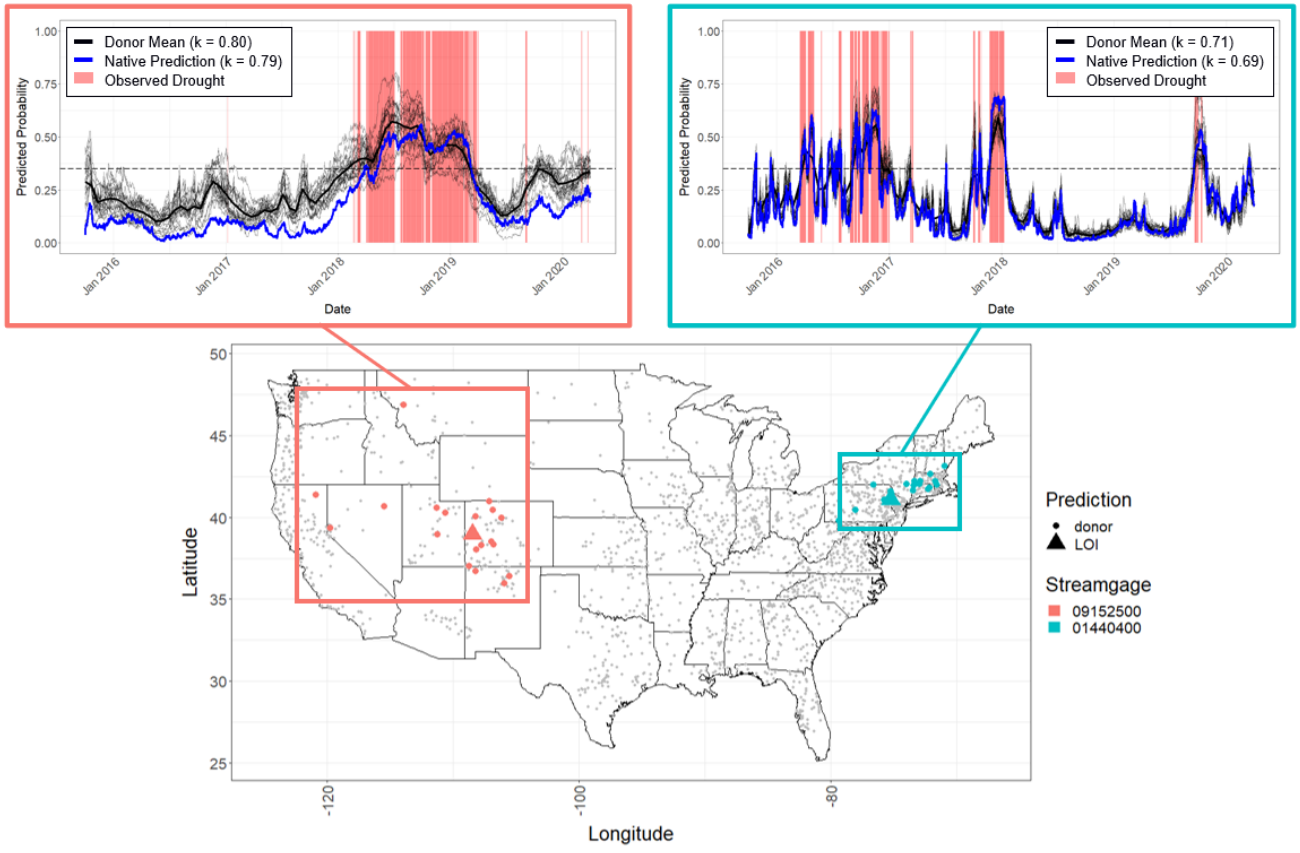


Figure 8. Two example prediction locations in the Southwest United States (U.S. Geological Survey [USGS] streamgage 09152500) and Northeast (USGS streamgage 01440400). The map displays the prediction locations as triangles, and the 19 selected donor gages for predicting at that location as circles of the same color. Time series of predicted streamflow drought probability for the model test period 2015-2020 are displayed for their corresponding prediction locations, with periods of observed streamflow drought displayed in red, the donors in black (mean donor prediction is bold), and the at-site model predictions in blue. A dashed line at 0.35 indicates the threshold for predicted streamflow drought status, where predictions above the line indicate streamflow drought and below indicate non-drought.

The drought events that occurred at these two locations differed in both longevity and frequency during the later test period (2016-2020), with droughts in the Southwest location appearing for greater lengths of time but at a lower frequency. Additionally, donor gages for the Northeast site were located a much shorter mean distance from the site at 219 km (donors for the Southwest site were located at a mean distance of 2,891 km). In addition, the relative agreement in predicted drought probability among the donors also differed, with donors generally in greater agreement for the Northeast location (mean standard deviation of 0.03) than the Southwest location (mean standard deviation of 0.06). However, despite these differences, the donor-based approach was able to perform comparably to the at-site model predictions in both circumstances, with a difference in Kappa of 0.07 and 0.01 for the western and eastern locations, respectively.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

4 Discussion

4.1 The common meteorologic drivers of streamflow drought across the CONUS

The random forest model variable importances indicated the top three important categories of variables among those tested for predicting streamflow drought in the CONUS were soil moisture, precipitation, and SPEI. These drivers, as fundamental indicators of moisture, are directly linked to streamflow and are therefore well-documented in their linkage to streamflow drought (McCabe et al., 2023; Pournasiri Poshtiri et al., 2019). Soil moisture importance, particularly at the deeper 10-40 and 40-100 cm depths, was especially high in places with relatively low aridity (i.e., regions outside of the Southwest). This is likely indicative of hydrologic memory effects that influence the persistence of drought through sustained dry periods, acting as a critical buffer to transient meteorological anomalies. Other variables, such as PET and SWE, showed a more region-specific (mainly, the Southwest and Northern Rockies) relevance to streamflow drought prediction. This is consistent with the dominant role of temperature and seasonal snowpack in the regulation of streamflow for basins in the arid Southwest and basins that are primarily snow-driven (Livneh and Badger, 2020). The teleconnection (AMO, PDO, ENSO, and PNA) and sunspots variables showed a generally low predictive importance across the CONUS. This is consistent with past findings that show these large-scale drivers have limited direct influence over streamflow at the local scale (Mallya et al., 2013; Piechota and Dracup, 1996). Notably, the Date variable shows elevated importance for certain sites in the western CONUS (South, Southwest, West regions). Given that streamflow was transformed to reduce seasonality in the models, it is possible that long-term drying trends due to persistent warming, snowpack loss, and reduced runoff efficiency are being captured here. This aligns with findings from prior literature (Livneh and Badger, 2020) documenting such shifts in many western catchments.

The principal component analysis allows for an examination of generalized drought typology in the CONUS, whereby individual meteorological and hydrologic variables can be examined by their relative fluctuations in importance across regions (Kim et al., 2021; Mainali and Pricope, 2017). Such regional drought patterns have been investigated before. Konapala and Mishra (2020) found three distinct drought regimes across the CONUS based on random forest algorithms for the period of 1979-2010 — longer duration, less frequent, and less intense droughts; moderate duration, moderately frequent, and moderately intense droughts; and shorter duration, more frequent, and more intense droughts. The basins organized into these drought regimes often exhibited similar spatial patterns to the regions that developed from the PCA in this analysis. Their study found high elevation, arid/semiarid, and/or snow-fed basins, typically found in the western CONUS (West, Southwest, Northwest, and Northern Rockies regions), were clustered into the first regime and contrast with the more humid basins in other clusters that demonstrate a more direct precipitation influence, typically found in the Northwest and eastern CONUS (Northeast, Central, and Southeast regions). Additionally, the largest loadings for PC2 in this study were largely the same variables in the positive (SPEI, precipitation) and negative (soil moisture, SWE) directions, but at different smoothing window lengths (365-day in the negative direction, 30-day in the positive direction). This is analogous to the varying drought lengths found by Konapala and Mishra (2020): models in the western CONUS were more influenced by longer smoothing window lengths for important drought predictor variables (365 days), and shorter lengths in the eastern CONUS (0 to 30 days).

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Through PCA, we found evidence of spatial groupings among broad categories of variables with relatively similar influence on streamflow drought. For example, PC1 described relatively high loadings among the teleconnections used in this study (PNA, ENSO, PDO), temperature, and PET. All these loadings were negative and are therefore associated with the region in Figure 5 populated with sites that have a negative (purple) PC value. This includes the West, Southwest, Northern Rockies, and non-coastal Northwest, as well as parts of the Southeast and Northeast regions. The outsized impact of these climate variables on drought conditions in the CONUS at regional scale has been well-documented (e.g., Baek et al., 2019; Cook et al., 2014; Woodhouse et al., 2016). Other notable examples include Singh et al. (2021) finding a significant modulation of ENSO effects on streamflow, and a PCA conducted by McCabe et al. (2004) finding a principal component highly correlated with PDO to explain 24% of total variance in drought frequency. Floriancic et al. (2021) found that high excess PET is a potential driver of low flows in catchments along the coastal Southeast and Northeast regions. The low importance of teleconnections variables at individual sites combined with high loadings and overall predominance of teleconnections variables within PC1 (demonstrating a strongly consistent, regional scale pattern) demonstrate that teleconnections act as strong background climate modulators at regional scales, influencing overall drought sensitivity without directly contributing to local drought conditions.

It is also possible that compounding interactions between the teleconnections variables is a factor in the dominance of teleconnections among the high magnitude loadings within PC1. Prior studies have shown that PDO may modify the relationship between ENSO and regional drought events by either strengthening or weakening the atmospheric circulation anomalies associated with ENSO (Gershunov and Barnett, 1998; Rao et al., 2019; Singh et al., 2021; Yeh et al., 2018). Similarly, a La Niña-induced negative PNA pattern may be intensified by the negative PDO (Hu and Huang, 2009; Nguyen et al., 2021; Wang et al., 2014). Given ideal phase timing, teleconnections may interact in compounding fashion to strengthen the severity of streamflow drought conditions at a regional scale.

We observe similar high-loading driver components and interactions in other PCs as well. PC2 combines precipitation and moisture-related drought drivers (SPEI, soil moisture), and PC3 is primarily associated with SWE and soil moisture-related drivers. Through sign analysis, we find that short-term precipitation, SPEI, and soil moisture are positive in PC2, and therefore associated with the Northeast region where PC2 is positive (green) in Figure 5. A precipitation deficit is generally considered to be the primary driver of drought conditions, though long-term changes to precipitation show a high degree of spatiotemporal variability (Naumann et al., 2018). The intermediate storage mode of precipitation, either in the soil or in snowpack, is likely the discerning difference between PC3 positive and negative regions. In the Northeast region, SWE is a high magnitude positive loading, and soil moisture is a high magnitude negative loading. Spatially, we see this in the delineation of basins into groups of snow-driven and non-snow-driven streamflow, such as between the Southeast region and mountainous parts of the western CONUS.

A difficulty that arises when linking meteorology to streamflow drought is the overlapping influence of multiple driver components within a region. From Figure 5, we observe that drought in the western CONUS is driven by numerous meteorological drivers. PC1, PC2, and PC3 all describe drought combinations of drivers centered in this area. Many individual

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

drivers are represented here: teleconnections, PET, temperature, and SWE, among others. Contrast this with the Northeast region, where we note a dominance of PC2 (precipitation-related) drivers over all others. It is possible that a major contributor to the relative difficulty in untangling predictors of streamflow drought in the western CONUS (apart from the difficulty in including water management operations), arises from this large pool of overlapping streamflow drought typologies that can affect the region. Future inclusions to the workflow, such as partial dependence analysis or Shapley Additive Explanations (SHAP) values not originally in the scope of this investigation, may prove useful in disentangling these effects.

4.2 The connection between basin physiography and streamflow drought drivers

To better understand the spatial structure of variable importance patterns in streamflow drought prediction found in the PCA, we regressed the principal components against static basin characteristics (Table 4). The resulting variability in connection strength between PCs and basin characteristics, where R2 for the first three principal component regressions showed increasing explanatory power (0.42 for PC1, 0.67 for PC2, and 0.74 for PC3), is likely indicative of the fact that basin characteristics more effectively explain the spatial variation in local, physically grounded predictors like soil moisture, precipitation, and snowpack, rather than indirect, regional-scale predictors like teleconnections, which were dominant loadings in PC1.

The high magnitude regression coefficients for the teleconnections and temperature-dominated PC1 typified the majority of the eastern CONUS (where PC1 is strongly positive) as forested or cultivated (FORESTNLCD06, PLANTNLCD06), rainy (PPTAVG_BASIN), warm (T_AVG_BASIN), and with permeable soils (PERMAVE). This is in contrast with the western CONUS (where PC1 is strongly negative), where elevations may be higher (ELEV_MEAN_M_BASIN), precipitation is more seasonal (PRECIP_SEAS_IND) and basins are drier (PET, WDMIN_BASIN). These relationships suggest a greater importance for teleconnections and energy demand-related variables in basins with less buffering capacity (low soil moisture retention and minimal snowpack), and a greater sensitivity overall towards larger-scale climate patterns such as ENSO and PNA which comprise the high-loading drivers for the PC1 drought combination. Peña-Gallardo et al. (2019) conducted a PCA investigating correlations between propagation from climate forcings (SPEI) and hydrological drought and found similar results.

PC2, dominated by moisture-related variables (precipitation, SPEI, and soil moisture), showed a contrast between the Northeast region (where PC2 is generally positive), and the western CONUS (where PC2 is negative). High magnitude coefficients associated with positive PC2 values included later freeze dates (FST32F_BASIN, LST32F_BASIN), more days with precipitation (WDMIN_BASIN), and a higher rainfall and runoff factor (RFACT). In the more humid Northeast, the strong influence of these longer-term soil moisture metrics show basins in this region retain water over longer periods. In contrast, the western CONUS associated negative PC2 values with high magnitude negative sign coefficients like elevation (ELEV_MEAN_M_BASIN) and PET, suggesting a diminished capacity for water storage over long periods. These results contextualize PC2 as a moisture-regime axis, framed by physiographic controls on catchment memory and responsiveness to shifts in moisture input.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Finally, PC3 distinguished between basins with soil moisture-driven (negative PC3 values) variables in the South and Southeast regions, and SWE-driven (positive PC3 values) variables in the Northeast, Northwest, Northern Rockies, and Southwest regions, as well as interior parts of the West region along the Sierra Nevada mountain range. From the regression, positive PC3 values were associated with high magnitude basin characteristics including snow as a percent of total precipitation (SNOW_PCT_PRECIP), timing of basin average day of first and last freeze (FST32F_BASIN, LST32F_BASIN), latitude (LAT_GAGE), and basin average temperature (T_AVG_BASIN), while negative PC3 values were associated with baseflow index (BFI_AVE) and PET. These regression relationships for PC3 point towards a dichotomy in seasonal storage between flow regimes in snow-dominated systems found in the western CONUS, and baseflow-driven systems in the South and Southeast regions where soil moisture plays a more dominant role. In the snowpack-driven western CONUS basins vulnerable to streamflow drought driven by a reduction in SWE, the effects of earlier (and therefore slower) snowmelt highlight the importance of freezing conditions timing and air temperature, as evidenced by findings that earlier, slower snowmelt decreases subsurface flow and streamflow production (Barnhart et al., 2016).

4.3 The performance and applicability of the donor-gage method

Beyond its use as an analytical tool for the investigation of streamflow drought causes and corollaries, this method was evaluated for its potential as a predictive tool in pseudo-ungaged basins. The primary motivator lies in the idea that the complexity of streamflow drought, in terms of both the interconnections and overlapping of meteorological drivers, as well as the mitigating and exacerbating factors of the land surface as the signal propagates to the basin outlet, imply a need for a model workflow capable of independent and flexibly-derived predictions.

Here, the term “independent” is defined in contrast to regional models that may be insensitive to sub-regional scale, highly heterogeneous, and influential drought drivers affecting individual basins within the region. Basins which are not well-represented by overall regional characteristics, which become increasingly prevalent as the modeled processes increase in spatiotemporal variability, may be poorly represented by such models. This is particularly true in regions that encompass a wide range of basin characteristics, such as the western CONUS. Our donor-based method addresses this issue by providing a dynamic regionalization approach to optimize a sparse-model architecture. In other words, it reduces the complexity of a regional tree-based model by pruning the number of donor gages back to only those that are similar to the ungaged location of interest.

Similarly, the term “flexibly-derived” pertains to overcoming the limitations introduced by models designed and calibrated to specific circumstances, locations, and prior human understanding of the modeled process(es). The process of combining meteorological drivers in the context of the response variable, such as with PCA, is likely to become an increasingly important step in an age where models are becoming larger in scope, and datasets more varied and numerous. Drought models forced with meteorological datasets without a contextually-based transformation or filtering built on the response variable(s) are likely to be less accurate than those with selected and pretreated input data. Machine learning-based approaches do not suffer the rigidity of process-based modeling approaches yet are sensitive to the datasets they are supplied. This opens the door to

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

potential future applications in different conditions, such as for modeling streamflow drought response to future climate scenarios.

One oft-cited concern with machine learning-based approaches comes from its characterization as a "black box" (Rudin, 2019; Welchowski et al., 2022). That is, the internal functions of a machine learning model are not easily interpretable. This can be an issue in situations where the model may be drawing misleading or inaccurate conclusions from the input data. Given this concern, our compartmentalization of the workflow into discrete steps provides several points along the workflow for analysis, interpretation, and adjustment. For example, this method provides a mechanism to examine and manipulate the specific inputs to the model (e.g., donor gages). For every set of predictions at ungaged locations using this approach, the modeler is able to analyze the change in predictions or predictive accuracy given a new set of donor gages. Through this permutation of donors, the importance of each donor in the model can be evaluated, providing crucial information on the relative importance of various model parameters on the final predictions.

5 Conclusion

In this study, we developed a machine learning-based methodology to advance understanding of the meteorological drivers of streamflow drought, as well as the land surface characteristics that influence its propagation to the basin outlet. We have also investigated the potential for this donor-based method as an approach for drought prediction in pseudo-ungaged basins using a dynamic regionalization. We conclude that the meteorological drivers of streamflow drought are highly interconnected, and the reorganization of these drivers with specific reference to streamflow drought through PCA transformation of random forest importance values provides contextually based insights into the influence of these drivers. Overall, we found that the greatest importance values across the CONUS were held by precipitation, soil moisture, and SPEI. From PCA, we found that, within PC1, teleconnections, temperature, and PET are highly interrelated in the context of streamflow drought. Similarly, we found from analysis of PC2 that the importance of precipitation, SPEI, and soil moisture to streamflow drought predictability are also highly interrelated at consistent smoothing window lengths. These PCs, along with the SWE- and soil-moisture-driven PC3, exhibit the strongest regional differences across the CONUS.

Additionally, we found regional differences in overlap of streamflow drought drivers, with the western CONUS subject to teleconnections, temperature, precipitation, and snow, and the Northeast region dominated by precipitation-related drivers alone. We also conclude static climatic and physiographic basin characteristics have an important and complex influence on drought propagation during the transitional period in the hydrologic cycle between the atmosphere and observed streamflow.

Finally, we found potential in the dynamic regionalization of a donor-based method for predicting streamflow drought at pseudo-ungaged locations. We found a comparable performance in predicting drought between an at-site random forest model, and a dynamically selected and weighted set of donors which share common drought response characteristics.

The results from this investigation provide new insights into the spatiotemporal distribution of influential streamflow drought drivers and may serve as a guiding framework towards future advances in streamflow drought prediction. Further, this method

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

can provide greater potential for improving the interpretability of machine learning by parsing the prediction-model donors out and allowing them to be analyzed individually.

Data availability and access

All data used or produced in the analyses described in this paper are publicly available and meet USGS Fundamental Science Practices (https://pubs.usgs.gov/circ/1367/). Daily streamflow percentiles and associated streamflow drought events are available from Simeone (2022). All associated model, input data, and output data have been archived according to USGS policy and are available at Heldmyer and Sando (2025).

USGS Disclaimer

This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the U.S. Geological Survey (USGS), it does not represent any official USGS finding or policy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author Contributions

AH prepared the original draft. AH and RS conceptualized the research, developed the methodology, and conducted the formal analysis. CS, MW, SH, PG, RM, JD, DW, BP, AS, KH, and JH reviewed and edited the published work. JH administered the project.

Acknowledgements

This research was funded by the U.S. Geological Survey (USGS) Water Availability and Use Science Program as part of the Water Resources Mission Area Data-Driven Drought Prediction Project. Computing resources were provided by USGS Cloud Hosting Solutions and USGS Core Science Systems. Thanks to reviewers for their suggestions and edits that improved this article. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

References

Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological applications and modelling, Int. J. Climatol., 33, 121–131, https://doi.org/10.1002/joc.3413, 2013.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., and Clark, M. P.: A Ranking of Hydrological Signatures Based on Their Predictability in Space, Water Resour. Res., 54, 8792–8812, https://doi.org/10.1029/2018WR022606, 2018.

Archer, K. J. and Kimes, R. V.: Empirical characterization of random forest variable importance measures, Comput. Stat. Data Anal., 52, 2249–2260, https://doi.org/10.1016/j.csda.2007.08.015, 2008.

Baek, S. H., Steiger, N. J., Smerdon, J. E., and Seager, R.: Oceanic Drivers of Widespread Summer Droughts in the United States Over the Common Era, Geophys. Res. Lett., 46, 8271–8280, https://doi.org/10.1029/2019GL082838, 2019.

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., and Schneider, D.: Snowmelt rate dictates streamflow, Geophys. Res. Lett., 43, 8006–8016, https://doi.org/10.1002/2016GL069690, 2016.

Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, Mon. Weather Rev., 115, 1083–1126, https://doi.org/10.1175/1520-0493(1987)115%253C1083:CSAPOL%253E2.0.CO;2, 1987.

Basara, J. B., Maybourn, J. N., Peirano, C. M., Tate, J. E., Brown, P. J., Hoey, J. D., and Smith, B. R.: Drought and Associated Impacts in the Great Plains of the United States—A Review, 2013, https://doi.org/10.4236/ijg.2013.46A2009, 2013.

Beven, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5, 1–12, https://doi.org/10.5194/hess-5-1-2001, 2001.

Biau, G. and Scornet, E.: A random forest guided tour, TEST, 25, 197–227, https://doi.org/10.1007/s11749-016-0481-7, 2016.

Bjerknes, J.: ATMOSPHERIC TELECONNECTIONS FROM THE EQUATORIAL PACIFIC, Mon. Weather Rev., 97, 163–172, https://doi.org/10.1175/1520-0493(1969)097%253C0163:ATFTEP%253E2.3.CO;2, 1969.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Broxton, P., Zeng, X., and Dawson, N.: Daily 4 km Gridded SWE and Snow Depth from Assimilated In-Situ and Modeled Data over the Conterminous US (1), https://doi.org/10.5067/0GGPB220EX6A, 2019.

Calle, M. L. and Urrea, V.: Letter to the Editor: Stability of Random Forest importance measures, Brief. Bioinform., 12, 86–89, https://doi.org/10.1093/bib/bbq011, 2011.

Carpenter, D. H. and Hayes, D. C.: Low-flow Characteristics of Streams in Maryland and Delaware, U.S. Geological Survey, 120 pp., 1996.

Clette, F., Cliver, E. W., Lefèvre, L., Svalgaard, L., and Vaquero, J. M.: Revision of the Sunspot Number(s), Space Weather, 13, 529–530, https://doi.org/10.1002/2015SW001264, 2015.

Cohen, J.: A Coefficient of Agreement for Nominal Scales, Educ. Psychol. Meas., 20, 37–46, https://doi.org/10.1177/001316446002000104, 1960.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Cook, B. I., Smerdon, J. E., Seager, R., and Cook, E. R.: Pan-Continental Droughts in North America over the Last Millennium, J. Clim., 27, 383–397, https://doi.org/10.1175/JCLI-D-13-00100.1, 2014.

Deweber, J. T., Tsang, Y.-P., Krueger, D. M., Whittier, J. B., Wagner, T., Infante, D. M., and Whelan, G.: Importance of Understanding Landscape Biases in USGS Gage Locations: Implications and Solutions for Managers, Fisheries, 39, 155–163, https://doi.org/10.1080/03632415.2014.891503, 2014.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J.: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S., Geophys. Res. Lett., 28, 2077–2080, https://doi.org/10.1029/2000GL012745, 2001.

Falcone, J. A.: GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow, U.S. Geological Survey, Reston, VA, 2011.

Feaster, T. D. and Lee, K. G.: Low-flow frequency and flow-duration characteristics of selected streams in Alabama through March 2014, Scientific Investigations Report, U.S. Geological Survey, https://doi.org/10.3133/sir20175083, 2017.

Fleiss, J. L., Levin, B., and Paik, M. C.: Statistical Methods for Rates and Proportions, 1st ed., Wiley, https://doi.org/10.1002/0471445428, 2003.

Floriancic, M. G., Berghuijs, W. R., Molnar, P., and Kirchner, J. W.: Seasonality and Drivers of Low Flows Across Europe and the United States, Water Resour. Res., 57, e2019WR026928, https://doi.org/10.1029/2019WR026928, 2021.

Frost, H. R., Li, Z., and Moore, J. H.: Principal component gene set enrichment (PCGSE), BioData Min., 8, 25, https://doi.org/10.1186/s13040-015-0059-z, 2015.

Gbedawo, V., Owusu Agyeman, G., Ankah, C., and Daabo, M.: An Overview of Computer Memory Systems and Emerging Trends, Am. J. Electr. Comput. Eng., 7, 19–26, https://doi.org/10.11648/j.ajece.20230702.11, 2023.

Gershunov, A. and Barnett, T.: Interdecadal modulation of ENSO teleconnections, Bull. Am. Meteorol. Soc., 79, 2715–2725, https://doi.org/10.1175/1520-0477(1998)079%253C2715:IMOET%253E2.0.CO;2, 1998.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, 2016.

Griffin, D. and Anchukaitis, K. J.: How unusual is the 2012–2014 California drought?, Geophys. Res. Lett., 41, 9017–9023, https://doi.org/10.1002/2014GL062433, 2014.

Hammond, J. C., Simeone, C., Hecht, J. S., Hodgkins, G. A., Lombard, M., McCabe, G., Wolock, D., Wieczorek, M., Olson, C., Caldwell, T., Dudley, R., and Price, A. N.: Going Beyond Low Flows: Streamflow Drought Deficit and Duration Illuminate Distinct Spatiotemporal Drought Patterns and Trends in the U.S. During the Last Century, Water Resour. Res., 58, e2022WR031930, https://doi.org/10.1029/2022WR031930, 2022.

Hamshaw, S., Goodling, P., Hafen, K., Hammond, J., McShane, R., Sando, R., Shastry, A., Simeone, C., Watkins, D., White, E., and Wieczorek, M.: Regional Streamflow Drought Forecasting in the Colorado River Basin using Deep Neural Network Models, Proc. 2023 SedHyd Conf. St Louis MO, 2023.

Heldmyer, A. and Sando, R.: Random forest and donor-based simulated streamflow drought for 3,198 gages across the CONUS, 1982-2020, https://doi.org/10.5066/P1FSRPMU, 2025.

Hotelling, H.: Analysis of a complex of statistical variables into principal components, J. Educ. Psychol., 24, 417–441, https://doi.org/10.1037/h0071325, 1933.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Hu, Z.-Z. and Huang, B.: Interferential Impact of ENSO and PDO on Dry and Wet Conditions in the U.S. Great Plains, J. Clim., 22, 6047–6065, https://doi.org/10.1175/2009JCLI2798.1, 2009.

Huang, S., Leng, G., Huang, Q., Xie, Y., Liu, S., Meng, E., and Li, P.: The asymmetric impact of global warming on US drought types and distributions in a large ensemble of 97 hydro-climatic simulations, Sci. Rep., 7, 5891, https://doi.org/10.1038/s41598-017-06302-z, 2017.

van Huijgevoort, M. H. J., Hazenberg, P., van Lanen, H. a. J., and Uijlenhoet, R.: A generic method for hydrological drought identification across different climate regions, Hydrol. Earth Syst. Sci., 16, 2437–2451, https://doi.org/10.5194/hess-16-2437-2012, 2012.

Jiang, S., Zheng, Y., Wang, C., and Babovic, V.: Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments, Water Resour. Res., 58, e2021WR030185, https://doi.org/10.1029/2021WR030185, 2022.

Jolliffe, I. T. (Ed.): Principal Component Analysis for Special Types of Data, in: Principal Component Analysis, Springer, New York, NY, 338–372, https://doi.org/10.1007/0-387-22440-8_13, 2002.

Kiang, J., Stewart, D., Archfield, S., Osborne, E., and Eng, K.: A National Streamflow Network Gap Analysis, 2013.

Kim, J. E., Yu, J., Ryu, J.-H., Lee, J.-H., and Kim, T.-W.: Assessment of regional drought vulnerability and risk using principal component analysis and a Gaussian mixture model, Nat. Hazards, 109, 707–724, https://doi.org/10.1007/s11069-021-04854-y, 2021.

Konapala, G. and Mishra, A.: Quantifying Climate and Catchment Control on Hydrological Drought in the Continental United States, Water Resour. Res., 56, e2018WR024620, https://doi.org/10.1029/2018WR024620, 2020.

Krabbenholft, C., Allen, G., Lin, P., Godsey, S., Allen, D., Burrows, R., DelVecchia, A., Fritz, K., Shanafield, M., Burgin, A., Zimmer, M., Datry, T., Dodds, W., Jones, N., Mimms, M., Franklin, C., Hammond, J., Zipper, S., Ward, A., Costigan, K., Beck, H., and Olden, J.: Assessing placement bias of the global river gauge network, Nat. Publ., 5, 586–592, https://doi.org/10.1038/s41893-022-00873-0, 2022.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T.: caret: Classification and Regression Training, 2023.

Landis, J. R. and Koch, G. G.: The measurement of observer agreement for categorical data, Biometrics, 33, 159–174, 1977.

Li, X., Khandelwal, A., Jia, X., Cutler, K., Ghosh, R., Renganathan, A., Xu, S., Tayal, K., Nieber, J., Duffy, C., Steinbach, M., and Kumar, V.: Regionalization in a Global Hydrologic Deep Learning Model: From Physical Descriptors to Random Vectors, Water Resour. Res., 58, e2021WR031794, https://doi.org/10.1029/2021WR031794, 2022.

Livneh, B. and Badger, A. M.: Drought less predictable under declining future snowpack, Nat. Clim. Change, 10, 452–458, https://doi.org/10.1038/s41558-020-0754-8, 2020.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Mainali, J. and Pricope, N. G.: High-resolution spatial assessment of population vulnerability to climate change in Nepal, Appl. Geogr., 82, 66–82, https://doi.org/10.1016/j.apgeog.2017.03.008, 2017.

Mallya, G., Zhao, L., Song, X. C., Niyogi, D., and Govindaraju, R. S.: 2012 Midwest Drought in the United States, J. Hydrol. Eng., 18, 737–745, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000786, 2013.

Mantua, N.: The Pacific Decadal Oscillation. A brief overview for non-specialists, Encycl. Environ. Change, 1999.

Manuel, J.: Drought in the Southeast: Lessons for Water Management, Environ. Health Perspect., 116, A168–A171, https://doi.org/10.1289/ehp.116-a168, 2008.

McCabe, G. J., Palecki, M. A., and Betancourt, J. L.: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States, Proc. Natl. Acad. Sci., 101, 4136–4141, https://doi.org/10.1073/pnas.0306738101, 2004.

McCabe, G. J., Wolock, D. M., Lombard, M., Dudley, R. W., Hammond, J. C., Hecht, J. S., Hodgkins, G. A., Olson, C., Sando, R., Simeone, C., and Wieczorek, M.: A hydrologic perspective of major U.S. droughts, Int. J. Climatol., 43, 1234–1250, https://doi.org/10.1002/joc.7904, 2023.

McDonald, R.: gdptools, 2022.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resour. Res., 41, https://doi.org/10.1029/2005WR004289, 2005.

Mishra, A. K. and Singh, V. P.: A review of drought concepts, J. Hydrol., 391, 202–216, https://doi.org/10.1016/j.jhydrol.2010.07.012, 2010.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J. Geophys. Res. Atmospheres, 109, https://doi.org/10.1029/2003JD003823, 2004.

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R. A., Carrao, H., Spinoni, J., Vogt, J., and Feyen, L.: Global Changes in Drought Conditions Under Different Levels of Warming, Geophys. Res. Lett., 45, 3285–3296, https://doi.org/10.1002/2017GL076521, 2018.

Nguyen, P.-L., Min, S.-K., and Kim, Y.-H.: Combined impacts of the El Niño-Southern Oscillation and Pacific Decadal Oscillation on global droughts assessed using the standardized precipitation evapotranspiration index, Int. J. Climatol., 41, E1645–E1662, https://doi.org/10.1002/joc.6796, 2021.

Nunes Carvalho, T. M., Lima Neto, I. E., and Souza Filho, F. de A.: Uncovering the influence of hydrological and climate variables in chlorophyll-A concentration in tropical reservoirs with machine learning, Environ. Sci. Pollut. Res., 29, 74967–74982, https://doi.org/10.1007/s11356-022-21168-z, 2022.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., and McIntyre, N.: Investigating regionalization techniques for large-scale hydrological modelling, J. Hydrol., 570, 220–235, https://doi.org/10.1016/j.jhydrol.2018.12.071, 2019.

Peña-Gallardo, M., Vicente-Serrano, S. M., Hannaford, J., Lorenzo-Lacruz, J., Svoboda, M., Domínguez-Castro, F., Maneta, M., Tomas-Burguera, M., and Kenawy, A. E.: Complex influences of meteorological drought time-scales on hydrological

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

droughts in natural basins of the contiguous Unites States, J. Hydrol., 568, 611–625, https://doi.org/10.1016/j.jhydrol.2018.11.026, 2019.

Piechota, T. C. and Dracup, J. A.: Drought and Regional Hydrologic Variation in the United States: Associations with the El Niño-Southern Oscillation, Water Resour. Res., 32, 1359–1373, https://doi.org/10.1029/96WR00353, 1996.

Pournasiri Poshtiri, M., Pal, I., Lall, U., Naveau, P., and Towler, E.: Variability patterns of the annual frequency and timing of low streamflow days across the United States and their linkage to regional and large-scale climate, Hydrol. Process., 33, 1569–1578, https://doi.org/10.1002/hyp.13422, 2019.

R Core Team: R: A language and environment for statistical computing, 2021.

Rao, J., Ren, R., Xia, X., Shi, C., and Guo, D.: Combined Impact of El Niño–Southern Oscillation and Pacific Decadal Oscillation on the Northern Winter Stratosphere, Atmosphere, 10, 211, https://doi.org/10.3390/atmos10040211, 2019.

Rice, J. S., Emanuel, R. E., Vose, J. M., and Nelson, S. A. C.: Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics, Water Resour. Res., 51, 6262–6275, https://doi.org/10.1002/2014WR016367, 2015.

Rudin, C.: Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nat. Mach. Intell., 1, 206–215, https://doi.org/10.1038/s42256-019-0048-x, 2019.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., and Zhong, C.: Interpretable machine learning: Fundamental principles and 10 grand challenges, Stat. Surv., 16, 1–85, https://doi.org/10.1214/21-SS133, 2022.

Schlögl, M. and Laaha, G.: Extreme weather exposure identification for road networks – a comparative assessment of statistical methods, Nat. Hazards Earth Syst. Sci., 17, 515–531, https://doi.org/10.5194/nhess-17-515-2017, 2017.

Seneviratne, S. I.: Historical drought trends revisited, Nature, 491, 338–339, https://doi.org/10.1038/491338a, 2012.

Shen, C., Chen, X., and Laloy, E.: Editorial: Broadening the Use of Machine Learning in Hydrology, Front. Water, 3, https://doi.org/10.3389/frwa.2021.681023, 2021.

Simeone, C., Foks, S., Towler, E., Hodson, T., and Over, T.: Evaluating Hydrologic Model Performance for Characterizing Streamflow Drought in the Conterminous United States, Water, 16, 2996, https://doi.org/10.3390/w16202996, 2024.

Simeone, C. E.: Streamflow Drought Metrics for Select United States Geological Survey Streamgages for Three Different Time Periods from 1921 - 2020, https://doi.org/10.5066/P92FAASD, 2022.

Singh, S., Abebe, A., Srivastava, P., and Chaubey, I.: Effect of ENSO modulation by decadal and multi-decadal climatic oscillations on contiguous United States streamflows, J. Hydrol. Reg. Stud., 36, 100876, https://doi.org/10.1016/j.ejrh.2021.100876, 2021.

Stoelzle, M., Stahl, K., Morhard, A., and Weiler, M.: Streamflow sensitivity to drought scenarios in catchments with different geology, Geophys. Res. Lett., 41, 6174–6183, https://doi.org/10.1002/2014GL061344, 2014.

Tahmasebi, P., Kamrava, S., Bai, T., and Sahimi, M.: Machine learning in geo- and environmental sciences: From small to large scale, Adv. Water Resour., 142, 103619, https://doi.org/10.1016/j.advwatres.2020.103619, 2020.

https://doi.org/10.5194/egusphere-2025-6064
Preprint. Discussion started: 21 January 2026

Tyralis, H., Papacharalampous, G., and Langousis, A.: A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources, Water, 11, 910, https://doi.org/10.3390/w11050910, 2019.

U.S. Geological Survey: USGS 01440400 Brodhead Creek near Analomink Pennsylvania, https://doi.org/10.5066/F7P55KJN, 2024a.

U.S. Geological Survey: USGS 09152500 Gunnison River near Grand Junction Colorado, https://doi.org/10.5066/F7P55KJN, 2024b.

Wang, S., Huang, J., He, Y., and Guan, Y.: Combined effects of the Pacific Decadal Oscillation and El Niño-Southern Oscillation on Global Land Dry–Wet Changes, Sci. Rep., 4, 6651, https://doi.org/10.1038/srep06651, 2014.

Welchowski, T., Maloney, K. O., Mitchell, R., and Schmid, M.: Techniques to Improve Ecological Interpretability of Black-Box Machine Learning Models, J. Agric. Biol. Environ. Stat., 27, 175–197, https://doi.org/10.1007/s13253-021-00479-7, 2022.

Williams, A. P., Seager, R., Abatzoglou, J. T., Cook, B. I., Smerdon, J. E., and Cook, E. R.: Contribution of anthropogenic warming to California drought during 2012–2014, Geophys. Res. Lett., 42, 6819–6828, https://doi.org/10.1002/2015GL064924, 2015.

Wlostowski, A. N., Jennings, K. S., Bash, R. E., Burkhardt, J., Wobus, C. W., and Aggett, G.: Dry landscapes and parched economies: A review of how drought impacts nonagricultural socioeconomic sectors in the US Intermountain West, WIREs Water, 9, e1571, https://doi.org/10.1002/wat2.1571, 2022.

Woodhouse, C. A., Pederson, G. T., Morino, K., McAfee, S. A., and McCabe, G. J.: Increasing influence of air temperature on upper Colorado River streamflow, Geophys. Res. Lett., 43, 2174–2181, https://doi.org/10.1002/2015GL067613, 2016.

Yang, R. and Xing, B.: Possible Linkages of Hydrological Variables to Ocean–Atmosphere Signals and Sunspot Activity in the Upstream Yangtze River Basin, Atmosphere, 12, 1361, https://doi.org/10.3390/atmos12101361, 2021.

Yeh, S.-W., Cai, W., Min, S.-K., McPhaden, M. J., Dommenget, D., Dewitte, B., Collins, M., Ashok, K., An, S.-I., Yim, B.-Y., and Kug, J.-S.: ENSO Atmospheric Teleconnections and Their Response to Greenhouse Gas Forcing, Rev. Geophys., 56, 185–206, https://doi.org/10.1002/2017RG000568, 2018.

Zhang, F., Biederman, J. A., Dannenberg, M. P., Yan, D., Reed, S. C., and Smith, W. K.: Five Decades of Observed Daily Precipitation Reveal Longer and More Variable Drought Events Across Much of the Western United States, Geophys. Res. Lett., 48, e2020GL092293, https://doi.org/10.1029/2020GL092293, 2021.

Zhu, P., Abramoff, R., Makowski, D., and Ciais, P.: Uncovering the Past and Future Climate Drivers of Wheat Yield Shocks in Europe With Machine Learning, Earths Future, 9, e2020EF001815, https://doi.org/10.1029/2020EF001815, 2021.

Zou, Z., Xiao, X., Dong, J., Qin, Y., Doughty, R. B., Menarguez, M. A., Zhang, G., and Wang, J.: Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016, Proc. Natl. Acad. Sci., 115, 3810–3815, https://doi.org/10.1073/pnas.1719275115, 2018.