Regional Base-Flow Index in Arid Landscapes Using Machine Learning and Instrumented Records

Caelum Mroczeka\*, Abraham E Springera, Neha Guptab, Temuulen Sankeyc, Benjamin Lucasd

a School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ, USA

b Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

c School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

d Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ, USA

Abstract

**Study Region**

This study focuses on Arizona, a dryland state in the southwestern United States characterized by pronounced variability in climate, elevation, and hydrogeology. Arizona encompasses two physiographic regions, the Colorado Plateau and the Basin and Range, each exhibiting distinct hydrologic behavior. The state’s arid to semi-arid conditions, combined with limited surface water, make groundwater discharge a vital component of streamflow. However, sparse streamgage coverage limits the ability to assess base-flow contributions across diverse catchments.

**Study Focus**

We quantify long-term base-flow index (BFI) patterns and trends across Arizona and develop a predictive framework for ungauged basins. BFI was first calculated at 205 USGS streamgages using a recursive digital filter applied to multi-decadal streamflow records. We trained an eXtreme Gradient Boosting (XGBoost) model using physiographic, hydroclimate, and land surface data to predict long-term BFI at the 8-digit HUC scale.

**New Hydrological Insights for the Region**

Approximately 32% of Arizona’s streamflow is sustained by groundwater, with considerable spatial variability influenced by elevation, land cover, and climate. High BFI values occur in spring-fed and snowmelt-driven headwaters, while low values dominate arid lowlands. Declining BFI trends are most common in monsoon-dominated, warm-dry, and low-slope regions. Precipitation trends align most closely with BFI trends, highlighting the climate sensitivity of dryland base flow. This framework improves understanding of groundwater–surface water interactions and offers a transferable tool for data-scarce dryland regions globally.

Keywords: Base Flow, Groundwater–Surface Water Interactions, Dryland Hydrology, Machine Learning, Ungauged Catchments

## Introduction

Dryland regions, encompassing arid, semi-arid, hyper-arid, and dry sub-humid systems, account for 40% of the Earth’s land surface. These regions are home to approximately 2 billion people globally and constitute the largest terrestrial biome (IUCN 2019). Despite supporting diverse ecosystems and human populations, dryland regions face mounting hydrologic challenges exacerbated by increasing urbanization, expanding agricultural activities, and climate-induced amplification of precipitation patterns (Taylor et al. 2013). This water scarcity is intensifying due to the compounding effects of climate variability and increased groundwater extraction (Taylor et al. 2013). Groundwater serves as a vital resource in drylands for sustaining ecological functions and supporting human livelihoods (Scanlon et al. 2006; Yao et al. 2018).

Base flow is the sustained portion of streamflow in the absence of runoff that is derived from groundwater discharge (USGS 2018). Base flow is critical to maintaining seasonal low-flow regimes, supporting aquatic ecosystems, and facilitating the transport of nutrients and chemicals. Base-flow contribution to streamflow can be highly variable spatially (Singh et al. 2018; Bosch et al. 2017; Beck et al. 2013), and temporally (Ficklin et al. 2016; Tan et al. 2020). Increasing groundwater extraction, changes in land cover/land use, and changes in precipitation patterns due to climate change affect the timing and volumes of base flow (Tan et al. 2020; Taylor et al. 2013). Effective management of water quantity and quality depends on understanding seasonal and interannual base-flow patterns and long-term changes in base-flow behavior.

The Base-Flow Index (BFI) is the ratio of the long-term mean base-flow volume to the long-term total streamflow volume expressed as a percentage. BFI serves as a normalized measure of groundwater contribution interannually or between basins. BFI is determined by hydrograph separation and is influenced by the climate and physiographic characteristics of a catchment (Neff et al. 2005; Beck et al. 2013; Singh et al. 2018). Between catchments, base flow fluctuates according to changes in the moisture content of the vadose zone, influenced by varying levels of evapotranspiration and aquifer storage dynamics (Bosch et al. 2017). Since BFI calculations rely on instrumented stream records, it remains unknown for ungauged catchments, which encompass most of the earth’s land surface (Fekete et al. 2007). Addressing this information gap is integral to approaching a comprehensive understanding of groundwater dynamics globally.

Advancements in machine learning provide tools to predict hydrologic indices in ungauged basins, addressing the limitations of sparse streamgage networks. To tackle the challenge of quantifying base flow in ungauged catchments, numerous studies have applied both regression and machine learning methods. Ahiablame et al. (2013) found that using a regression model to estimate annual base flow of ungauged catchments was reasonably easy and accurate. Beck et al. (2013) overcame the nonlinearity of basin characteristics and improved results of multivariate analyses by using artificial neural networks (ANN) to estimate BFI globally. Singh et al. (2018) implemented a random forest algorithm to predict long-term BFI for ungauged catchments across New Zealand. These applications demonstrate the versatility and effectiveness of machine learning in capturing complex ecohydrologic dynamics and improving our understanding of groundwater contributions to streamflow.

Previous studies have examined base-flow regionalization and synthesis across various spatial scales, from global to continental (Beck et al. 2013; Santhi et al. 2008; Ayers et al. 2022; Singh et al. 2018). Such large-scale analyses often utilize generalized datasets and methodologies, resulting in limited applicability to regions with unique hydrogeologic and climatic conditions. Additionally, global and continental-scale studies tend to rely on streamgage networks that disproportionately represent large perennial rivers and regulated watersheds with dense human populations, while underrepresenting arid and semi-arid regions characterized by non-perennial flow regimes and smaller streams (Krabbenhoft et al. 2022). Thus, their effectiveness in accurately capturing groundwater-surface water interactions, particularly in critically water-stressed dryland regions, remains constrained.

Despite advancements, existing large-scale studies have limited applicability to the unique hydrogeologic conditions in arid and semi-arid regions like Arizona, leaving uncertainty regarding groundwater-surface water interactions in these critically water-stressed areas. This study addresses this crucial gap by developing a regionally tailored machine learning model specifically designed to estimate BFI for Arizona’s ungauged basins. Utilizing hydrogeologic characteristics and hydroclimatic data, we estimate long-term mean BFI (1991–2020) in ungauged catchments, filling spatial gaps in the sparse streamgage network. Furthermore, regional trends in base flow and BFI at instrumented sites are analyzed alongside coincident climate trends in precipitation, reference evapotranspiration (ETO), and temperature. The outcomes offer novel insights into groundwater contributions to streamflow, supporting water resource management in an area increasingly vulnerable to drought and climate change impacts, and contributing broadly to the understanding of hydrologic processes in dryland environments.

## Methods

### Study Area

Arizona, located in the southwestern United States, spans approximately 295,253 km² and encompasses a diverse range of landscapes, elevations, and climate regimes. The state includes portions of three major physiographic regions: the Colorado Plateau in the northeast, the Basin and Range province in the south and west, and the Central Highlands, a transitional zone between them. This heterogeneity results in substantial variation in environmental conditions across the state. The Colorado Plateau is characterized by high-elevation desert and mountain woodlands, averaging 1,936 masl (6,352 ft), with mean temperatures ranging from -6°C (20°F) to 26°C (80°F) and annual precipitation of about 580 mm (23 in). In contrast, the Basin and Range region is lower in elevation, averaging 490 masl (1,600 ft), and features a semi-arid to arid climate, with temperatures ranging from 15°C (60°F) to 43°C (110°F) and an average annual precipitation of 200 mm (8 in) (Arizona State Climate Office 2024). The Central Highlands feature a mix of mountainous terrain and interspersed basins, adding to the state’s topographic and hydroclimatic complexity.

|  |
| --- |
| Figure 1: Map of Arizona and US Geological Survey (USGS) streamgages used in this study. 8-digit HUC subbasin boundaries and physiographic regions shown. |

Arizona’s hydrology varies seasonally and spatially between its physiographic regions. In summer, localized and intense convective storms stem from the North American Monsoon; in winter, orographic precipitation is delivered by Pacific frontal systems (Eastoe et al. 2019). While monsoonal precipitation can account for up to 50% of annual precipitation, evaporation and dry preceding soil properties leads to most precipitation becoming runoff (Sheppard et al. 2002). As such, 94% of streams in Arizona are ephemeral or intermittent (Levick et al. 2008). Much of the hydrology of Arizona is snow-melt derived, driven by spring melt from the high-elevation Colorado Plateau winter snowpack. Though winter precipitation provides only 30% of annual averages, it provides the majority of water for natural reservoirs (Sheppard et al. 2002).

### Data

Daily observed streamflow data obtained from the United States Geological Survey (USGS) National Water Information System (NWIS) were used in this study. Streamgages were selected depending on criteria to ensure the applicability of each site. Following the findings of O’Donnell et al. (2016), which determined that 8–10 years of calibration data are necessary to account for climate variability in paired watershed studies in the region, a minimum record length of 10 years was required. Additionally, years with more than 30 missing days of streamflow data were excluded from the analysis. This study focuses on natural, streamflow-influencing dynamics, so streamgages affected by regulation or diversions were excluded. The regulated river streamgages were identified through annual reports on water data published by the USGS (USGS 2010). Furthermore, streamgages along the Colorado River were omitted because they represent managed flows governed by the Colorado River Compact. After applying these selection criteria, 205 USGS streamgages with acceptable periods of record were included in the study ([Figure 1](#fig-study-area)). Periods of record ranged from 10 to 112 years, with a median of 28 years.

Our data selection criteria ensure a robust analysis but also highlight notable spatial gaps that our machine learning model can address. Arizona has 184 active USGS streamflow stations (as of 2024) covering an area of 295,253 km². For comparison, Indiana (a humid state in the U.S.) maintains 189 active stations within a significantly smaller area of 94,326 km². This results in a streamgage density of approximately 2.004 gauges per 1,000 km² in Indiana, more than three times greater than Arizona’s density of 0.623 gauges per 1,000 km². Such disparities in gauge coverage are typical for dryland regions globally (Krabbenhoft et al. 2022), underscoring the necessity and relevance of this type of modeling approach in arid and semi-arid environments.

Watersheds across the United States are delineated by the USGS using a hydrologically-defined network. This system delineates the country using hierarchical hydrologic unit codes (HUCs), where each subsequent basin includes the digits of the enclosing basin. Here, 8-digit HUCs (HUC 8s) are used to divide Arizona into 84 sub-basins that are fully or partially in the state ([Figure 1](#fig-study-area)). These HUC 8 sub-basins are analogous to medium-sized river basins and are defined by surface water characteristics.

Annual precipitation and temperature data came from the PRISM climate group at Oregon State University at a resolution of 4 km (<https://prism.oregonstate.edu;> (Daly et al. 2008)). The PRISM dataset provides valuable insights into regional climate in ungauged regions and has been shown to perform well across the southwestern US (Buban et al. 2020). Instead of the water year, PRISM data uses a calendar-year format, which was adopted for consistency in the water balance. Although this may introduce challenges in the annual estimates due to inter-annual snow storage, the use of long-term annual averages reduces potential errors (Reitz et al. 2017).

Annual reference evapotranspiration (ETO) data came from TerraClimate, a 4-km grid climatological data set (Abatzoglou et al. 2018). TerraClimate uses a Penman-Monteith approach to generate a reference evapotranspiration. The ETO values were calculated assuming a reference grass surface across the landscape with unlimited water. In the drylands of the southwestern US, ETO typically exceeds precipitation annually (Zomer et al. 2022).

A 30-meter resolution Digital Elevation Model (DEM) of Arizona was used to derive key basin characteristics: basin area, average slope, and the proportion of each basin oriented toward north or south aspects. Various geospatial variables, such as aspect, were disaggregated then averaged to assess the areal percentage of each sub-variable within individual HUC 8 basins. By calculating these percentages, we derived a more comprehensive understanding of landscape composition across space. Land cover was acquired from USGS-NLCD (National Land Cover Database), hydrologic soil group from SSURGO (Soil Survey Geographic Database), and underlying geology and karst from USGS were all similarly averaged across the basins. Aggregating variables to align with the HUC 8 boundaries allowed for more precise predictions of BFI by integrating spatial variations within each basin.

### Base-flow separation

Directly measuring base flow and BFI presents unique challenges (Eckhardt 2008). The technique chosen to separate base flow has been shown to affect results, and the choice of base-flow separation method is subjective since ‘true’ BFI values are not known (Beck et al. 2013). However, many methods have been developed to estimate these values. These methods include the use of tracers (Gonzales et al. 2009), graphical interpolation (Hydrology 1980; Sloto et al. 1996), and digital filters (Arnold et al. 1995; Eckhardt 2005; Lyne et al. 1979; Nathan et al. 1990). These techniques have varying levels of applicability depending on the spatial scale, time span, and the scope of the study. Comparisons of various base-flow separation techniques have been made in previous studies (e.g. Eckhardt 2005, 2008; Nathan et al. 1990); this study does not explore the superiority of different methods.

Base flow was calculated using a single-parameter, recursive digital filter technique from Nathan et al. (1990). This base-flow separation technique is based on a recursive digital filter used in signal analysis that separates high-frequency signals (quickflow) from low-frequency signals (base flow) (Lyne et al. 1979). Eckhardt (2023) noted that recursive digital filters lack a physical basis, but as the method is easy to automate, objective, and repeatable, it is appropriate for a regional-scale study. The Lyne-Hollick filter has been used in multiple studies (Arnold et al. 2000; Santhi et al. 2008; Bloomfield et al. 2009; Singh et al. 2018), and it takes the form of

where is base flow, is the filter parameter, is the total streamflow, and is the time step. A filter parameter of 0.925 was used as in Nathan et al. (1990) and Fuka et al. (2014). The filter was run three times (forward, backward, forward) to attenuate the base-flow signal.

### Machine Learning

The implementation of machine learning models to predict hydrologic indices has been successful in past studies (Singh et al. 2018; Schmidt et al. 2020; Rozos et al. 2021). In this work, we used the eXtreme Gradient Boosting (XGBoost) algorithm (Chen et al. 2016) to predict BFI at ungauged locations using catchment characteristics as predictors [Table 1](#tbl-predictors). The XGBoost algorithm is a decision tree-based ensemble algorithm, which can be adapted for either regression or classification problems. This algorithm iteratively builds an ensemble of decision trees, where each tree corrects errors from previous trees to improve predictions (Chen et al. 2016). Its efficiency, scalability, and robustness have made it increasingly popular in recent years, with successful applications in environmental modeling tasks such as streamflow forecasting (Szczepanek 2022; Ni et al. 2020) and land use/land cover classification (Georganos et al. 2018).

XGBoost operates by leveraging gradient boosting on decision tree algorithms, combining multiple low-variance models to produce a robust overall prediction. Gradient boosting works iteratively: the initial tree is trained on the target values, while subsequent trees are trained on the residual errors of the preceding tree. Each tree is assigned a weight based on its contribution to reducing error, and these weights are used to determine the influence of each tree in the final model. The ultimate prediction is made by aggregating the outputs of all weighted trees in the ensemble. In this study, the trained XGBoost model is used to predict BFI in ungauged catchments based on geospatial and hydroclimate predictor variables ([Table 1](#tbl-predictors)). Certain features were further subdivided according to their areal coverage within each basin (e.g. land cover was divided into 16 subdivisions). This approach allowed the model to capture finer-scale spatial variability and improve predictive accuracy.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Basin-characteristic variables used as initial features in XGBoost model. Starred features are maintained in the final, dimensionality-reduced model.   |  | Variable | Source | Geoprocessing | | --- | --- | --- | --- | | Hydroclimate | Precipitation\* | PRISM | Basin average | |  | Mean Temperature\* | PRISM | Basin average | |  | Reference Evapotranspiration\* | TerraClimate | Basin average | | Geospatial | Elevation\* | DEM | Basin average | |  | Area | DEM | Basin average | |  | Slope | DEM | Basin average | |  | Aspect | DEM | Percent areal coverage | |  | Land Cover\* | NLCD | Percent areal coverage | |  | Hydrologic Soil Group\* | SSURGO | Percent areal coverage | |  | Geology | USGS | Percent areal coverage | |  | Karst | USGS | Percent areal coverage | |

Our initial training dataset comprised 7,724 observations across 45 variables. To optimize the model’s performance, we first conducted an exhaustive grid search combined with 5-fold cross-validation to identify the optimal hyperparameter values. The hyperparameters evaluated included the learning rate (), minimum split loss (), maximum tree depth, minimum child weight, and the number of trees. While not an exhaustive list of all possible XGBoost hyperparameters, this range of values provided sufficient variation to ensure the selection of a high-performing model. The optimal hyperparameters were determined to be: 700 trees, a learning rate () of 0.05, a minimum split loss () of 0.075, a maximum tree depth of 7, and a minimum child weight of 5. Using these values, the XGBoost model was trained on the dataset with 10-fold cross-validation.

-fold cross-validation provides an unbiased estimate of a model’s accuracy on unseen data, while also insuring against overfitting or underfitting. In this approach, the data are randomly divided into folds of equal size. The model is trained on folds and tested on the remaining fold, referred to as the validation set. This process is repeated times, with each fold serving as the validation set exactly once. Each iteration trains an independent model with the same hyperparameters but using a different subset of training data. By averaging the model performance across all folds, we achieve a robust and reliable estimate of predictive accuracy. Root mean squared error (RMSE) was used as the performance metric for both model optimization and evaluation. RMSE provided a consistent and interpretable measure of the model’s accuracy throughout the training and validation process.

|  |
| --- |
| Figure 2: Schematic of -fold cross validation. The dataset is randomly divided into stratified folds. Each fold serves as the validation set once, while the remaining folds are combined to create a training set for model development. Performance metrics for the test set are calculated and recorded, and this process is repeated for all folds. |

#### Feature Selection

Machine learning models are prone to overfitting, especially when provided with a large set of predictive features (Ying 2019). Overfitting can degrade model performance on unseen data and increase the demand for computational resources and memory storage (Li et al. 2017). Dimensionality reduction offers a robust solution to these challenges and generally falls into two broad categories: feature extraction and feature selection.

Feature extraction involves transforming the original dataset into a lower-dimensional feature space. However, this process generates new features that often lack the physical interpretability of the original variables. In contrast, feature selection identifies a subset of the original features, preserving their physical meaning while improving model readability and interpretability (Li et al. 2017). In this study, supervised feature selection was employed to reduce the number of predictors, which enhanced learning performance, reduced computational costs, and mitigated overfitting.

To begin, an initial model was trained using the full feature set of 45 predictors ([Table 1](#tbl-predictors)). A feature selection method based on feature importance was then applied to identify and remove less relevant and noisy features. Feature importance scores quantify the contribution of individual features—either positively or negatively—to the model’s predictions (Murdoch et al. 2019). In this analysis, SHAP (SHapley Additive exPlanations) values were used to compute feature importance scores (Lundberg and Lee 2017).

SHAP values is a method to explain the prediction of an individual instance by calculating the contribution of each feature to that prediction. The method is based on coalition game theory and is discussed further in Lundberg and Lee (2017). Here, SHAP values are used for global interpretation of feature importance and feature effects on the model. Global feature importance is produced by the absolute Shapley values of each feature across the dataset, providing a list of features in order of most to least important. Feature effects provide an indication of the relationship between the value of a predicting feature and its impact on the prediction.

The ten most important features [Figure 6](#fig-shap_values) were selected based on their SHAP values and used to train a subsequent model [Table 1](#tbl-predictors). This refined model demonstrated improved performance and reduced computational time compared to the initial model. The final model, trained on this optimized feature subset, was ultimately used for the analysis presented here.

### Statistical Analyses

Statistical analyses were conducted on annual BFI and base-flow values from instrumented streamgages to identify temporal trends using the Mann-Kendall nonparametric trend test (Kendall 1970; Mann 1945). This test detects monotonic trends in datasets that are non-parametric and assumes the absence of autocorrelation among observations. This test is widely used in studies of this nature (Ficklin et al. 2016; Ayers et al. 2019; Woodhouse et al. 2022).

To check for autocorrelation, we applied the Durbin-Watson test (Durbin and Watson 1950), which revealed significant autocorrelation at four streamgages on an annual basis. Of these, only one streamgage (09486500 - Santa Cruz River at Cortaro, AZ) showed a significant trend in BFI. This streamgage was excluded from the trend analysis, as autocorrelation could inflate the variance of the Mann-Kendall statistic, potentially leading to biased trend estimates (Hamed and Rao 1998). Trends with a are considered significant.

## Results

### Observed BFI in Gauged Catchments

The long-term BFI for the 205 gauged reaches across Arizona is illustrated in [Figure 3](#fig-instrumented-bfi) . The long-term mean BFI is 0.32, indicating that ~32% of long-term streamflow in Arizona likely originates from groundwater discharge and other delayed sources. The highest BFI values (>0.9) are found along the Grand Canyon in northwestern Arizona, where highly karstic geology facilitates rapid subsurface flow to surface water and spring outlets (Chambless et al. 2023). High BFI values (>0.8) are also found at the spring-fed headwaters of the Verde River and Fossil Creek, which have similar highly-karstic, snowmelt-driven recharge areas.

The stream reaches of the Little Colorado River Basin (northeastern Arizona) indicate consistently low BFI values (< 0.2). This is likely due to low-yielding perched aquifers underlying the Defiance Plateau in northeastern Arizona, which are hydrologically connected to surface streams, while the high-yield, confined regional aquifer is much deeper (Blanchard 2002).

|  |
| --- |
| Figure 3: Long-term BFI for the period of record from instrumented stream flow data. |

#### Trends in Base Flow and BFI

Trends in BFI over the period of record for each streamgage are illustrated in [Figure 4](#fig-instrumented-trend) and summarized in [Table 2](#tbl-trends). Base flow and BFI trends were analyzed across all instrumented sites over their respective periods of record using the Mann–Kendall test. Statistically significant trends were observed in both metrics, with a 72.2% coincidence rate between significant base flow and BFI trends, indicating a strong dependence of BFI on base-flow dynamics.

[Figure 4](#fig-instrumented-trend) illustrates the spatial variation in BFI trends across the study area. Statistically significant decreasing trends are observed at 16.1% of sites, while increasing trends are found at 8.8% of sites. No consistent regional patterns are evident.

|  |
| --- |
| Figure 4: Trends in BFI over full period of record for instrumented sites used in this study. Red downward (blue upward) arrows indicate a decreasing (increasing) trend at a significance level of 5%. White circles represent sites with no statistically significant trends. |

#### Classification Trends

Classifications presented in [Table 2](#tbl-trends) were determined based on precipitation regime, physiographic region, climate, and slope. The dominant precipitation regime (monsoon vs. snowmelt) was identified by analyzing streamflow hydrographs for each station, focusing on peak flow periods during the monsoon season (July–September) and the snowmelt season (March–June). Physiographic region was assigned based on which region the streamgage is located. Climate classifications were defined as warm (above the long-term median temperature of Arizona), cool (below the long-term median temperature), wet (above the long-term median precipitation), and dry (below the long-term median precipitation). Slope was categorized as high (above the median slope) and low (below the median slope).

Statistically significant decreasing trends in BFI were more common than increasing trends across all site classifications ([Table 2](#tbl-trends)). While decreasing trends dominate, both increasing and decreasing trends are observed within each classification. Monsoon-dominated regions exhibit a higher proportion of significant negative trends (24.1%) compared to snowmelt-dominated regions (10.2%), suggesting that monsoon-dominated systems are more consistently correlated with declining base flow. Among climate classifications, warm-dry climates have the highest proportion of negative trends (20.0%), followed by warm-wet climates (19.4%), indicating that regions with higher temperatures are more prone to base flow declines. Low-slope regions show a greater prevalence of negative trends (20.4%) compared to high-slope regions (11.8%). This suggests that flatter areas may be more susceptible to base-flow reductions, potentially due to differences in hydrologic connectivity and recharge dynamics.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Comparison of trends for BFI for all sites split by various classifications. Only sites with a significant () trend are included here as established by a Mann-Kendall test for monotonic trends across the full period of record. n is the number of sites, n\_pos (n\_neg) is the number of sites with positive (negative) trends, perc\_pos (perc\_neg) is the percentage of n with a positive (negative) trend.   | **Classification Group** | **n** | **n\_pos** | **n\_neg** | **perc\_pos** | **perc\_neg** | | --- | --- | --- | --- | --- | --- | | Precipitation - Monsoon Dominated | 87 | 8 | 21 | 0.092 | 0.241 | | Precipitation - Snowmelt Dominated | 118 | 9 | 12 | 0.076 | 0.102 | | Physiographic Region - Basin and Range | 156 | 14 | 26 | 0.090 | 0.167 | | Physiographic Region - Colorado Plateau | 49 | 3 | 7 | 0.061 | 0.143 | | Climate - Warm-Wet | 31 | 2 | 6 | 0.065 | 0.194 | | Climate - Warm-Dry | 55 | 6 | 11 | 0.109 | 0.200 | | Climate - Cool-Wet | 74 | 4 | 9 | 0.054 | 0.122 | | Climate - Cool-Dry | 45 | 5 | 7 | 0.111 | 0.156 | | Slope - High | 102 | 10 | 12 | 0.098 | 0.118 | | Slope - Low | 103 | 7 | 21 | 0.068 | 0.204 | |

#### Coincident Climate Trends

Coincident trends are defined as those in which base flow or BFI trends match in direction with climate variable trends over the same period. Results are summarized in [Table 3](#tbl-coincident-trends). This analysis includes both significant and non-significant trends, which is appropriate where the influence of complex, interconnected processes may not always manifest as statistically significant patterns over limited observational periods (Ficklin et al. 2016). Coincidence was highest with precipitation (64.88% for base flow; 53.17% for BFI), followed by ETO and then temperature. Temperature trends most often show an inverse relationship—i.e., warming associated with declining BFI and base flow.

Given the variations in the period of record across the instrumented network (see Figure S2), we analyzed trends in base flow and BFI in relation to coincident trends in climate variables. Trends are classified as coincident when the direction of the climate variable trend (positive or negative) aligns with the trend observed in base flow or BFI ([Table 3](#tbl-coincident-trends)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Coincident trends of climate variables (ETO, precipitation, temperature) with base flow and BFI trends.   |  | **Climate Variable** | **Coincidence Percentage (%)** | | --- | --- | --- | | BFI | ETO | 44.88 | |  | Precipitation | 53.17 | |  | Temperature | 47.32 | | Base Flow | ETO | 55.12 | |  | Precipitation | 64.88 | |  | Temperature | 39.02 | |

### BFI of Ungauged Catchments

#### Model Validation

Predicted values of BFI are plotted against observed values for the entire period of record of the instrumented dataset in [Figure 5](#fig-actual_predicted) . The agreement between “out-of-bag” predictions (blind cross-validation, treating each site as ungauged) and observed values indicate strong model performance across the full dataset (R2 = 0.764). The overall RMSE is 0.129 and the overall percent bias (pbias) is -5.6%.

Model performance metrics across various classifications are summarized in [Table 4](#tbl-performance). These metrics show consistent model performance across spatial and climatic classifications. However, the negative pbias values across all classifications indicate a systematic underprediction of BFI. Categories with relatively lower R2 and Nash-Sutcliffe Efficiency (NSE) values also exhibit higher biases, reflecting weaker model performance in those contexts.

|  |
| --- |
| Figure 5: Linear relationship between observed BFI and predicted BFI. The solid line is the 1:1 line, the dashed, green line is regressed to the data. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Performance of model predictions for BFI for all sites split by various classifications. n is number of observations, R2 is the coefficient of determination of a linear regression, MSE is mean-squared-error, RMSE is root-mean-squared-error, MAE is mean-absolute-error, NSE is Nash-Sutcliffe efficiency, and pbias is percent bias.   | **Classification Group** | **n** | **R2** | **MSE** | **RMSE** | **MAE** | **NSE** | **pbias** | | --- | --- | --- | --- | --- | --- | --- | --- | | Climate - Monsoon Dominated | 3039 | 0.633 | 0.016 | 0.126 | 0.074 | 0.619 | -13.7 | | Climate - Snowmelt Dominated | 4685 | 0.733 | 0.015 | 0.121 | 0.087 | 0.725 | -3.5 | | PhysRegion - Basin&Range | 6147 | 0.733 | 0.016 | 0.127 | 0.084 | 0.724 | -6.3 | | PhysRegion - CO Plateau | 1577 | 0.846 | 0.011 | 0.104 | 0.073 | 0.843 | -3.8 | | Climate - Warm-Wet | 1506 | 0.693 | 0.014 | 0.117 | 0.077 | 0.685 | -8.2 | | Climate - Warm-Dry | 2351 | 0.693 | 0.022 | 0.147 | 0.092 | 0.675 | -11.9 | | Climate - Cool-Wet | 2350 | 0.738 | 0.011 | 0.106 | 0.078 | 0.736 | -1.7 | | Climate - Cool-Dry | 1517 | 0.831 | 0.012 | 0.111 | 0.078 | 0.827 | -4.3 | | Slope - High | 3795 | 0.776 | 0.012 | 0.111 | 0.079 | 0.771 | -3.3 | | Slope - Low | 3929 | 0.724 | 0.018 | 0.133 | 0.085 | 0.713 | -9.1 | |

#### Predictor Importance

The predictors used to estimate BFI at ungauged sites were evaluated for their importance in the final XGBoost model, as illustrated in [Figure 6](#fig-shap_values). The most influential feature for predicting long-term BFI is basin elevation. While elevation itself does not directly affect base-flow characteristics, it has consistently been identified as a key predictor in previous BFI studies (Singh et al. 2018; Beck et al. 2013). The importance of elevation aligns with findings from Beck et al. (2013), highlighting its role as a proxy for climate variables such as temperature, precipitation, and snowpack duration. Seasonal snowpack duration, in particular, has been shown to strongly correlate with springflow and groundwater recharge in this region (Donovan et al. 2022).

|  |
| --- |
| Figure 6: SHAP value plot of features used in final model. Land cover features are indicated by the percentage of cover by each land cover type and soil types are defined by hydrologic soil group. |

Land cover and land use predictors also play a significant role in BFI estimation. Analysis of SHAP values indicates that a higher percentage of evergreen forest positively influences BFI predictions, while higher proportions of shrubland and developed land exert a negative influence. Similarly, hydrologic soil types show distinct trends in their impact on BFI. Soil Type C, characterized by moderately high runoff potential (20-40% clay), tends to negatively influence BFI. In contrast, Soil Type A, which has low runoff potential and facilitates rapid water infiltration, exhibits a mixed influence (USDA 2009).

#### Predicted BFI in Ungauged Basins

The regionalized (HUC-8) long-term BFI (1991–2020) is shown in [Figure 7](#fig-bfi-huc). Basins with high BFIs, such as those along the Grand Canyon in the northwestern part of the study area, indicate greater surface water and groundwater interaction. Elevated BFI values are also observed along portions of the Mogollon Rim, a heavily forested region with high precipitation that marks the transition between physiographic regions. Additionally, headwater regions of perennial rivers tend to exhibit higher BFI values. In contrast, low BFI values are found in areas like the Defiance Plateau in northeastern Arizona and the arid southern regions of the state.

|  |
| --- |
| Figure 7: Predicted long-term BFI values for 8-digit HUC (1991-2020) |

## Discussion

Previous studies have examined base-flow regionalization and synthesis across various spatial scales, some overlapping with our study area. Beck et al. (2013) explored global patterns of BFI using streamflow observations extrapolated through an ANN. The U.S. streamgages included in that analysis were drawn from the interim Model Parameter Estimation Experiment (MOPEX) dataset of 1862 USGS gauges (Beck et al. 2013; Schaake, Cong, and Duan 2006). Only 10 of these interim gauges overlap geographically with our study area. Similarly, Ayers et al. (2022) assessed climatic influences on monthly base flow across the continental United States (CONUS), utilizing a subset of GAGES-II streamflow gauges. Although exact gauge locations from that study are unavailable, published figures suggest minimal geographic overlap with our research domain. Santhi et al. (2008) employed interpolation methods for regional base-flow estimation across CONUS, but specific gauge locations or data are not available, limiting direct regional comparisons.

The existing streamgage networks Arizona are limited for base-flow analyses by their design and monitoring objectives. While the number of streamgages has increased to meet regulatory imperatives, such as those driven by the Clean Water Act, many gauges prioritize peak flow monitoring rather than low-flow conditions critical for base-flow studies (Maricopa County 2020). Newer non-USGS gauges, such as those installed by flood control districts (e.g., the ALERT system), are generally tailored for flood detection rather than monitoring sustained base flow. Recent USGS installations focus on addressing in-stream flow rights and future monitoring improvements, though low-flow capabilities remain limited.

This study addresses these monitoring gaps through an analysis of observed base flow and BFI at instrumented streamgages, highlighting groundwater-surface water interactions within Arizona’s arid and semi-arid landscapes. Observed spatial patterns in BFI clearly align with known hydrogeological and climatic variations. High BFI values coincide with areas characterized by spring-fed streams and pronounced physiographic transitions, such as the Grand Canyon and Mogollon Rim. Conversely, lower observed BFI values correspond to regions with limited groundwater contributions, notably the Defiance Plateau and southern Arizona, emphasizing vulnerability in these arid environments. A trend of high BFI in upstream reaches to lower BFI in downstream reaches [Figure 3](#fig-instrumented-bfi) likely indicates a transition from zones of groundwater discharge to zones of groundwater recharge and may indicate a transition from gaining stream reaches to losing stream reaches (Winter 2007).

Analysis of trends at instrumented sites revealed a predominance of declining BFI values, particularly within monsoon-dominated precipitation regimes, warm-dry climates, and low-slope basins. These negative trends likely reflect intensified climate stress, highlighting areas most vulnerable to groundwater depletion and surface-water scarcity. Precipitation was identified as the primary driver influencing base-flow variability, with temperature and evapotranspiration trends adding complexity. The weaker correlation between temperature and BFI trends suggests that temperature impacts are moderated by other environmental factors, underscoring the need for integrated climate and land-management strategies to protect groundwater-dependent ecosystems.

Given the limited coverage provided by existing gauges, a common issue in arid and semi-arid regions globally (Krabbenhoft et al. 2022), we utilized a machine learning approach to estimate BFI in ungauged catchments. The regional model exhibited strong predictive capability (overall R²=0.764), though systematic underprediction (negaitve pbias) was noted, particularly in monsoon-dominated and warm-dry climates. These biases suggest areas for improvement in model performance, potentially through the incorporation of additional predictors or more refined regional hydrologic representations. Although creating separate regional models (by physiographic region) was explored to enhance predictive accuracy, the state-wide model consistently outperformed these regionalized approaches, emphasizing the advantage of a comprehensive, state-wide dataset.

Elevation emerged as the most significant predictor, reinforcing its established role as a proxy for climate conditions influencing groundwater recharge. This aligns with findings from previous studies conducted in diverse landscapes, including New Zealand (Singh et al. 2018), across CONUS (Santhi et al. 2008), and globally (Beck et al. 2013). The mixed influence observed for Soil Type A indicates complexity in infiltration dynamics, suggesting that future research could benefit from explicitly incorporating temporal variables such as snowpack duration, soil moisture, and detailed land-cover characteristics.

## Conclusions

The long-term average BFI across Arizona is approximately 0.32, highlighting that groundwater discharge significantly contributes to surface water flows but exhibits substantial spatial variability. Our machine learning model (XGBoost) effectively predicts BFI in ungauged catchments (R² = 0.764), primarily driven by basin elevation, land cover, and hydrologic soil characteristics. Regions such as the Grand Canyon and Mogollon Rim exhibit notably high BFI values, signifying robust groundwater-surface water interactions, whereas areas such as the Defiance Plateau and southern Arizona show consistently lower BFI values, reflecting limited groundwater contributions.

Analysis of temporal trends indicates a prevailing pattern of declining base flow and BFI across the region, particularly in monsoon-dominated, warm-dry, and low-slope areas. Precipitation emerged as the strongest driver of base-flow variability, while temperature and evapotranspiration added additional layers of complexity. The observed trends underscore the sensitivity of Arizona’s hydrological systems to climatic variability and suggest increasing vulnerability under projected climate change scenarios and other stresses to aquifers, such as increased groundwater pumping.

This research highlights critical limitations in Arizona’s current streamgage network for monitoring base-flow processes, suggesting a need for targeted investments in instrumentation capable of capturing low-flow dynamics. Additionally, the demonstrated success of our modeling framework accentuates its broader applicability to similarly data-scarce dryland regions globally, providing a valuable tool for water resource management and climate adaptation strategies. By improving our understanding of groundwater-surface water interactions under changing climatic conditions, this study contributes substantially to addressing hydrological uncertainties in arid and semi-arid environments.

## References

Abatzoglou, John T. et al. 2018. “TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 19582015.” *Scientific Data* 5 (1): 170191. <https://doi.org/10.1038/sdata.2017.191>.

Ahiablame, Laurent et al. 2013. “Estimation of Annual Baseflow at Ungauged Sites in Indiana USA.” *Journal of Hydrology* 476: 13–27. <https://doi.org/10.1016/j.jhydrol.2012.10.002>.

Arizona State Climate Office. 2024. “Climate of Arizona.” <https://globalfutures.asu.edu/azclimate/climate/>.

Arnold, J. G. et al. 1995. “Automated Base Flow Separation and Recession Analysis Techniques.” *Ground Water* 33 (6): 1010–18. <https://doi.org/10.1111/j.1745-6584.1995.tb00046.x>.

——— et al. 2000. “Regional Estimation of Base Flow and Groundwater Recharge in the Upper Mississippi River Basin.” *Journal of Hydrology* 227 (1): 21–40. https://doi.org/<https://doi.org/10.1016/S0022-1694(99)00139-0>.

Ayers, Jessica R., Gabriele Villarini, Christopher Jones, and Keith Schilling. 2019. “Changes in Monthly Baseflow Across the U.S. Midwest.” *Hydrological Processes* 33 (5): 748–58. <https://doi.org/10.1002/hyp.13359>.

Ayers, Jessica R., Gabriele Villarini, Keith Schilling, Christopher Jones, Andrea Brookfield, Samuel C. Zipper, and William H. Farmer. 2022. “The Role of Climate in Monthly Baseflow Changes Across the Continental United States.” *Journal of Hydrologic Engineering* 27 (5): 04022006. <https://doi.org/10.1061/(ASCE)HE.1943-5584.0002170>.

Beck, Hylke E. et al. 2013. “Global Patterns in Base Flow Index and Recession Based on Streamflow Observations from 3394 Catchments: Global Patterns in Base Flow Characteristics.” *Water Resources Research* 49 (12): 7843–63. <https://doi.org/10.1002/2013WR013918>.

Blanchard, Paul J. 2002. “Assessments of Aquifer Sensitivity on Navajo Nation and Adjacent Lands and Ground-Water Vulnerability to Pesticide Contamination on the Navajo Indian Irrigation Project, Arizona, New Mexico, and Utah.” Water-Resources Investigations Report 02-4051. USGS. <https://doi.org/10.3133/wri024051>.

Bloomfield, J. P. et al. 2009. “Examining Geological Controls on Baseflow Index (BFI) Using Regression Analysis: An Illustration from the Thames Basin, UK.” *Journal of Hydrology* 373 (1-2): 164–76. <https://doi.org/10.1016/j.jhydrol.2009.04.025>.

Bosch, David D. et al. 2017. “Temporal Variations in Baseflow for the Little River Experimental Watershed in South Georgia, USA.” *Journal of Hydrology: Regional Studies* 10 (April): 110–21. <https://doi.org/10.1016/j.ejrh.2017.02.002>.

Buban, Michael S. et al. 2020. “A Comparison of the u.s. Climate Reference Network Precipitation Data to the Parameter-Elevation Regressions on Independent Slopes Model (PRISM).” *Journal of Hydrometeorology* 21 (10): 2391–2400. <https://doi.org/10.1175/JHM-D-19-0232.1>.

Chambless, Hannah E. et al. 2023. “Deep-Karst Aquifer Spring-Flow Trends in a Water-Limited System, Grand Canyon National Park, USA.” *Hydrogeology Journal* 31 (7): 1755–71. <https://doi.org/10.1007/s10040-023-02702-w>.

Chen, Tianqi et al. 2016. “Xgboost: A Scalable Tree Boosting System.” In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Daly, Christopher et al. 2008. “Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States.” *International Journal of Climatology* 28 (15): 2031–64. <https://doi.org/10.1002/joc.1688>.

Donovan, Keegan M. et al. 2022. “Karst Spring Processes and Storage Implications in High Elevation, Semiarid Southwestern United States.” In *Geophysical Monograph Series*, edited by Matthew J. Currell and Brian G. Katz, 1st ed., 35–50. Wiley. <https://doi.org/10.1002/9781119818625.ch4>.

Durbin, James, and Geoffrey S Watson. 1950. “Testing for Serial Correlation in Least Squares Regression: i.” *Biometrika* 37 (3/4): 409–28.

Eastoe, Christopher J. et al. 2019. “Hydrology of Mountain Blocks in Arizona and New Mexico as Revealed by Isotopes in Groundwater and Precipitation.” *Geosciences* 9 (11). <https://doi.org/10.3390/geosciences9110461>.

Eckhardt, K. 2005. “How to Construct Recursive Digital Filters for Baseflow Separation.” *Hydrological Processes* 19 (2): 507–15. <https://doi.org/10.1002/hyp.5675>.

———. 2008. “A Comparison of Baseflow Indices, Which Were Calculated with Seven Different Baseflow Separation Methods.” *Journal of Hydrology* 352 (1-2): 168–73. <https://doi.org/10.1016/j.jhydrol.2008.01.005>.

———. 2023. “Technical Note: How Physically Based Is Hydrograph Separation by Recursive Digital Filtering?” *Hydrology and Earth System Sciences* 27 (2): 495–99. <https://doi.org/10.5194/hess-27-495-2023>.

Fekete, Balázs M et al. 2007. “The Current Status of Global River Discharge Monitoring and Potential New Technologies Complementing Traditional Discharge Measurements.” *IAHS Publications*.

Ficklin, Darren L. et al. 2016. “Impacts of Recent Climate Change on Trends in Baseflow and Stormflow in United States Watersheds.” *Geophysical Research Letters* 43 (10): 5079–88. <https://doi.org/10.1002/2016GL069121>.

Fuka, DR et al. 2014. “EcoHydRology: A Community Modeling Foundation for Eco-Hydrology.” *R Package Version 0.4* 12.

Georganos, S. et al. 2018. “Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting.” *IEEE Geoscience and Remote Sensing Letters* 15: 607–11. <https://doi.org/10.1109/LGRS.2018.2803259>.

Gonzales, A. L. et al. 2009. “Comparison of Different Base Flow Separation Methods in a Lowland Catchment.” *Hydrology and Earth System Sciences* 13 (11): 2055–68. <https://doi.org/10.5194/hess-13-2055-2009>.

Hamed, Khaled H., and A. Rao. 1998. “A Modified Mann-Kendall Trend Test for Autocorrelated Data.” *Journal of Hydrology* 204 (1): 182–96. https://doi.org/<https://doi.org/10.1016/S0022-1694(97)00125-X>.

Hydrology, Institute of. 1980. “Low Flow Studies Report 3 Catchment Characteristic Estimation Manual.” Institute of Hydrology.

IUCN. 2019. “Drylands and Climate Change.” <https://www.iucn.org/resources/issues-brief/drylands-and-climate-change>.

Kendall, Maurice G. 1970. *Rank Correlation Methods*. 4th ed. London: Griffin.

Krabbenhoft, Corey A., George H. Allen, Peirong Lin, Sarah E. Godsey, Daniel C. Allen, Ryan M. Burrows, Amanda G. DelVecchia, et al. 2022. “Assessing placement bias of the global river gauge network.” *Nature Sustainability* 5 (7): 586–92. <https://doi.org/10.1038/s41893-022-00873-0>.

Levick, Lainie R. et al. 2008. “The ecological and hydrological significance of ephemeral and intermittent streams in the arid and semi-arid American Southwest.” *U.S. Environmental Protection Agency and USDA/ARS Southwest Watershed Research Center*, no. EPA/600/R-08/134 (January).

Li, Jundong, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. 2017. “Feature Selection: A Data Perspective.” *ACM Comput. Surv.* 50 (6). <https://doi.org/10.1145/3136625>.

Lundberg, Scott M, and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” Edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett 30. <https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>.

Lyne, V et al. 1979. “Institute of Engineers Australia National Conference.” In, 79:89–93. Barton, Australia: Institute of Engineers Australia.

Mann, Henry B. 1945. “Nonparametric Tests Against Trend.” *Econometrica: Journal of the Econometric Society*, 245–59.

Maricopa County, Flood Control District of. 2020. “Flood Warning Program - Program Review.” Maricopa County Flood Control District; Maricopa County Flood Control District. <https://alert.fcd.maricopa.gov/alert/FWP_report_FINAL.pdf>.

Murdoch, W. James, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. 2019. “Definitions, Methods, and Applications in Interpretable Machine Learning.” *Proceedings of the National Academy of Sciences* 116 (44): 22071–80. <https://doi.org/10.1073/pnas.1900654116>.

Nathan, R. J. et al. 1990. “Evaluation of Automated Techniques for Base Flow and Recession Analyses.” *Water Resources Research* 26 (7): 1465–73. <https://doi.org/10.1029/WR026i007p01465>.

Neff, B. P. et al. 2005. “Base Flow in the Great Lakes Basin.” Report 2005-5217. Reston, VA: USGS. <https://doi.org/10.3133/sir20055217>.

Ni, Lingling et al. 2020. “Streamflow Forecasting Using Extreme Gradient Boosting Model Coupled with Gaussian Mixture Model.” *Journal of Hydrology* 586: 124901.

O’Donnell, Frances C et al. 2016. “12th Biennial Conference of Research on the Colorado Plateau.” In.

Reitz, M. et al. 2017. “Annual Estimates of Recharge, Quick-Flow Runoff, and Evapotranspiration for the Contiguous US Using Empirical Regression Equations.” *JAWRA Journal of the American Water Resources Association* 53 (4): 961–83. <https://doi.org/10.1111/1752-1688.12546>.

Rozos, E. et al. 2021. “Machine Learning in Assessing the Performance of Hydrological Models.” *Hydrology*. <https://doi.org/10.3390/hydrology9010005>.

Santhi, C. et al. 2008. “Regional Estimation of Base Flow for the Conterminous United States by Hydrologic Landscape Regions.” *Journal of Hydrology* 351 (1-2): 139–53. <https://doi.org/10.1016/j.jhydrol.2007.12.018>.

Scanlon, Bridget R. et al. 2006. “Global Synthesis of Groundwater Recharge in Semiarid and Arid Regions.” *Hydrological Processes* 20 (15): 3335–70. <https://doi.org/10.1002/hyp.6335>.

Schaake, J, S Cong, and Q Duan. 2006. “U.s. MOPEX DATA SET.” *IAHS Publication Series, Vol. 307, N/A, November 1, 2006, Pp. 9-28*, May. <https://www.osti.gov/biblio/899413>.

Schmidt, Lennart et al. 2020. “Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany.” *Water Resources Research* 56 (5): e2019WR025924. https://doi.org/<https://doi.org/10.1029/2019WR025924>.

Sheppard, Pr et al. 2002. “The Climate of the US Southwest.” *Climate Research* 21: 219–38. <https://doi.org/10.3354/cr021219>.

Singh, Shailesh Kumar et al. 2018. “Towards Baseflow Index Characterisation at National Scale in New Zealand.” *Journal of Hydrology* 568: 646–57. <https://doi.org/10.1016/j.jhydrol.2018.11.025>.

Sloto, Ronald et al. 1996. “HYSEP: A Computer Program for Streamflow Hydrograph Separation and Analysis.” 96-4040. USGS. <https://doi.org/10.3133/wri964040>.

Szczepanek, Robert. 2022. “Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost.” *Hydrology* 9 (12): 226.

Tan, Xuejin et al. 2020. “Global Changes in Baseflow Under the Impacts of Changing Climate and Vegetation.” *Water Resources Research* 56 (9): e2020WR027349. <https://doi.org/10.1029/2020WR027349>.

Taylor, Richard G. et al. 2013. “Ground Water and Climate Change.” *Nature Climate Change* 3 (4): 322–29. <https://doi.org/10.1038/nclimate1744>.

USDA. 2009. *Part 630 Hydrology National Engineering Handbook, Chapter 7: Hydrologic Soil Groups*. USDA Natural Resources Conservation Service.

USGS. 2010. “Water Resources Data for the United States Water Year 2007.” <http://wdr.water.usgs.gov/wy2007/search.jsp>.

USGS. 2018. “Water Science Glossary.” <https://www.usgs.gov/special-topics/water-science-school/science/water-science-glossary>.

Winter, Thomas C. 2007. “The Role of Ground Water in Generating Streamflow in Headwater Areas and in Maintaining Base Flow1.” *JAWRA Journal of the American Water Resources Association* 43 (1): 15–25. <https://doi.org/10.1111/j.1752-1688.2007.00003.x>.

Woodhouse, Connie A. et al. 2022. “Upper Gila, Salt, and Verde Rivers: Arid Land Rivers in a Changing Climate.” *Earth Interactions* 26 (1): 1–14. <https://doi.org/10.1175/EI-D-21-0014.1>.

Yao, Yingying et al. 2018. “Role of Groundwater in the Dryland Ecohydrological System: A Case Study of the Heihe River Basin.” *Journal of Geophysical Research: Atmospheres* 123 (13): 6760–76. <https://doi.org/10.1029/2018JD028432>.

Ying, Xue. 2019. “An Overview of Overfitting and Its Solutions.” *Journal of Physics: Conference Series* 1168 (2): 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>.

Zomer, Robert J. et al. 2022. “Version 3 of the Global Aridity Index and Potential Evapotranspiration Database.” *Scientific Data* 9 (1): 409. <https://doi.org/10.1038/s41597-022-01493-1>.