

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

**3      Caelum Mroczek<sup>1</sup>, Abraham Springer<sup>1</sup>, Neha Gupta<sup>2</sup>, Benjamin Lucas<sup>3</sup>**

**4**      <sup>1</sup>School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ, USA,

**5**      <sup>2</sup>Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA,

**6**      <sup>3</sup>Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ, USA,

7      **Abstract**

8      Base flow, sustained by groundwater discharge, is a vital component of river ecosystems,  
 9      particularly in drylands, where water resources are limited. This study analyzes  
 10     the instrumented streamflow record in Arizona to assess long-term base-flow  
 11     index (BFI) trends across gauged catchments. Results indicate that approximately  
 12     32% of Arizona's streamflow originates from groundwater discharge, with significant  
 13     spatial variability driven by landscape and climatic factors. Base flow relationships  
 14     are analyzed with coincident trends in climate variables such as precipitation, evap-  
 15     otranspiration, and temperature. Spatial and climatic trends reveal variability in  
 16     base-flow contributions, providing insight into groundwater-surface water interac-  
 17     tions in arid and semi-arid landscapes. Building on this analysis, we applied machine  
 18     learning methods to predict BFI in ungauged basins, addressing the challenges  
 19     of Arizona's sparse streamgage network. Using the eXtreme Gradient Boosting  
 20     (XGBoost) algorithm trained on hydroclimate and physiographic predictors, we  
 21     estimated long-term BFI from 1991 to 2020. This combined approach integrates ob-  
 22     servational data with predictive modeling to enhance our understanding of base-flow  
 23     processes and provide a framework for water resource management in data-limited  
 24     regions.

25      **Plain Language Summary**

26      Rivers in drylands, such as Arizona, depend on groundwater contributions to main-  
 27     tain flow during dry periods. This portion of streamflow, called base flow, plays a  
 28     critical role in supporting ecosystems and water availability in these regions. This  
 29     study examines long-term patterns of base flow in Arizona using data from stream-  
 30     flow monitoring stations. We found that about 32% of Arizona's river flow comes  
 31     from groundwater, though this varies significantly across the state depending on  
 32     local climate and landscape features. We also analyzed how base flow is influenced  
 33     by changes in precipitation, evapotranspiration, and temperature, providing insights  
 34     into how groundwater and surface water interact in these complex environments. To  
 35     estimate base flow in basins without monitoring stations, we used machine learning  
 36     techniques. By training a model on data from monitored sites and corresponding hy-  
 37     droclimate data, we predicted long-term base flow for ungauged areas across Arizona.  
 38     This innovative approach addresses the challenge of Arizona's sparse monitoring net-  
 39     work and provides a valuable tool for understanding and managing water resources  
 40     in regions with limited data.

41      **1 Introduction**

42      Dryland regions, encompassing arid, semi-arid, hyper-arid, and dry sub-humid sys-  
 43     tems, account for 40% of the Earth's land surface. These regions are home to ap-  
 44     proximately 2 billion people globally and constitute the largest terrestrial biome  
 45     (IUCN, 2019). Despite supporting diverse ecosystems and human populations,  
 46     dryland regions face mounting hydrologic challenges exacerbated by increasing ur-  
 47     banization, expanding agricultural activities, and climate-induced amplification of  
 48     precipitation patterns (Taylor et al., 2013). This water scarcity is intensifying due  
 49     to the compounding effects of climate variability and increased groundwater extrac-  
 50     tion (Taylor et al., 2013). In drylands, groundwater serves as a vital resource for  
 51     sustaining ecosystems and meeting human needs (Scanlon et al., 2006; Yao et al.,  
 52     2018).

53      Base flow is the sustained portion of streamflow in the absence of runoff that is de-  
 54     rived from groundwater discharge (USGS, 2018). Base flow is critical to maintaining  
 55     seasonal low-flow regimes, supporting aquatic ecosystems, and facilitating the trans-  
 56     port of nutrients and chemicals. Base-flow contribution to streamflow can be highly  
 57     variable spatially (Beck et al., 2013; Bosch et al., 2017; Singh et al., 2018), and tem-  
 58     porally (Ficklin et al., 2016; Tan et al., 2020). Increasing groundwater extraction,

59 changes in land cover/land use, and changes in precipitation patterns due to climate  
 60 change affect the timing and volumes of base flow (Tan et al., 2020; Taylor et al.,  
 61 2013). Effective management of water quantity and quality depends on understanding  
 62 seasonal and interannual base-flow patterns and long-term changes in base-flow  
 63 behavior.

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

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

88 This study develops a technique for estimating BFI in ungauged basins across Ari-  
 89 zona and evaluates the state's long-term BFI. Regional trends in base flow and BFI  
 90 at instrumented sites are also analyzed, with these trends being linked to coincident  
 91 trends in precipitation, reference evapotranspiration ( $ET_O$ ), and temperature over  
 92 the same periods. Using a machine learning model trained on the hydrogeologic  
 93 characteristics of surface water basins, we estimate the mean BFI for ungauged  
 94 basins from 1991 to 2020. This approach helps address the spatial gaps in the  
 95 streamgage network, which is relatively sparse across the state. The results offer  
 96 novel insights into low-flow processes in both gauged and ungauged basins, enhanc-  
 97 ing our understanding of climate controls on consistent flows in Arizona. This study  
 98 contributes to a more comprehensive view of hydrological dynamics in the context of  
 99 arid and semi-arid landscapes.

## 100 2 Data & Methods

### 101 2.1 Study Area

102 The state of Arizona, located in the southwest United States, covers a total area  
 103 of 295,253 km<sup>2</sup>. Arizona is divided into two primary physiographic provinces: the  
 104 Colorado Plateau in the northeast, and the Basin-and-Range region in the west and  
 105 south. The Central Highlands is a transition zone consisting of scattered basins sep-  
 106 arated by the mountainous foothills of the Mogollon Rim. The Colorado Plateau is  
 107 dominated by high-elevation desert with an average elevation of 1,936 masl (6,352  
 108 ft). Mean temperatures range from -6°C (20°F) to 26°C (80°F) and it averages 580  
 109 mm (23 in) of precipitation. The Basin-and-Range region has a semi-arid to arid  
 110 climate with an average elevation of 490 masl (1600 ft). Average temperature ranges

111 from 15°C (60°F) to 43°C (110°F) and the region averages 200 mm (8 in) of precipi-  
 112 tation annually (Arizona State Climate Office, 2024).

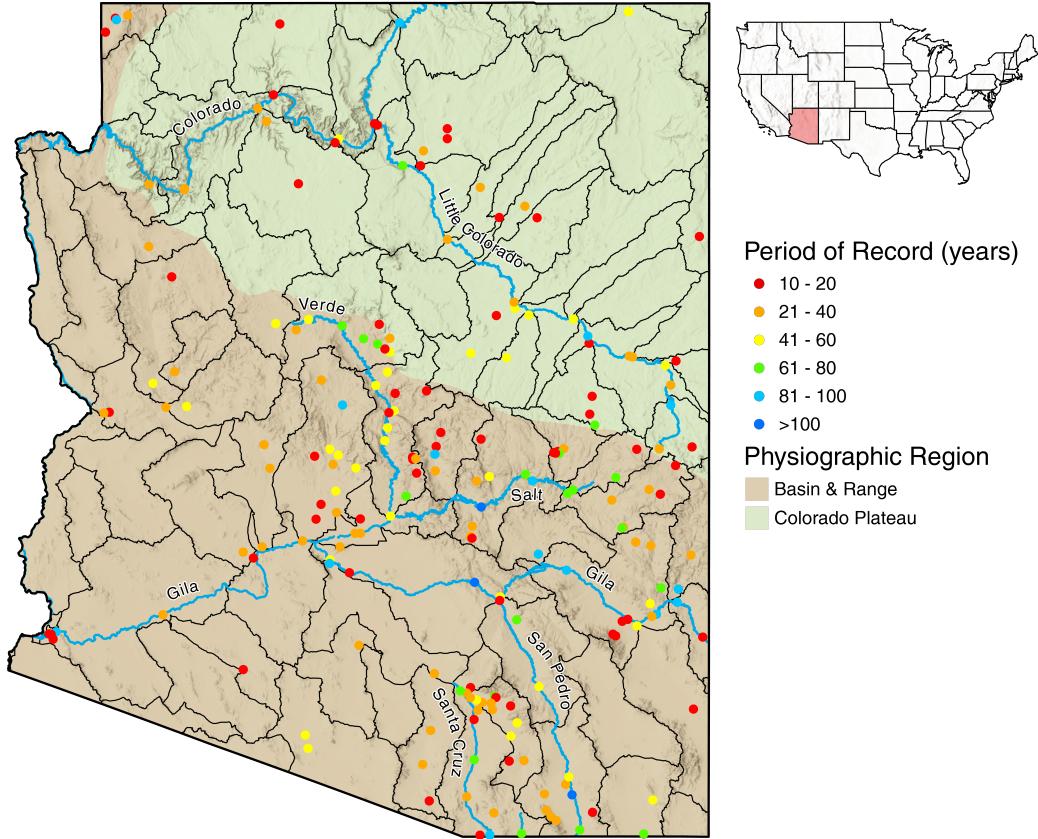


Figure 1: Map of Arizona indicating US Geological Survey (USGS) streamgages used in this study. 8-digit HUC subbasin boundaries and physiographic regions shown.

113 Arizona's hydrology varies seasonally and spatially between its physiographic regions.  
 114 In the summer, localized and intense convective precipitation events are driven by  
 115 the North American Monsoon, while in the winter, orographic precipitation comes  
 116 from Pacific frontal systems (Eastoe et al., 2019). While monsoonal precipitation  
 117 can account for up to 50% of annual precipitation, evaporation and dry preceding  
 118 soil properties leads to most precipitation becoming runoff (Sheppard et al., 2002).  
 119 As such, 94% of streams in Arizona are ephemeral or intermittent (Levick et al.,  
 120 2008). Much of the hydrology of Arizona is snow-melt derived, driven by spring melt  
 121 from the high-elevation Colorado Plateau winter snowpack. While winter precipi-  
 122 tation provides only 30% of annual averages, it provides the majority of water for  
 123 natural reservoirs (Sheppard et al., 2002).

## 2.2 Data

124 Daily observed streamflow data obtained from the United States Geological Sur-  
 125 vey (USGS) National Water Information System (NWIS) were used in this study.  
 126 Streamgages were selected depending on criteria to ensure the applicability of each  
 127 site. Following the findings of O'Donnell et al. (2016), which determined that 8–10  
 128 years of calibration data are necessary to account for climate variability in paired wa-  
 129 tershed studies in the region, a minimum record length of 10 years was required. Ad-

ditionally, 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). Periods of record ranged from 10 to 112 years, with a median of 28 years.

While our data selection ensures a robust record for this analysis, it also highlights gaps in spatial coverage that the machine learning model can address. Arizona has 184 active USGS streamflow stations (as of 2024) across an area of 295,253 km<sup>2</sup>, whereas Indiana (a humid state in the U.S.) has 189 active stations within an area of 94,326 km<sup>2</sup>. This translates to a streamgage density of approximately 2.004 gages per 1,000 km<sup>2</sup> in Indiana, over three times higher than Arizona's density of 0.623 gages per 1,000 km<sup>2</sup>. This stark difference underscores the relative sparsity of Arizona's streamgage network, particularly in the context of its larger geographic area and the unique hydrological challenges posed by its arid and semi-arid landscapes.

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). 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 is likely to reduce any potential errors (Reitz et al., 2017).

Annual reference evapotranspiration (ET<sub>O</sub>) 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 ET<sub>O</sub> values were calculated assuming a reference grass surface across the landscape with unlimited water. In the drylands of the southwestern US, ET<sub>O</sub> 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 aimed to get 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.

181    **2.3 Base-flow separation**

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

193    Base flow was calculated using a single-parameter, recursive digital filter technique  
 194    from Nathan et al. (1990). This base-flow separation technique is based on a re-  
 195    cursive digital filter used in signal analysis that separates high-frequency signals  
 196    (quickflow) from low-frequency signals (base flow) (Lyne et al., 1979). Eckhardt  
 197    (2023) noted that recursive digital filters lack a physical basis, but as the method  
 198    is easy to automate, objective, and repeatable, it is appropriate for a regional-scale  
 199    study. The Lyne-Hollick filter has been used in multiple studies (Arnold et al., 2000;  
 200    Bloomfield et al., 2009; Santhi et al., 2008; Singh et al., 2018), and it takes the form  
 201    of

$$b = \alpha b_{k-1} + \frac{1-\alpha}{2}(Q_k + Q_{k-1}) \quad (1)$$

202    where  $b$  is base flow,  $\alpha$  is the filter parameter,  $Q$  is the total streamflow, and  $k$  is  
 203    the time step. A filter parameter  $\alpha$  of 0.925 was used as in Nathan et al. (1990) and  
 204    Fuka et al. (2014). The filter was run three times (forward, backward, forward) to  
 205    attenuate the base-flow signal.

206    **2.4 Machine Learning**

207    The implementation of machine learning models to predict hydrologic indices has  
 208    been successful in past studies (Rozos et al., 2021; Schmidt et al., 2020; Singh et al.,  
 209    2018). In this work, we used the eXtreme Gradient Boosting (XGBoost) algorithm  
 210    (Chen et al., 2016) to predict BFI at ungauged locations using catchment charac-  
 211    teristics as predictors Table 1. The XGBoost algorithm is a decision tree-based  
 212    ensemble algorithm, which can be adapted for either regression or classification prob-  
 213    lems. This algorithm iteratively builds an ensemble of decision trees, where each  
 214    tree corrects errors from previous trees to improve predictions (Chen et al., 2016).  
 215    Its efficiency, scalability, and robustness have made it increasingly popular in re-  
 216    cent years, with successful applications in environmental modeling tasks such as  
 217    streamflow forecasting (Ni et al., 2020; Szczepanek, 2022) and land use/land cover  
 218    classification (Georganos et al., 2018).

219    XGBoost operates by leveraging gradient boosting on decision tree algorithms, com-  
 220    bining multiple low-variance models to produce a robust overall prediction. Gradient  
 221    boosting works iteratively: the initial tree is trained on the target values, while sub-  
 222    sequent trees are trained on the residual errors of the preceding tree. Each tree is  
 223    assigned a weight based on its contribution to reducing error, and these weights are  
 224    used to determine the influence of each tree in the final model. The ultimate predic-  
 225    tion is made by aggregating the outputs of all  $n$  weighted trees in the ensemble. In  
 226    this study, the trained XGBoost model is used to predict BFI in ungauged catch-  
 227    ments based on geospatial and hydroclimate predictor variables (Table 1). Certain  
 228    features were further subdivided according to their areal coverage within each basin  
 229    (e.g. land cover was divided into 16 subdivisions). This approach allowed the model  
 230    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 ( $\eta$ ), minimum split loss ( $\gamma$ ), maximum tree depth, minimum child weight, and the number of trees. While not an exhaustive list of all possible XGBoost hyperparameters, the range of values explored provided sufficient variation to ensure the selection of a high-performing model. The optimal hyperparameters were determined to be: 700 trees, a learning rate ( $\eta$ ) of 0.05, a minimum split loss ( $\gamma$ ) 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.

$K$ -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 is randomly divided into  $k$  folds of equal size. The model is trained on  $k - 1$  folds and tested on the remaining fold, referred to as the validation set. This process is repeated  $k$  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  $k$  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.

#### 2.4.1 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.

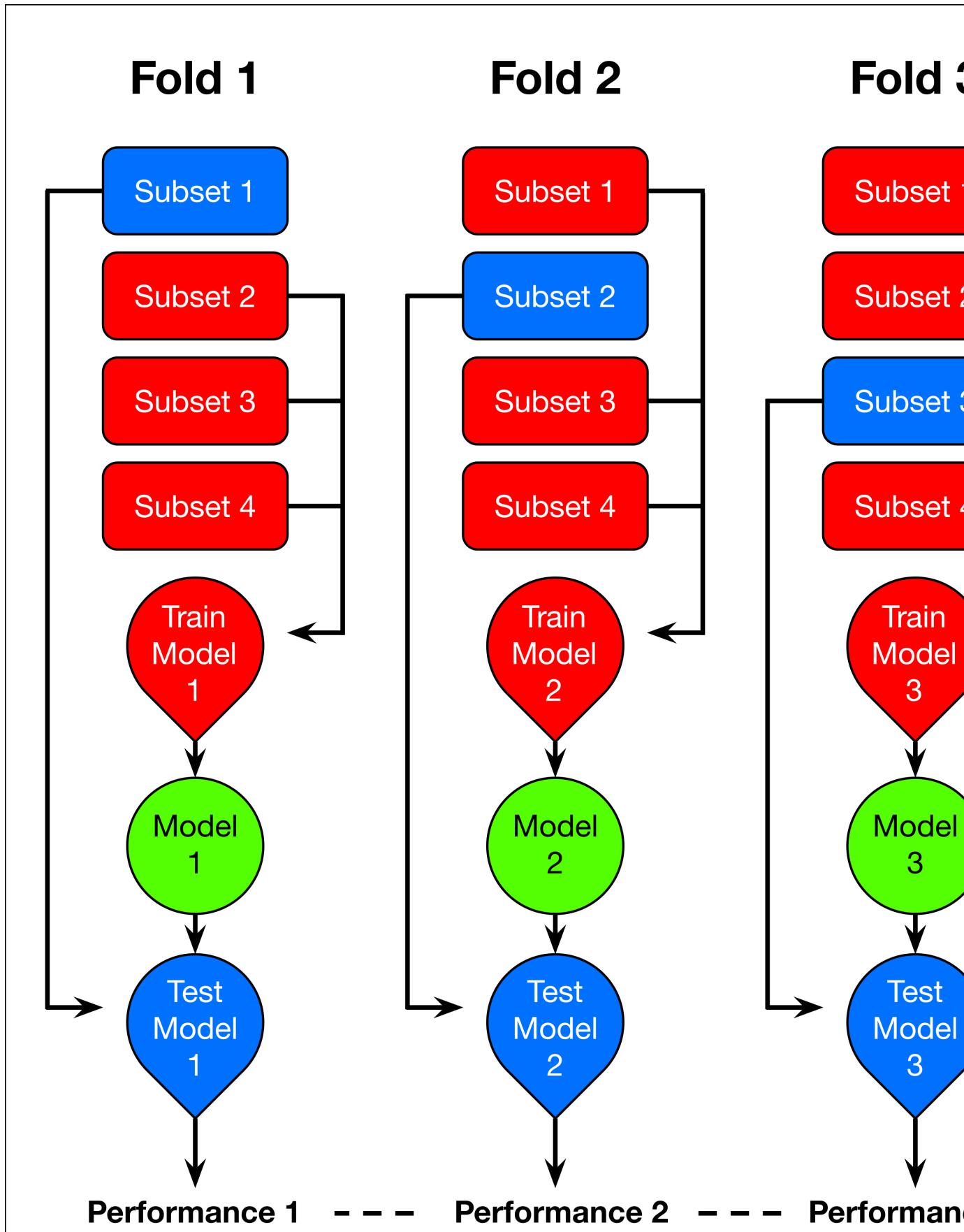


Figure 2: Schematic of  $k$ -fold cross validation. The dataset is randomly divided into  $k$  stratified folds. Each fold serves as the validation set once, while the remaining folds are combined to create a training set for model development. Performance metrics for the test set are calculated and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

To begin, an initial model was trained using the full feature set of 45 predictors (Table 1). 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 & 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 & 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 4 were selected based on their SHAP values and used to train a subsequent model Table 1. 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.

## 2.5 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 (Ayers et al., 2019; Ficklin et al., 2016; Woodhouse et al., 2022).

To check for autocorrelation, we applied the Durbin-Watson test, 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 & Rao, 1998). Trends with a  $\rho \leq 0.05$  are considered significant.

## 3 Results

### 3.1 BFI of Ungauged Catchments

#### 3.1.1 Model Validation

Predicted values of BFI are plotted against observed values for the entire period of record of the instrumented dataset in Figure 3 . The agreement between “out-of-bag” predictions (blind cross-validation, treating each site as ungauged) and observed values is acceptable ( $R^2 = 0.764$ ) indicating that the model performs well across the full dataset. 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 2. These performance metrics demonstrate that the regional model performs consistently well across different spatial and climatic classifications. However, the negative pbias values across all classifications, along with the overall pbias, indicate a systematic underprediction of BFI by the model. Categories with relatively lower  $R^2$  and Nash-Sutcliffe Efficiency (NSE) values also exhibit higher biases, reflecting weaker model performance in those specific contexts.

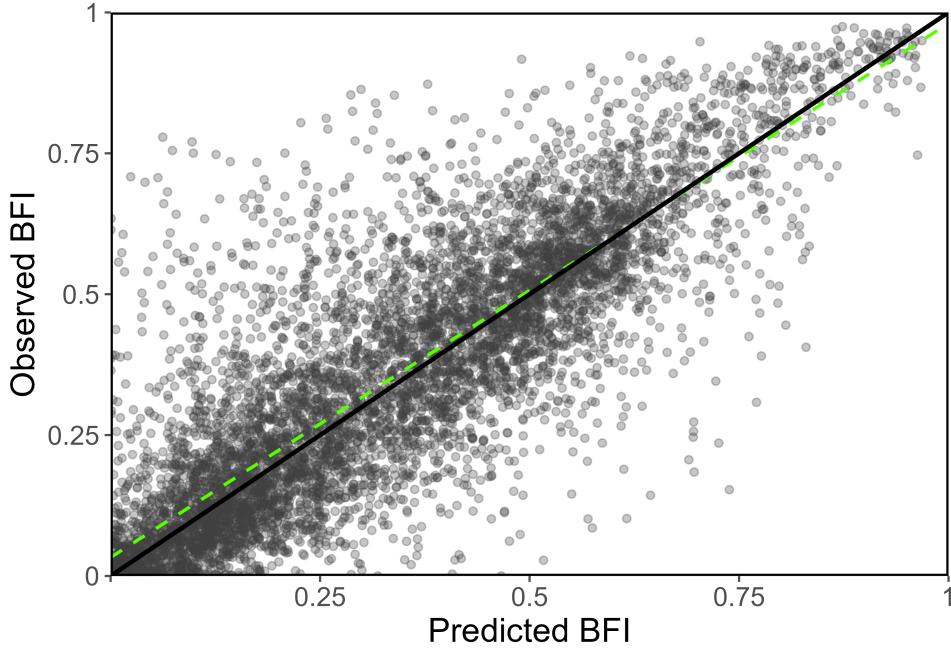


Figure 3: Linear relationship between observed BFI and predicted BFI. The solid line is the 1:1 line, the dashed, green line is regressed to the data.

Table 2: Performance of model predictions for BFI for all sites split by various classifications. n is number of observations, R<sup>2</sup> 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.

<b>Classification</b>							
<b>Group</b>	<b>n</b>	<b>R2</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>NSE</b>	<b>pbias</b>
Climate - Monsoon	3039	0.633	0.016	0.126	0.074	0.619	-13.7
Dominated							
Climate - Snowmelt	4685	0.733	0.015	0.121	0.087	0.725	-3.5
Dominated							
PhysRegion6147 - Basin&Range	6147	0.733	0.016	0.127	0.084	0.724	-6.3
PhysRegion1577 - CO	1577	0.846	0.011	0.104	0.073	0.843	-3.8
Plateau							

<b>Classification</b>		<b>n</b>	<b>R2</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>NSE</b>	<b>pbias</b>
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	

### 316      **3.1.2 Predictor Importance**

317      The predictors used to estimate BFI at ungauged sites were evaluated for their im-  
 318      portance in the final XGBoost model, as illustrated in Figure 4. The most influential  
 319      feature for predicting long-term BFI is basin elevation. While elevation itself does  
 320      not directly affect base-flow characteristics, it has consistently been identified as a  
 321      key predictor in previous BFI studies (Beck et al., 2013; Singh et al., 2018). The  
 322      importance of elevation aligns with findings from Beck et al. (2013), highlighting  
 323      its role as a proxy for climate variables such as temperature, precipitation, and  
 324      snowpack duration. Seasonal snowpack duration, in particular, has been shown to  
 325      strongly correlate with springflow and groundwater recharge in this region (Donovan  
 326      et al., 2022). This relationship is further supported by hydroclimate features where  
 327      higher temperatures tend to negatively influence BFI, while precipitation exhibits a  
 328      mixed influence. In some cases, higher precipitation values correlate with lower BFI,  
 329      likely due to a larger proportion of precipitation contributing to runoff rather than  
 330      infiltration and groundwater recharge.

331      Land cover and land use predictors also play a significant role in BFI estimation.  
 332      Analysis of SHAP values indicates that a higher percentage of evergreen forest  
 333      positively influences BFI predictions, while higher proportions of shrubland and de-  
 334      veloped land exert a negative influence. Similarly, hydrologic soil types show distinct  
 335      trends in their impact on BFI. Soil Type C, characterized by moderately high runoff  
 336      potential (20-40% clay), tends to negatively influence BFI. In contrast, Soil Type  
 337      A, which has low runoff potential and facilitates rapid water infiltration, exhibits  
 338      a mixed influence (USDA, 2009). The mixed effects of Soil Type A are likely due  
 339      to SHAP values capturing interactions between features rather than direct rela-  
 340      tionships. For example, regions dominated by Soil Type A may also have steep slopes  
 341      or sparse vegetation. Additionally, the relationship between Soil Type A and BFI  
 342      is likely non-linear and influenced by the complex dynamics of regional base-flow  
 343      processes.

### 344      **3.1.3 Predicted Long-term BFI**

345      The regionalized (HUC-8) long-term BFI (1991–2020) is shown in Figure 5, while  
 346      the long-term BFI for stream reaches with a Strahler stream order of 3 or greater  
 347      is presented in Figure 6. Basins with high BFIs, such as those along the Grand  
 348      Canyon in the northwestern part of the study area, indicate greater surface water  
 349      and groundwater interaction. Elevated BFI values are also observed along portions

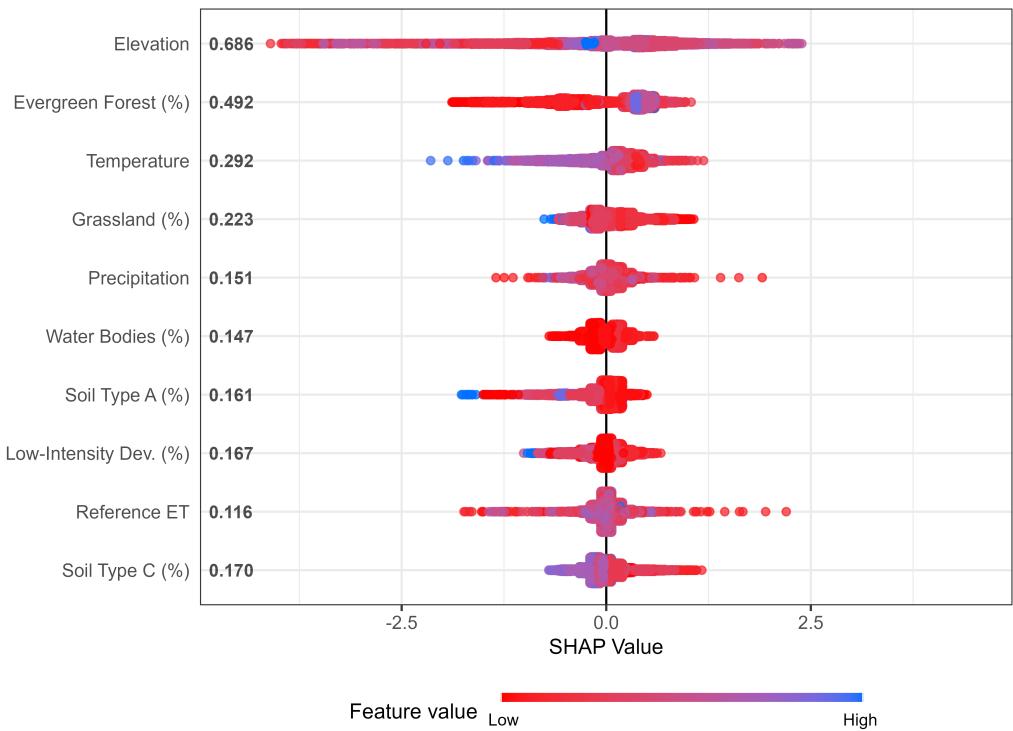


Figure 4: SHAP value plot of features used in final model. Land cover features are indicated by the percentage of cover by each land cover type and soil types are defined by hydrologic soil group.

350 of the Mogollon Rim, a heavily forested region with high precipitation that marks  
 351 the transition between physiographic regions. Additionally, headwater regions of  
 352 perennial rivers tend to exhibit higher BFI values. In contrast, low BFI values are  
 353 found in areas like the Defiance Uplift in northeastern Arizona and the arid southern  
 354 regions of the state.

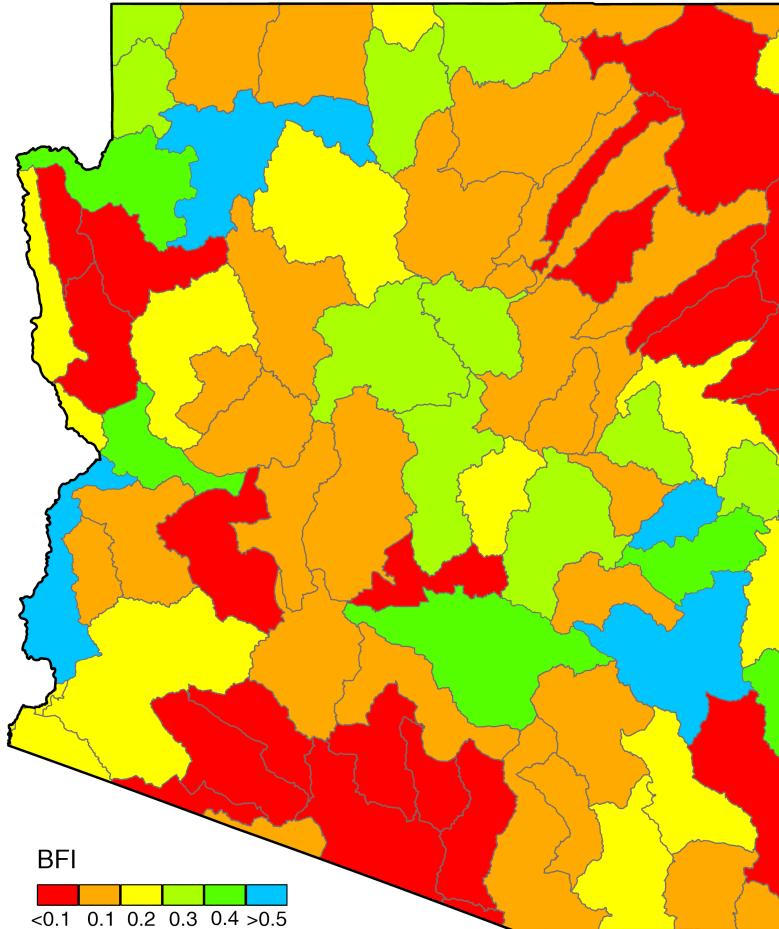


Figure 5: Predicted long-term BFI values for 8-digit HUC (1991-2020)

355 Figure 6 indicates the predicted long-term BFI for stream reaches in Arizona of  
 356 a Strahler stream order of 3 or greater. Stream-level BFI values often correspond  
 357 to the long-term BFI observed in HUC-8 basins, as geospatial predictor data were  
 358 aggregated at the basin level. Figure 6 highlights stream density across the region,  
 359 revealing areas with higher concentrations of perennial streams, particularly in re-  
 360 gions with favorable hydrogeologic conditions, such as spring-fed systems in the  
 361 Verde River and the Salt River headwaters. Conversely, lower BFI values are ob-  
 362 served in arid regions and those with lower stream densities such as parts of the  
 363 Little Colorado River Basin, where groundwater recharge is limited (Flint & Flint,  
 364 2007). While BFI values and their spatial patterns may align with certain terrain  
 365 features or climatic patterns, it is important to recognize that BFI is driven by com-  
 366 plex interactions among multiple factors Figure 4.

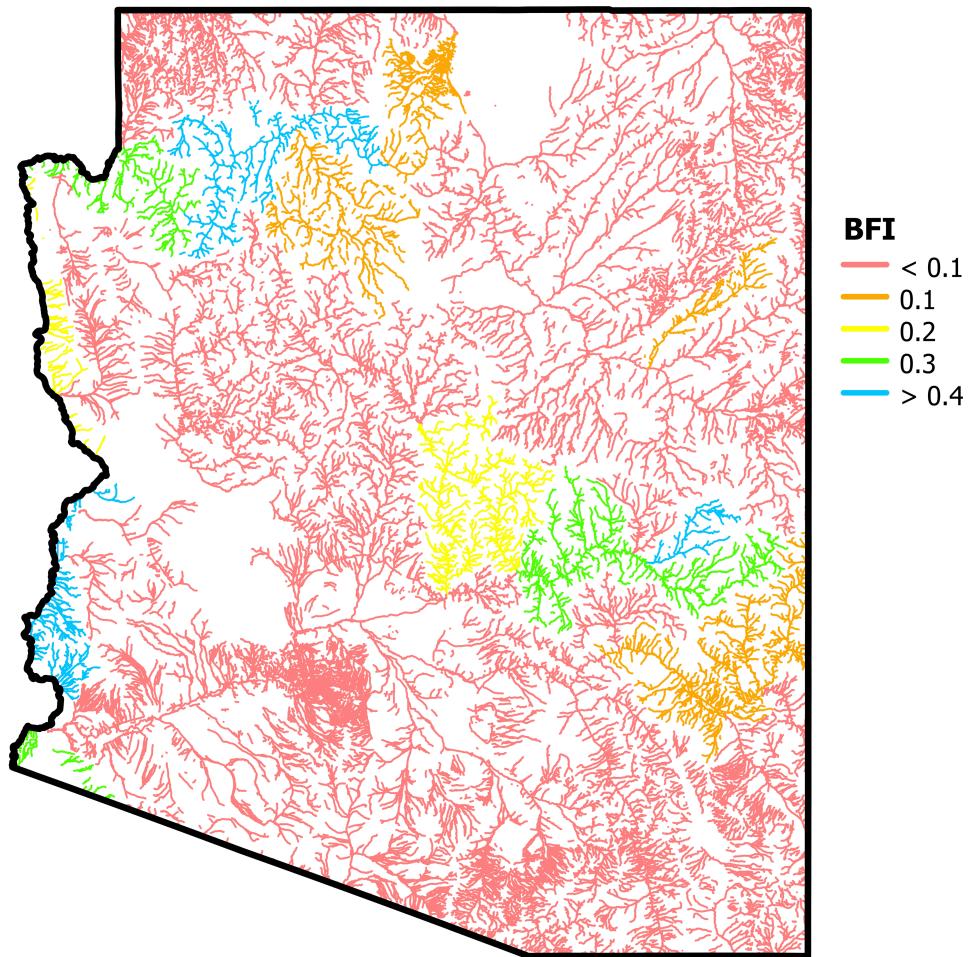
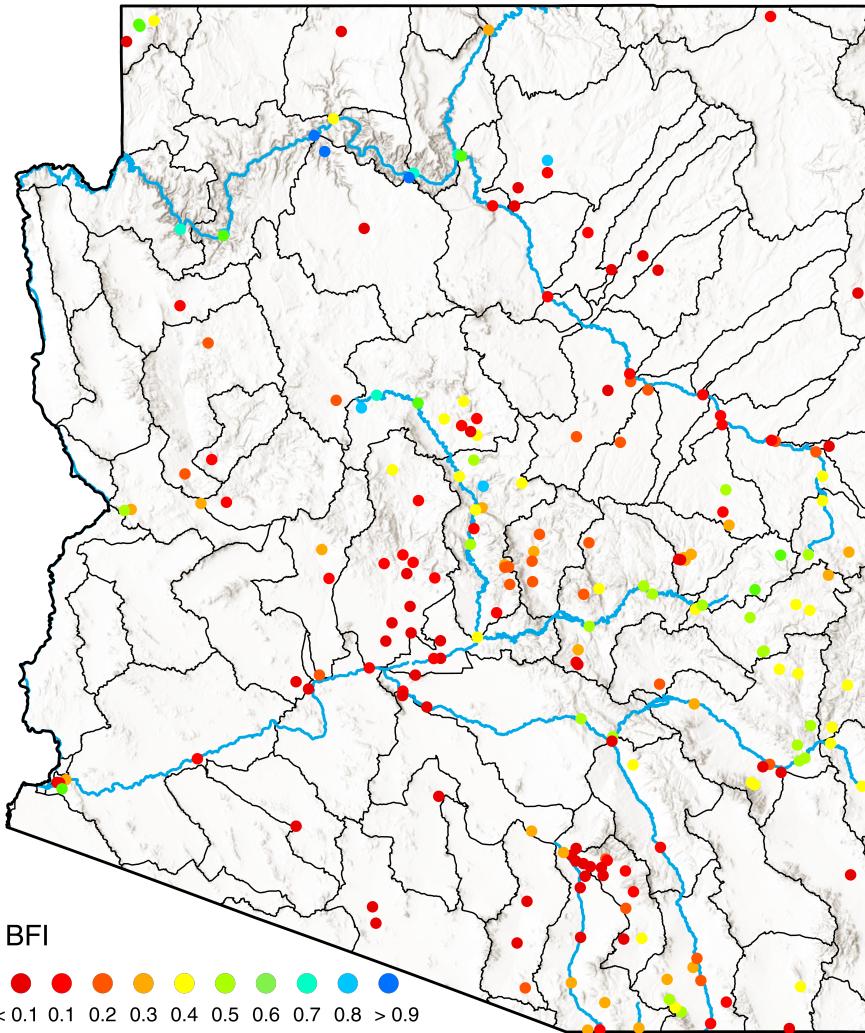


Figure 6: Predicted long-term BFI values for all stream reaches of Strahler stream order 3 and greater (1991-2020)

367 **3.2 BFI of Gauged Catchments**368  
369  
370  
371  
372  
373  
374  
375  
376  
Figure 7: Long-term BFI for the period of record from instrumented stream flow data.

The long-term BFI for the 205 gauged reaches across Arizona is illustrated in Figure 7. 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. The highly karstic geology of this region facilitates the rapid movement of subsurface flow to surface water and spring outlets (Chambless et al., 2023). Relatively high BFI values ( $>0.8$ ) are found at the spring-fed headwaters of the Verde River (Del Rio Spring) and the spring-fed headwaters of Fossil Creek. These results are consistent with interpolated BFI values reported by Wolock (2003).

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 Uplift in northeastern Arizona, which are hydrologically connected to surface streams, while the high-yield, confined regional aquifer is much deeper (Blanchard, 2002). A notable tendency emerges along most major

382 rivers in the study area: upstream reaches tend to exhibit higher BFI values, while  
 383 downstream reaches display lower values. This pattern is presumed to result from  
 384 greater groundwater-surface water interactions at stream headwaters, influenced by  
 385 spring outlets, and the dilution of base flow as water moves downstream. This trend  
 386 is particularly evident along the Gila River, Verde River, and Little Colorado River  
 387 in Arizona.

### 388 ***3.2.1 Trends in Base Flow & BFI***

389 Trends in BFI over the period of record for each streamgage in this analysis are illus-  
 390 trated in Figure 8 and Table 3. Base flow and BFI trends were analyzed across all  
 391 instrumented sites over their respective periods of record using the Mann-Kendall  
 392 test. Statistically significant trends were observed in both metrics, with a 72.20%  
 393 coincidence rate between significant base flow and BFI trends, indicating a strong  
 394 dependence of BFI on base-flow dynamics.

395 Figure 8 illustrates the spatial variation in BFI trends across the study area. Statisti-  
 396 cally significant decreasing trends are observed at 16.1% of sites, while increasing  
 397 trends are found at 8.8% of sites, with no clear regional patterns for either. In the  
 398 Basin and Range physiographic region, 9% of sites show increasing trends, while  
 399 16.7% exhibit decreasing trends. In the Colorado Plateau region, increasing trends  
 400 occur at 6.1% of sites, and decreasing trends are observed at 14.3% of sites.

### 401 ***3.2.2 Classification Trends***

402 Classifications presented in Table 3 were determined based on precipitation regime,  
 403 physiographic region, climate, and slope. The dominant precipitation regime (mon-  
 404soon vs. snowmelt) was identified by analyzing streamflow hydrographs for each  
 405 station, focusing on peak flow periods during the monsoon season (July–September)  
 406 and the snowmelt season (March–June). Physiographic region was assigned based on  
 407 which region the streamgage is located. Climate classifications were defined as warm  
 408 (above the long-term median temperature of Arizona), cool (below the long-term  
 409 median temperature), wet (above the long-term median precipitation), and dry (be-  
 410 low the long-term median precipitation). Slope was categorized as high (above the  
 411 median slope) and low (below the median slope).

412 Statistically significant decreasing trends in BFI were more common than increas-  
 413 ing trends across all site classifications (Table 3). While decreasing trends domi-  
 414 nate, both increasing and decreasing trends are observed within each classification.  
 415 Monsoon-dominated regions exhibit a higher proportion of significant negative  
 416 trends (24.1%) compared to snowmelt-dominated regions (10.2%), suggesting that  
 417 monsoon-dominated systems are more consistently correlated with declining base  
 418 flow. Among climate classifications, warm-dry climates have the highest proportion  
 419 of negative trends (20.0%), followed by warm-wet climates (19.4%), indicating that  
 420 regions with higher temperatures are more prone to base flow declines. Low-slope  
 421 regions show a greater prevalence of negative trends (20.4%) compared to high-  
 422 slope regions (11.8%). This suggests that flatter areas may be more susceptible to  
 423 base-flow reductions, potentially due to differences in hydrologic connectivity and  
 424 recharge dynamics.

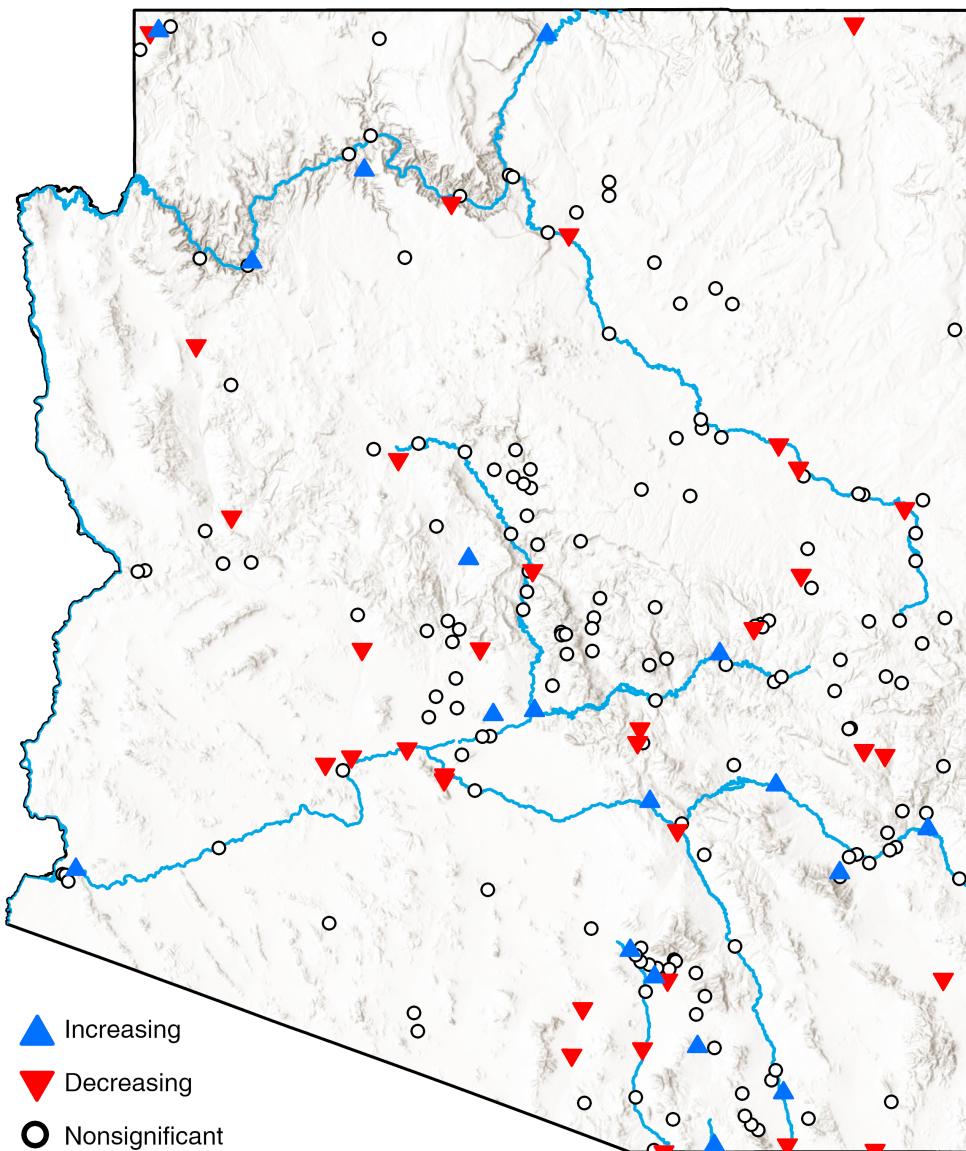


Figure 8: Trends in BFI over full period of record for instrumented sites used in this study. Red upward (blue downward) arrows indicate an increasing (decreasing) trend at a significance level of 5%. White circles represent sites with no statistically significant trends.

Table 3: Comparison of trends for BFI for all sites split by various classifications. Only sites with a significant ( $\rho \leq 0.05$ ) 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.

<b>Classification</b>					
<b>Group</b>	<b>n</b>	<b>n_pos</b>	<b>n_neg</b>	<b>perc_pos</b>	<b>perc_neg</b>
Precipitation	87	8	21	0.092	0.241
- Monsoon					
Dominated					
Precipitation	118	9	12	0.076	0.102
- Snowmelt					
Dominated					
Physiographic	156	14	26	0.090	0.167
Region -					
Basin and					
Range					
Physiographic	49	3	7	0.061	0.143
Region -					
Colorado					
Plateau					
Climate -	31	2	6	0.065	0.194
Warm-Wet					
Climate -	55	6	11	0.109	0.200
Warm-Dry					
Climate -	74	4	9	0.054	0.122
Cool-Wet					
Climate -	45	5	7	0.111	0.156
Cool-Dry					
Slope - High	102	10	12	0.098	0.118
Slope - Low	103	7	21	0.068	0.204

### 3.2.3 Coincident Climate Trends

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 4). 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).

Table 4: Coincident trends of climate variables ( $ET_O$ , precipitation, temperature) with base flow and BFI trends.

	<b>Climate Variable</b>	<b>Coincidence Percentage (%)</b>
BFI	$ET_O$	44.88
	Precipitation	53.17
	Temperature	47.32
	$ET_O$	55.12
	Precipitation	64.88
Base Flow		

Climate Variable	Coincidence Percentage (%)
Temperature	39.02

434 The analysis shows that base flow and BFI trends most frequently align with precip-  
 435 itation trends (64.88% and 53.17%, respectively), emphasizing precipitation as the  
 436 primary driver of local groundwater recharge and discharge. Coincidence with  $ET_O$   
 437 (reference evapotranspiration) trends (55.12% for base flow and 44.88% for BFI) sug-  
 438 gests that evapotranspiration also plays a significant role, particularly in arid regions  
 439 where it can reduce recharge or base flow during dry periods. In contrast, tempera-  
 440 ture trends show lower percentages of coincidence, often opposing base flow and BFI  
 441 trends. Specifically, positive (negative) temperature trends are frequently associated  
 442 with negative (positive) base-flow trends (60.98%) and BFI trends (52.68%). These  
 443 results highlight the complex interplay between climatic variables and hydrological  
 444 processes. While precipitation exerts the strongest influence on base flow and BFI,  
 445 evapotranspiration and temperature add further variability in specific environmental  
 446 contexts.

#### 447 4 Summary & Conclusions

448 This study provides new insights into base-flow dynamics and groundwater con-  
 449 tributions in Arizona's dryland rivers by combining an analysis of instrumented  
 450 streamflow records with machine learning predictions for ungauged basins. The  
 451 results highlight significant spatial variability in BFI, with approximately 32% of  
 452 Arizona's long-term streamflow originating from groundwater discharge. Regions  
 453 such as the Grand Canyon and Mogollon Rim demonstrate high BFI values due to  
 454 strong groundwater-surface water interactions, while areas like the Little Colorado  
 455 River Basin exhibit low BFI values, reflecting limited groundwater recharge.

456 Using an XGBoost machine learning algorithm, we successfully predicted long-term  
 457 BFI in ungauged basins, achieving strong model performance ( $R^2 = 0.764$ , RMSE  
 458 = 0.129). The model performed well across all classifications, demonstrating its ro-  
 459 bustness in capturing base-flow dynamics across a region with substantial variability  
 460 in climate, elevation, and physiographic characteristics. Key predictors included  
 461 elevation, land cover, and soil type, highlighting the importance of integrating hydro-  
 462 climate and physiographic characteristics into regional hydrological models. These  
 463 predictions address the limitations posed by Arizona's sparse streamgage network,  
 464 offering a scalable approach to estimate BFI in data-limited regions.

465 Our analysis of BFI trends in gauged catchments revealed that precipitation is the  
 466 primary driver of base-flow variability, with evapotranspiration and temperature  
 467 contributing additional complexity. These findings emphasize the critical role of  
 468 climate-hydrology interactions in shaping groundwater contributions to streamflow.  
 469 Inverse trends between temperature and BFI suggest that further warming could  
 470 reduce groundwater contributions to streamflow. Coincident trends in precipitation  
 471 and BFI further underscore the importance of understanding recharge processes,  
 472 especially in arid and semi-arid landscapes, where precipitation events play key roles  
 473 in replenishing groundwater.

474 The utility of the streamgage networks for base-flow analyses in Arizona is limited  
 475 by their design and focus. There has been an increase in the number of streamgages  
 476 in the region to meet regulatory imperatives, such as the Clean Water Act. Even  
 477 so, many of these gages are not suited for base-flow studies because they emphasize  
 478 peak flow monitoring and lack the ability to accurately measure low-flow dynam-  
 479 ics (Maricopa County, 2020). New, non-USGS streamgages are typically installed  
 480 by flood control districts (e.g. the ALERT system) and are designed to track flood  
 481 flows rather than base flow. New USGS streamgages have been added over the past

482 decade to address instream flow rights and to improve the density of monitoring in  
 483 the future.

484 This study demonstrates the benefits of combining observational records with ma-  
 485 chine learning to improve our understanding of streamflow processes in drylands.  
 486 Future work should explore the projected effects of climate change on base-flow pro-  
 487 cesses and developing models for other data-poor regions. The framework presented  
 488 here has broad applicability to other arid and semi-arid regions worldwide and can  
 489 inform water resource management strategies aimed at addressing water scarcity and  
 490 adapting to climate variability.

## 491 References

- 492 Abatzoglou, J. T. et al. (2018). TerraClimate, a high-resolution global dataset of  
 493 monthly climate and climatic water balance from 1958–2015. *Scientific Data*,  
 494 5(1), 170191. <https://doi.org/10.1038/sdata.2017.191>
- 495 Ahiablame, L. et al. (2013). Estimation of annual baseflow at ungauged sites in  
 496 Indiana USA. *Journal of Hydrology*, 476, 13–27. <https://doi.org/10.1016/j.jhydrol.2012.10.002>
- 497 Arizona State Climate Office. (2024). Climate of Arizona. Retrieved from <https://globalfutures.asu.edu/azclimate/climate/>
- 498 Arnold, J. G. et al. (1995). Automated Base Flow Separation and Recession Analy-  
 499 sis Techniques. *Ground Water*, 33(6), 1010–1018. <https://doi.org/10.1111/j.1745-6584.1995.tb00046.x>
- 500 Arnold, J. G. et al. (2000). Regional estimation of base flow and groundwater  
 501 recharge in the upper mississippi river basin. *Journal of Hydrology*, 227(1),  
 502 21–40. [https://doi.org/10.1016/S0022-1694\(99\)00139-0](https://doi.org/10.1016/S0022-1694(99)00139-0)
- 503 Ayers, J. R., Villarini, G., Jones, C., & Schilling, K. (2019). Changes in monthly  
 504 baseflow across the U.S. Midwest. *Hydrological Processes*, 33(5), 748–758.  
<https://doi.org/10.1002/hyp.13359>
- 505 Beck, H. E. et al. (2013). Global patterns in base flow index and recession based  
 506 on streamflow observations from 3394 catchments: Global Patterns in Base  
 507 Flow Characteristics. *Water Resources Research*, 49(12), 7843–7863. <https://doi.org/10.1002/2013WR013918>
- 508 Blanchard, P. J. (2002). *Assessments of aquifer sensitivity on Navajo Nation and  
 509 adjacent lands and ground-water vulnerability to pesticide contamination on the  
 510 Navajo Indian Irrigation Project, Arizona, New Mexico, and Utah* (No. Water-  
 511 Resources Investigations Report 02-4051). USGS. <https://doi.org/10.3133/wri024051>
- 512 Bloomfield, J. P. et al. (2009). Examining geological controls on baseflow in-  
 513 dex (BFI) using regression analysis: An illustration from the Thames Basin,  
 514 UK. *Journal of Hydrology*, 373(1-2), 164–176. <https://doi.org/10.1016/j.jhydrol.2009.04.025>
- 515 Bosch, D. D. et al. (2017). Temporal variations in baseflow for the Little River ex-  
 516 perimental watershed in South Georgia, USA. *Journal of Hydrology: Regional  
 517 Studies*, 10, 110–121. <https://doi.org/10.1016/j.ejrh.2017.02.002>
- 518 Buban, M. S. et al. (2020). A comparison of the u.s. Climate reference network  
 519 precipitation data to the parameter-elevation regressions on independent slopes  
 520 model (PRISM). *Journal of Hydrometeorology*, 21(10), 2391–2400. <https://doi.org/10.1175/JHM-D-19-0232.1>
- 521 Chambliss, H. E. et al. (2023). Deep-karst aquifer spring-flow trends in a water-  
 522 limited system, grand canyon national park, USA. *Hydrogeology Journal*, 31(7),  
 523 1755–1771. <https://doi.org/10.1007/s10040-023-02702-w>
- 524 Chen, T. et al. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of  
 525 the 22nd acm sigkdd international conference on knowledge discovery and data  
 526 mining* (pp. 785–794).

- 535 Daly, C. et al. (2008). Physiographically sensitive mapping of climatological tem-  
 536 perature and precipitation across the conterminous United States. *International*  
 537 *Journal of Climatology*, 28(15), 2031–2064. <https://doi.org/10.1002/joc.1688>
- 539 Donovan, K. M. et al. (2022). Karst spring processes and storage implications  
 540 in high elevation, semiarid southwestern united states. In M. J. Currell &  
 541 B. G. Katz (Eds.), *Geophysical monograph series* (1st ed., pp. 35–50). Wiley.  
 542 <https://doi.org/10.1002/9781119818625.ch4>
- 543 Eastoe, C. J. et al. (2019). Hydrology of mountain blocks in arizona and new mex-  
 544 ico as revealed by isotopes in groundwater and precipitation. *Geosciences*, 9(11).  
 545 <https://doi.org/10.3390/geosciences9110461>
- 546 Eckhardt, K. (2005). How to construct recursive digital filters for baseflow sepa-  
 547 ration. *Hydrological Processes*, 19(2), 507–515. <https://doi.org/10.1002/hyp.5675>
- 549 Eckhardt, K. (2008). A comparison of baseflow indices, which were calculated with  
 550 seven different baseflow separation methods. *Journal of Hydrology*, 352(1-2),  
 551 168–173. <https://doi.org/10.1016/j.jhydrol.2008.01.005>
- 552 Eckhardt, K. (2023). Technical note: How physically based is hydrograph separa-  
 553 tion by recursive digital filtering? *Hydrology and Earth System Sciences*, 27(2),  
 554 495–499. <https://doi.org/10.5194/hess-27-495-2023>
- 555 Fekete, B. M. et al. (2007). The current status of global river discharge monitoring  
 556 and potential new technologies complementing traditional discharge measure-  
 557 ments.
- 558 Ficklin, D. L. et al. (2016). Impacts of recent climate change on trends in baseflow  
 559 and stormflow in United States watersheds. *Geophysical Research Letters*, 43(10),  
 560 5079–5088. <https://doi.org/10.1002/2016GL069121>
- 561 Flint, L. E., & Flint, A. L. (2007). *Regional Analysis of Ground-Water Recharge*.
- 562 Fuka, D. et al. (2014). EcoHydRology: A community modeling foundation for eco-  
 563 hydrology. *R Package Version 0.4, 12*.
- 564 Georganos, S. et al. (2018). Very high resolution object-based land use–land cover  
 565 urban classification using extreme gradient boosting. *IEEE Geoscience and Re-  
 566 mote Sensing Letters*, 15, 607–611. <https://doi.org/10.1109/LGRS.2018.2803259>
- 568 Gonzales, A. L. et al. (2009). Comparison of different base flow separation methods  
 569 in a lowland catchment. *Hydrology and Earth System Sciences*, 13(11), 2055–  
 570 2068. <https://doi.org/10.5194/hess-13-2055-2009>
- 571 Hamed, K. H., & Rao, A. (1998). A modified mann-kendall trend test for autocor-  
 572 related data. *Journal of Hydrology*, 204(1), 182–196. [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X)
- 574 Institute of Hydrology. (1980). *Low flow studies report 3 catchment characteristic  
 575 estimation manual*.
- 576 IUCN. (2019, September). Drylands and climate change. Retrieved from <https://www.iucn.org/resources/issues-brief/drylands-and-climate-change>
- 578 Kendall, M. G. (1970). *Rank correlation methods* (4th ed). London: Griffin.
- 579 Levick, L. R. et al. (2008). The ecological and hydrological significance of ephemeral  
 580 and intermittent streams in the arid and semi-arid American Southwest. *U.S.*  
 581 *Environmental Protection Agency and USDA/ARS Southwest Watershed Re-  
 582 search Center*, (EPA/600/R-08/134).
- 583 Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H.  
 584 (2017). Feature selection: A data perspective. *ACM Comput. Surv.*, 50(6).  
 585 <https://doi.org/10.1145/3136625>
- 586 Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model pre-  
 587 dictions, 30. Retrieved from [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf)

- 589 Lyne, V. et al. (1979). Institute of engineers australia national conference. In (Vol.  
 590 79, pp. 89–93). Barton, Australia: Institute of Engineers Australia.
- 591 Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of  
 592 the Econometric Society*, 245–259.
- 593 Maricopa County, F. C. D. of. (2020). *Flood Warning Program - Program Re-  
 594 view*. Maricopa County Flood Control District. Retrieved from [https://alert.fcd.maricopa.gov/alert/FWP\\_report\\_FINAL.pdf](https://alert.fcd.maricopa.gov/alert/FWP_report_FINAL.pdf)
- 595 Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions,  
 596 methods, and applications in interpretable machine learning. *Proceedings  
 597 of the National Academy of Sciences*, 116(44), 22071–22080. <https://doi.org/10.1073/pnas.1900654116>
- 598 Nathan, R. J. et al. (1990). Evaluation of automated techniques for base flow and  
 599 recession analyses. *Water Resources Research*, 26(7), 1465–1473. <https://doi.org/10.1029/WR026i007p01465>
- 600 Neff, B. P. et al. (2005). *Base flow in the great lakes basin* (Report No. 2005-5217).  
 601 Reston, VA. <https://doi.org/10.3133/sir20055217>
- 602 Ni, L. et al. (2020). Streamflow forecasting using extreme gradient boosting model  
 603 coupled with gaussian mixture model. *Journal of Hydrology*, 586, 124901.
- 604 O'Donnell, F. C. et al. (2016). 12th Biennial Conference of Research on the Col-  
 605 orado Plateau. In.
- 606 Reitz, M. et al. (2017). Annual Estimates of Recharge, Quick-Flow Runoff, and  
 607 Evapotranspiration for the Contiguous US Using Empirical Regression Equations.  
 608 *JAWRA Journal of the American Water Resources Association*, 53(4), 961–983.  
<https://doi.org/10.1111/1752-1688.12546>
- 609 Rozos, E. et al. (2021). Machine learning in assessing the performance of hydrologi-  
 610 cal models. *Hydrology*. <https://doi.org/10.3390/hydrology9010005>
- 611 Santhi, C. et al. (2008). Regional estimation of base flow for the conterminous  
 612 United States by hydrologic landscape regions. *Journal of Hydrology*, 351(1-2),  
 613 139–153. <https://doi.org/10.1016/j.jhydrol.2007.12.018>
- 614 Scanlon, B. R. et al. (2006). Global synthesis of groundwater recharge in semiarid  
 615 and arid regions. *Hydrological Processes*, 20(15), 3335–3370. <https://doi.org/10.1002/hyp.6335>
- 616 Schmidt, L. et al. (2020). Challenges in applying machine learning models for hy-  
 617 drological inference: A case study for flooding events across germany. *Water  
 618 Resources Research*, 56(5), e2019WR025924. <https://doi.org/https://doi.org/10.1029/2019WR025924>
- 619 Sheppard, P. et al. (2002). The climate of the US Southwest. *Climate Research*, 21,  
 620 219–238. <https://doi.org/10.3354/cr021219>
- 621 Singh, S. K. et al. (2018). Towards baseflow index characterisation at national  
 622 scale in New Zealand. *Journal of Hydrology*, 568, 646–657. <https://doi.org/10.1016/j.jhydrol.2018.11.025>
- 623 Sloto, R. et al. (1996). *HYSEP: A Computer Program for Streamflow Hydrograph  
 624 Separation and Analysis* (No. 96-4040). USGS. <https://doi.org/10.3133/wri964040>
- 625 Szczepanek, R. (2022). Daily streamflow forecasting in mountainous catchment  
 626 using XGBoost, LightGBM and CatBoost. *Hydrology*, 9(12), 226.
- 627 Tan, X. et al. (2020). Global Changes in Baseflow Under the Impacts of Changing  
 628 Climate and Vegetation. *Water Resources Research*, 56(9), e2020WR027349.  
<https://doi.org/10.1029/2020WR027349>
- 629 Taylor, R. G. et al. (2013). Ground water and climate change. *Nature Climate  
 630 Change*, 3(4), 322–329. <https://doi.org/10.1038/nclimate1744>
- 631 USDA. (2009). *Part 630 hydrology national engineering handbook, chapter 7: Hydro-  
 632 logic soil groups*. USDA Natural Resources Conservation Service.
- 633 USGS. (2010). Water resources data for the united states water year 2007. <http://wdr.water.usgs.gov/wy2007/search.jsp>.

- 644      USGS. (2018). Water science glossary. Retrieved from <https://www.usgs.gov/special-topics/water-science-school/science/water-science-glossary>
- 645      Wolock, D. M. (2003). Estimated mean annual natural ground-water recharge in the  
646      conterminous united states. <https://doi.org/10.5066/P9FSSVF3>.
- 647      Woodhouse, C. A. et al. (2022). Upper Gila, Salt, and Verde Rivers: Arid Land  
648      Rivers in a Changing Climate. *Earth Interactions*, 26(1), 1–14. <https://doi.org/10.1175/EI-D-21-0014.1>
- 649      Yao, Y. et al. (2018). Role of Groundwater in the Dryland Ecohydrological Sys-  
650      tem: A Case Study of the Heihe River Basin. *Journal of Geophysical Research: Atmospheres*, 123(13), 6760–6776. <https://doi.org/10.1029/2018JD028432>
- 651      Ying, X. (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>
- 652      Zomer, R. J. et al. (2022). Version 3 of the Global Aridity Index and Potential  
653      Evapotranspiration Database. *Scientific Data*, 9(1), 409. <https://doi.org/10.1038/s41597-022-01493-1>
- 654
- 655
- 656
- 657
- 658
- 659