THESIS

ARE WILD AND SCENIC RIVERS REALLY "FREE-FLOWING"?

Submitted by

Kathryn Rosalie Willi

Department of Ecosystem Science and Sustainability

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2021

Master's Committee:

Advisor: Stephanie K. Kampf
Co-Advisor: Matt R. V. Ross

Lynn Badia

ABSTRACT

ARE WILD AND SCENIC RIVERS REALLY "FREE-FLOWING"?

This study quantified the "*free-flowing*" character of wild and scenic river watersheds by first developing linear regression models to predict the "*natural condition*" of a river's magnitude, timing, frequency, and variability of flows. We compared these estimates of "*natural*" flow to the observed values for stream gages within wild and scenic river watersheds and found that nearly half (45.1%) had at least one altered flow metric. This was significantly lower ($p < 0.05$) than the fraction of altered gages outside wild and scenic river watersheds, and supported our other conclusion that wild and scenic rivers are associated with protected areas. On the other hand, wild and scenic river watersheds had a significantly higher ($p < 0.05$) fraction of gages with dam storage densities ≥ 100 megaliters$\cdot$km^{-2} than gages outside wild and scenic river watersheds. Because the Wild and Scenic Rivers Act was designed as a complement to dam development, many wild and scenic rivers are designated in direct response to the threat of dam construction, or to counterbalance special rivers that have already been dammed. We posit that this biases wild and scenic river designations towards locations where dam development is common. Our study's findings expose a paradox in how a wild and scenic river designation can fully "*protect and enhance"* a river's free-flowing character. True protection of these special resources does not stop at designation, and requires additional support from managing agencies and stewardship groups to make improvements to their watersheds.

ACKNOWLEDGEMENTS

I would first like to sincerely thank my advisors, Stephanie Kampf and Matt Ross. I would not have been able to make it through this journey without their insight, guidance, and extreme kindness. I am also grateful to Abby Eurich, who was instrumental in helping me first navigate the world of linear regression modelling in R. Lynn Badia’s literature recommendations were also critical in the completion of this work, particularly in helping give context to my findings. I would also like to take this opportunity to thank Jennifer Back with the National Park Service, who has been a mentor to me throughout this wild (and scenic!) experience and who made this research possible.

TABLE OF CONTENTS

1. INTRODUCTION

The National Wild and Scenic River System is a collection of rivers that are supposed to receive the highest level of river protection in the United States (US). From the far reaches of Alaska to the jungles of Puerto Rico, the National System currently "*protects*" over 500 streams for their free-flowing character, their water quality, and the "*outstandingly remarkable values*" that make them unique at a regional or national scale (16 U.S. Code 1271-1278). Safeguarded under the National Wild and Scenic Rivers Act of 1968, this collection of rivers shared between the US Forest Service, the US Fish and Wildlife Service, the Bureau of Land Management, and the National Park Service is seen internationally as a model method for river conservation at a time when river protection has become a global priority (Palmer, 2017; IUCN, 2020).

Under the Wild and Scenic Rivers Act, river managing agencies must "*protect and enhance*" a wild and scenic river's free-flowing character, which the Act defines as "*existing or flowing in natural condition without impoundment, diversion, straightening, rip-rapping, or other modification of the waterway*" (16 U.S. Code 1271-1278). To maintain a wild and scenic river's free-flowing status, the Act prohibits federally funded water resource projects "*on or directly affecting*" the river, and projects up or downstream of the designated corridor that "*invade the area or unreasonably diminish*" the values that had it designated. Although the designated corridor is protected from federally assisted water resource projects, a wild and scenic river designation does not fundamentally protect its entire watershed. Upstream reaches, and even designated reaches, may still have impoundments, diversions, or other circumstances that have the potential to modify the "*natural condition*" of wild and scenic rivers by altering the magnitude, timing, and variability of their flow regimes (Poff et al., 1997). Therefore, the goal of this study is to assess the extent to which wild and scenic river corridors and their contributing watersheds are truly “*free-flowing*”.

2. BACKGROUND

Currently there are 230 designated wild and scenic rivers in the system[1] (Figure 1), though many designations incorporate more than one river. For example, the Smith Wild and Scenic River in California embodies the Smith River as well as 62 of its tributaries, while the Wekiva Wild and Scenic River in Florida includes the Wekiva River as well as Rock Springs Run and Blackwater Creek. Consequently, the system's true number of protected streams lies closer to 520 (updated from Palmer, 2017). The National Wild and Scenic River System captures many different types of rivers: from the historic Delaware River in the mid-Atlantic, to the desert oasis of Surprise Canyon Creek in Death Valley, to the mountainous headwaters of the Snake River in Wyoming. Spatially, there are large concentrations of wild and scenic rivers in Alaska, the Pacific Northwest, and the Great Lakes regions, though 40 states and Puerto Rico have at least one wild and scenic river. The lengths of designated reaches range from under one kilometer to over 600 kilometers. Watershed sizes also vary substantially; the wild and scenic river with the largest contributing area is the Missouri Wild and Scenic River in South Dakota and Nebraska (785,119 km^2), while the smallest is Spring Creek in Oregon (2.1 km^2). Regardless of these differences, all wild and scenic rivers are protected for the values that got them designated: their water quality, their outstandingly remarkable river values, and their free-flowing condition.

[1] The official number of designated wild and scenic rivers is 226. However, the Delaware Wild and Scenic River, the Rio Grande Wild and Scenic River, and the Klamath Wild and Scenic River have been split into separate wild and scenic rivers for this analysis based on how the Interagency Wild and Scenic River Coordinating Council treats them.

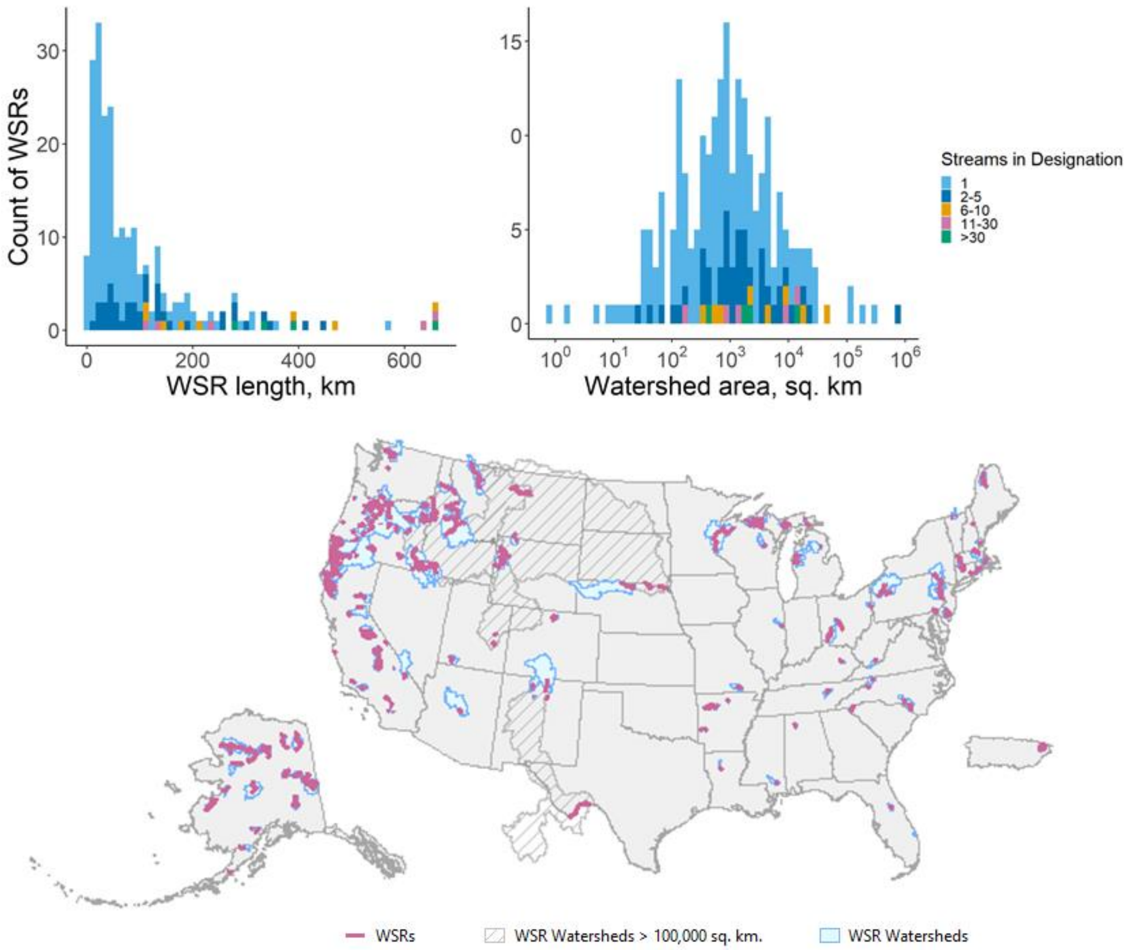

Figure 1. Histograms depicting the distribution of wild and scenic river (WSR) lengths and watershed sizes, and a map of the WSR System and its contributing watersheds. The River Styx in Oregon, WSRs in Alaska and Puerto Rico, and WSRs whose watersheds are > 100,000 km^2 were excluded from our analysis.

Because streamflow is a dominant driver of a river's water quality, ecological diversity, and channel geomorphology, it is often considered the "master variable" of river function (Power et al., 1995; Poff et al., 2010). Therefore, the "*natural condition*" of streamflow is not only a critical component of the free-flowing requirements of the Wild and Scenic Rivers Act, it is also intimately linked to a wild and scenic river's water quality and the status of many outstandingly remarkable values. The fluctuations in streamflow and their timing have major implications on the concentrations of chemical constituents in the stream, how they are transported, and how they are transformed (Kagawa 1992; Tu, 2009; Wei et al., 2009; Fantin-Cruz et al., 2016). Many outstandingly remarkable values also directly rely on the magnitude, timing, and range of natural flows. For example, the Virgin Wild and Scenic River in Utah has been designated in part for its exceptional ecological value; the rare plant communities and cottonwood galleries that exist there depend on the river's seasonal flooding (US Department of Interior, 2013b). The range and variability of streamflow, coupled with changing sediment loads, are often critical in supporting a rich and diverse community of native plants and animals (Junk, Bayley, and Sparks, 1989; Poff et al., 1997; Hart and Finelli, 1999). Another example is the Snake River Headwaters Wild and Scenic River, whose geologic values rely on specific flow magnitudes for the unique geomorphologic features found there (US Department of Interior, 2013a). Dynamic flows can also support a wide range of recreation opportunities (Brown, Taylor, and Shelby, 1992). For instance, the Bluestone Wild and Scenic River's recreational values rely on high flows for paddling activities, and lower flows for safe wading, fishing, and swimming (Nadeau et al., 2018).

Yet, there is limited information regarding the extent to which streams have been modified within wild and scenic river watersheds. Some assessments have been conducted for individual wild and scenic rivers (e.g., Narvaez and Homsey, 2016; OARS for the Assabet, Sudbury, and Concord Rivers, 2019; Elmore et al., 2020), but there has been no comprehensive effort to assess the flow conditions of the entire system. At a broader level of investigation, Grill et al. (2019) found that 63% of the world's large rivers are no longer free-flowing, while Harrison-Atlas et al. (2017) found that 49% of all river miles in the Western US are hydrologically altered. Moreover, a recent report by the US Geological Survey (USGS) found that 80% of the USGS stream gages they assessed had altered flow (Eng et al., 2019). Extrapolating from Eng et al. (2019)'s findings, Carlisle et al. (2019) posits that over one third of all streams in the contiguous US have human-modified streamflow. An analysis of wild and scenic river water quality also revealed that states have identified multiple wild and scenic rivers as having impaired streamflow under the Clean Water Act (Willi and Back, 2018). Collectively these findings suggest that hydrologic alteration may be widespread in the system, despite wild and scenic river legal protections.

Though not specific to wild and scenic rivers, several frameworks have been developed to address questions surrounding the hydrologic character of streams and their watersheds. Thornbrugh et al. (2018) developed a watershed assessment framework that uses nationally available datasets on land use, dam density, fertilizer application rates, and other watershed characteristics to assess the likely extent of hydrologic modification. Along those lines, the US Environmental Protection Agency (EPA) developed the Watershed Index Online tool that allows users to develop watershed condition indices based on a similar set of nationally available datasets for any HUC-12 watershed in the conterminous US (EPA, 2017). Because these watershed assessment strategies are intended to be used for all streams of the US with a consistent approach, they do not incorporate actual streamflow observations into their frameworks.

An example of a localized approach for measuring a stream's hydrologic character is the Nature Conservancy (2009)'s Indicators of Hydrologic Alteration tool, where users input daily streamflow data to calculate ecologically relevant flow metrics through time. With information related to changes made to a stream's watershed, such as dam development or a change in land use, the tool calculates the mean or variance of the flow statistics across the time period before and after the disturbance. Comparative statistical analysis between these periods then allows the user to determine whether the disturbance significantly changed streamflow. Because this tool requires knowledge of when changes in the watershed occurred, it is most useful for answering river-specific questions related to streamflow.

Another common way of assessing hydrologic character is to develop models that predict natural streamflow across individual states or regions, and compare the modeled values to observed streamflow (e.g., Sanborn and Bledsoe, 2006; Ries et al., 2017; Eurich, 2020). One such application has been implemented as part of the National Hydrography Dataset, which provides estimates of mean monthly and annual streamflow across all stream features in its dataset (McKay et al., 2012; Moore et al., 2019). These streamflow metrics are first estimated by using a water balance model (McCabe and Wolock, 2011), thereby providing an estimate of natural streamflow. These natural estimates can then be adjusted using a regression equation to match nearby stream gage records, providing a value more consistent with observed streamflow patterns.

In France, Snelder et al. (2009) developed boosted regression trees that grouped streams by their flow regimes to predict natural flow metrics related to the frequency, magnitude, timing, and variability of daily flows. Using unmodified streamflow stations to train the models, they found watershed characteristics such as slope, watershed area, temperature, and soil permeability were effective predictor variables for identifying a stream's flow regime.

Similarly, Eng et al. (2019) developed random forest models predicting natural streamflow metrics across the contiguous US that could be compared to observed streamflow, landscape characteristics including urbanization, dam density and agriculture, and climate variability. Their assessment used stream gages in the GAGES-II database (Falcone, 2011), which includes both small headwater watersheds as well as large river basins up to 49,600 km^2. They found that 80% of the 3,355 gages analyzed were altered to some degree, with landscape characteristics being a more dominant driver of changes in streamflow compared to climate variability. Here, we build on the methods of Eng et al. (2019) by developing regional regression models that capture the "*natural condition*" of the flow regime for USGS stream gages within and outside wild and scenic river watersheds.

3. METHODS

This analysis focuses on wild and scenic river watersheds within the contiguous US; including watersheds within Alaska and Puerto Rico would have required datasets that are not as readily available and nationally consistent. For example, there are only four USGS stream gages in the entire wild and scenic river system of Alaska, which spans a collective area of 131,618 km^2. We also excluded the Snake Wild and Scenic River in Oregon and Idaho, the Missouri Wild and Scenic River in Montana, South Dakota, and Nebraska, the Green Wild and Scenic River in Utah, and the Rio Grande Wild and Scenic River in Texas from our study because their watersheds are so large (>100,000 km^2) that this type of analysis would not be appropriate. Furthermore, we already know these large wild and scenic river watersheds are altered due to their national importance for irrigation, commerce, and hydropower (Reisner, 1993). The River Styx in Oregon was also excluded because both the designated reach and its watershed are almost exclusively underground. This reduces our analysis to 196 of the 230 designated wild and scenic rivers (Figure 1).

To characterize the "*free-flowing*" status of wild and scenic river watersheds, we evaluated whether their flow regimes had been substantially altered from their "*natural condition*". We define "*natural condition*" as an unmodified river's mean annual flow, high and low flow magnitude (Q99, Q1) and variability (CV_{high}, CV_{low}), frequency (F_{high}, F_{low}) and duration (D_{high}, D_{low}) of high and low flow pulses, and the seasonal distribution (S) of daily flows. To quantify these components of the flow regime, we developed linear regression models using reference-quality USGS stream gages to predict the "*natural condition*" of streams within wild and scenic river watersheds.

3.1 Calculating "Free-Flow" Metrics

Daily streamflow for each stream gage used in our study was downloaded from the USGS National Water Information System using the 'dataRetrieval' package in R (De Cicco et al., 2018; USGS, 2020). All data was then area-normalized to $mm \cdot day^{-1}$. To best represent current flow conditions, we only analyzed streamflow data from the period of 1980 to 2019. 1980 was selected as the cutoff because the development of most major water projects that could significantly modify flow (e.g., large reservoirs and large-scale diversions) occurred before then, and the major patterns in land and water management have not changed significantly since then (Eng et al., 2019).

The one-day low and high magnitudes (Q1 and Q99, respectively) were calculated as the 1^{st} and 99^{th} percentile non-exceedance daily flow from each stream gage's chosen period of record (i.e., 1980-2019). Annual high and low flow variability (CV_{high} and CV_{low}, respectively) were calculated by identifying each water year's maximum and minimum daily flow then determining their coefficient of variation:

$$CV_{low} = sd_{low} / mean_{low}$$

$$CV_{high} = sd_{high} / mean_{high}$$

where sd_x represents the annual standard deviation of the minimum or maximum flows and $mean_x$ represents the annual mean low and high flow. High and low flow frequency (F_{high} and F_{low}, respectively) were calculated by computing each year's total number of high or low pulses, then averaging the number of high or low pulses across all years. A high pulse was defined as a set of consecutive days (≥ 2 days) in which the flow was at or above the 90^{th} percentile flow, and a low pulse was defined as a set of consecutive days in which the flow was at or below the 10^{th} percentile non-exceedance daily flow. High flow duration (D_{high}) was calculated as the mean number of days per water year that the flow was at or above the 90th percentile non-exceedance daily flow. Low flow duration (D_{low}) was calculated as the mean number of days per water year that the flow was at or below the 10th percentile flow. Seasonality was calculated as the fraction of the total annual flow (in mm) that falls within a given season, for each year. Those fractions were then averaged across each year to get the mean fraction of flow for each season. See Table 1 for a list of all flow metrics analyzed in our study.

Table 1. Flow metrics used to represent a stream's *"natural condition*". Asterisks represent flow metrics used in Eng et al. (2019).

Flow Metric	**Description**
Q_{mean}	Mean annual flow
Autumn	Mean annual fraction of total flow in September, October, and November
Winter Flows	Mean annual fraction of total flow in December, January, and February
Spring Flows	Mean annual fraction of total flow in March, April, and May
Summer	Mean annual fraction of flow in June, July, August
Q99*	99th percentile non-exceedance flow
Q1*	1st percentile non-exceedance flow
CV_{high}*	Coefficient of variation of annual maximum daily flows
CV_{low}*	Coefficient of variation of annual minimum daily flows
F_{high}*	Mean number of annual flow pulses greater than the 90th percentile non-exceedance flow
F_{low}*	Mean number of annual flow pulses less than the 10th percentile non-exceedance flow
D_{high}*	Mean annual duration of flow pulses greater than the 90th percentile non-exceedance flow
D_{low}*	Mean annual duration of flow pulses less than the 10th percentile non-exceedance flow

3.2 Selecting Reference Watersheds

Models for predicting natural streamflow were developed using a subset of stream gages identified as being of reference quality by Falcone (2011). From Falcone (2011), we selected only stream gages whose watersheds were smaller than 1,500 km^2 to minimize the potential for within-basin variability in streamflow generation (Hammond and Kampf, 2020; Eurich, 2020). Stream gages whose watersheds crossed international boundaries were removed from the analysis due to limitations in the spatial extent of most datasets in this study. Reference watersheds were further screened for characteristics with known impacts to the natural flow regime including transbasin diversions, dams, major wastewater treatment facilities, urban land cover, and cropland.

To identify transbasin diversions we used the High-Resolution National Hydrography Dataset (Moore et al., 2019) to detect ditches, canals, and pipelines that crossed watershed boundaries. We also identified watersheds with dam storage densities of over 100 megaliters·km^{-2}, which Eng et al. (2019) used as the threshold at which a watershed is categorized as dam altered. Using a similar methodology to Harrison-Atlas et al. (2017), we used the National Inventory of Dams' normal storage variable (USACE, 2019) to compute dam storage densities. Major wastewater treatment facilities that discharge $\geq$ 1 million gallons·day^{-1} were identified using the EPA's Wastewater Treatment Plants geodatabase (EPA, 2020). Land cover characteristics including percent cropland and imperviousness were computed from the 2011 National Landcover Dataset (Wickham et al., 2014). Urban watersheds were defined as watersheds with 5% or greater mean impervious land cover as suggested by Bhaskar and Welty (2012). We considered a watershed with 2.5% cropland or greater to be agricultural. This cutoff was based on visual inspection of the mean annual flow of a stream gage's watershed against the mean aridity index of the watershed; at a given aridity index, watersheds with over 2.5% cropland tend to have lower mean annual streamflow than those with less than 2.5% cropland (Figure A1 in the Appendix).

The final dataset included watersheds with no transbasin diversions, dam storage densities under 100 megaliters·km^{-2}, no major wastewater treatment facilities, less than 5% mean urban impervious cover, and less than 2.5% of crop land cover. Beyond these watershed characteristic requirements, only stream gages with at least 20 years of complete data (i.e., a 100% complete period of record for a water year) between the 1980-2019 period were included, leaving a dataset of 642 reference gages.

3.3 Grouping Watersheds of the Contiguous US

Because of the diversity of climate, soil, and land cover conditions for the contiguous US, strong models could not be developed using all of the gages together. Instead, we chose to divide them into regions, each analyzed individually. However, our strict criteria for selecting reference watersheds created a set of stream gages that were not evenly distributed in space, leaving large swaths of the US unrepresented, particularly the Midwest (Figure 2). If we based regionalization on the National Rivers and Stream Assessment ecoregions (EPA, 2016), for instance, there would be only one viable stream gage in the Temperate Plains region, one viable stream gage in the Northern Plains region, and only 11 viable stream gages in the Upper Midwest. Therefore, we developed a grouping mechanism that relied on other factors that are linked to streamflow generation. We grouped watersheds together based on their mean aridity index, soil sand content, and geographic location. Each watershed was identified as having a mean aridity index above or below one, having a mean soil content above or below 53% sand, and being either predominantly east or west of the 100th meridian.

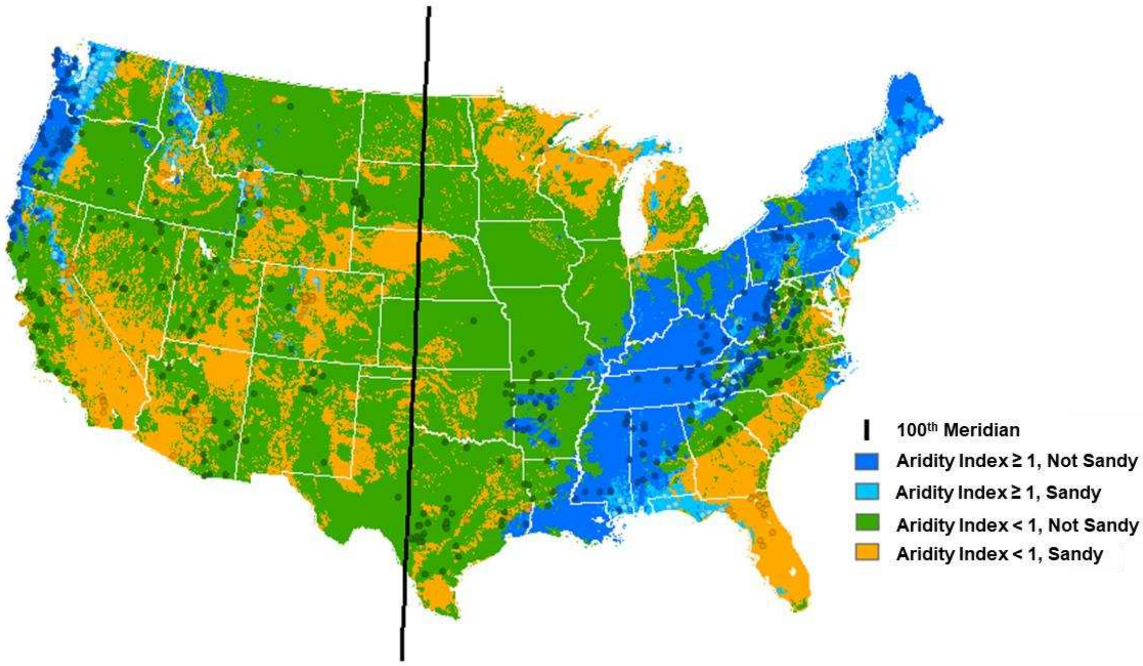

Figure 2. Map showing model regions, and the stream gages associated with them. We divided the contiguous US into eight regions: east, aridity index ≥ 1 (surplus, S), not sandy (E-S-NS, n=146); east, aridity index ≥ 1, sandy (E-S-S, n=53); east, aridity index < 1 (deficit, D), not sandy (E-D-NS, n=97); east, aridity index < 1, sandy (E-D-S, n=29); west, aridity index ≥ 1, not sandy (W-S-NS, n=74); west, aridity index ≥ 1, sandy (W-S-S, n=58); west, aridity index < 1, not sandy (W-D-NS, n=111); and west, aridity index < 1, sandy (W-D-S, n=74).

The mean aridity index of a watershed, which is calculated as the mean ratio between precipitation and potential evapotranspiration, was used to define sub-regions because it broadly captures a given watershed's relationship to water; if a watershed's mean aridity index falls below one, it indicates that the watershed's ability to lose water to the atmosphere is greater than the amount of precipitation it receives (i.e., water deficit, D). Inversely, if a watershed's mean aridity index is greater than one, the watershed generally receives more water than it can lose to the atmosphere (i.e., water surplus, S).

The sand content of soil was used because locations with sandy soils tend to behave differently than non-sandy soils when assessing regional patterns of streamflow; sandy soils typically lead to faster rates of infiltration and higher hydraulic conductivities (Twarakavi, Šimůnek, and Schaap, 2010). The chosen threshold of 53% sand was selected based on the USDA Natural Resource Conservation Service's soil texture triangle; all soil textures are considered "sandy" past 53% sand (Vanlear, 2019). Moreover, this closely captures Twarakavi, Šimůnek, and Schaap (2010)'s sand-dominated hydraulic classification groups. Lastly, the 100th meridian was selected as a geographic grouping mechanism because it approximates the point at which the elevation begins to rise on its way towards the Rocky Mountains, and levels of precipitation tend to change (Gesch et al., 2002). Ultimately, the combination of these aggregating mechanisms resulted in a total of eight unique watershed groups (Figure 2): east, aridity index ≥ 1 (water surplus, S), not sandy (E-S-NS, n=146); east, aridity index ≥ 1, sandy (E-S-S, n=53); east, aridity index < 1 (water deficit, D), not sandy (E-D-NS, n=97); east, aridity index < 1, sandy (E-D-S, n=29); west, aridity index ≥ 1, not sandy (W-S-NS, n=74); west, aridity index ≥ 1, sandy (W-S-S, n=58); west, aridity index < 1, not sandy (W-D-NS, n=111); and west, aridity index < 1, sandy (W-D-S, n=74). For each of the eight watershed groups, 13 linear models were developed to predict each of the flow metrics of interest, resulting in 104 unique models.

3.4 Developing the Models

Watershed boundaries for the stream gages used in model development were obtained from the GAGES-II database (Falcone, 2011). Using the geospatial software ArcGIS (ESRI, 2017), 32 variables related to climate, topography, soil properties, geology, and land cover were computed for each watershed (see Table 2). These variables were chosen as potential predictors because they have been used in previous models that predict streamflow statistics, and/or have an established relationship to streamflow generation (Hortness and Berenbrock, 2001; Snelder et al., 2009; Carlisle et al., 2010; Gotvald, 2017; Eurich, 2020).

Table 2. Watershed and stream attributes used as predictor variables in the regression models.

Variable	Source
Stream Gage Latitude (NAD83)	NWIS (USGS, 2020)
Stream Gage Longitude (NAD83)	NWIS (USGS, 2020)
Stream Gage Elevation (m)	National Elevation Dataset (NED; Gesch et al., 2002)
Watershed Area (km^2)	NED (Gesch et al., 2002)
Mean Watershed Elevation (m)	NED (Gesch et al., 2002)
Median Watershed Elevation (m)	NED (Gesch et al., 2002)
Mean Watershed Slope (%)	NED (Gesch et al., 2002)
Percent Watershed with Slope over 30 (%)	NED (Gesch et al., 2002)
Dominant Aspect	NED (Gesch et al., 2002)
Mean Base-Flow Index	USGS (Wolock, 2003)
Mean Erodibility Factor	SSURGO (NRCS USDA, 2019)
Mean Soil Organic Matter (%)	SSURGO (NRCS USDA, 2019)
Mean Soil Permeability (mm)	SSURGO (NRCS USDA, 2019)
Mean Clay (%)	SSURGO (NRCS USDA, 2019)
Mean Sand (%)	SSURGO (NRCS USDA, 2019)
Mean Silt (%)	SSURGO (NRCS USDA, 2019)
Mean Annual Precipitation (mm)	PRISM (Daly, 2013)
Mean Annual Temperature (℃)	PRISM (Daly, 2013)
Mean Spring (March-May) Precipitation (mm)	PRISM (Daly, 2013)
Mean Spring (March-May) Temperature (℃)	PRISM (Daly, 2013)
Mean Summer (June-August) Precipitation (mm)	PRISM (Daly, 2013)
Mean Summer (June-August) Temperature (℃)	PRISM (Daly, 2013)
Mean Fall (September-October) Precipitation (mm)	PRISM (Daly, 2013)
Mean Fall (September-October) Temperature (℃)	PRISM (Daly, 2013)
Mean Winter (November-January) Precipitation (mm)	PRISM (Daly, 2013)
Mean Winter (November-January) Temperature (℃)	PRISM (Daly, 2013)
Mean Aridity Index	GridMET (Abatzoglou, 2013), PRISM (Daly, 2013)
Mean Snow Persistence (%)	MODIS (Hammond, Saavedra, and Kampf, 2020)
Percent Forest as Landcover (%)	NLCD 2011 (Wickham et al., 2014)
Percent Herbaceous as Landcover (%)	NLCD 2011 (Wickham et al., 2014)
Percent Shrubland as Landcover (%)	NLCD 2011 (Wickham et al., 2014)
Percent Wetland as Landcover (%)	NLCD 2011 (Wickham et al., 2014)

For each watershed group, we reduced the number of predictor variables to be included in the model development process by removing those that had over 85% correlation with other predictor variables. This was performed in R using the 'caret' package, which calculates the mean absolute correlation of each variable and removes those with the largest mean absolute correlation (Kuhn et al., 2020). To best meet the model assumptions of normality and equal variance of the residuals, all streamflow statistics were square-root transformed. Multiple linear regression models were then applied using the independent variables to predict each of the square-root transformed streamflow response variables. The final multiple linear regression models were selected using Akaike's Information Criterion (AIC) with the 'MuMIN' package in R (Bartón, 2020). All possible subsets of the predictor variables were created and ranked by their AIC value, and the model with the lowest AIC was selected as the final model for each flow metric.

Model performance was primarily assessed using the Nash-Sutcliffe Efficiency (NSE), percent bias (PBIAS), and the ratio of the root mean square error to the standard deviation of the measured data (RSR). Based on Moriasi et al. (2007), a model is satisfactory if NSE > 0.50, RSR < 0.70, and PBIAS < 25. If a model did not meet these thresholds, it was removed from our analysis. Models were also removed from the analysis if assumptions of normality and equal variance could not be met even after square-root transforming. For a visual example of a satisfactory model versus an insufficient model, see Figure A2 in the Appendix.

Due to the limited size of our datasets, we did not divide data into training and test datasets. Instead, we tested the model performance using the predicted sum of squares (PRESS) statistic, which is a leave-one-out measure of cross validation that expresses how sensitive a model is to any single observation used in the analysis (Allen, 1971).

3.5 Predicting and Assessing "Natural" Flow

We used the subset of models that met our performance criteria to predict natural flow statistics at USGS stream gage sites within wild and scenic river watersheds. These gages were then classified based on their location within a wild and scenic river using the R package 'nhdplusTools' (Blodgett, 2019). Like the stream gages used to develop the models, only those with over 20 years of complete data between the 1980-2019 period and with watersheds under 1,500 km^2 were used. These stream gages were also screened to remove any that were not monitoring stream channels (i.e., those on canals, ditches, or springs). For wild and scenic river stream gages that were not within the GAGES-II database, watersheds were delineated using the Watershed tool in ArcGIS (ESRI, 2017). Collectively, 763 USGS-maintained gages were found within wild and scenic river watersheds, though only 324 met our data requirements. These gages were in only 90 of the 196 wild and scenic river watersheds assessed in this study. 138 gages (42.7%) were directly along or within 10 km of a wild and scenic river corridor, while the rest were over 10 km upstream from a designated reach. 98 (30.3%) of the gages were of reference quality and had been used in the development of the models (Figure 3).

The final models were also used to predict flow statistics for stream gages outside wild and scenic river watersheds as a means of evaluating how wild and scenic river flow conditions compare to other rivers in the nation. We selected all other stream gages in Falcone (2011) that were not used in the model's development that still met our data record requirements (i.e., at least 20 complete years of record from 1980-2019 and were under 1,500 km^2), resulting in a comparison dataset of 2,696 nonreference stream gages.

The predicted values for each stream gage in our analysis were compared to the observed values to determine the extent that a flow metric has deviated from its "*natural condition*". This deviation was calculated as the ratio of the observed (O) to the predicted (P) flow metric. If the 95% prediction interval of the predicted value did not include the observed value, it was considered to have altered flow, which was categorized as either decreased or increased.

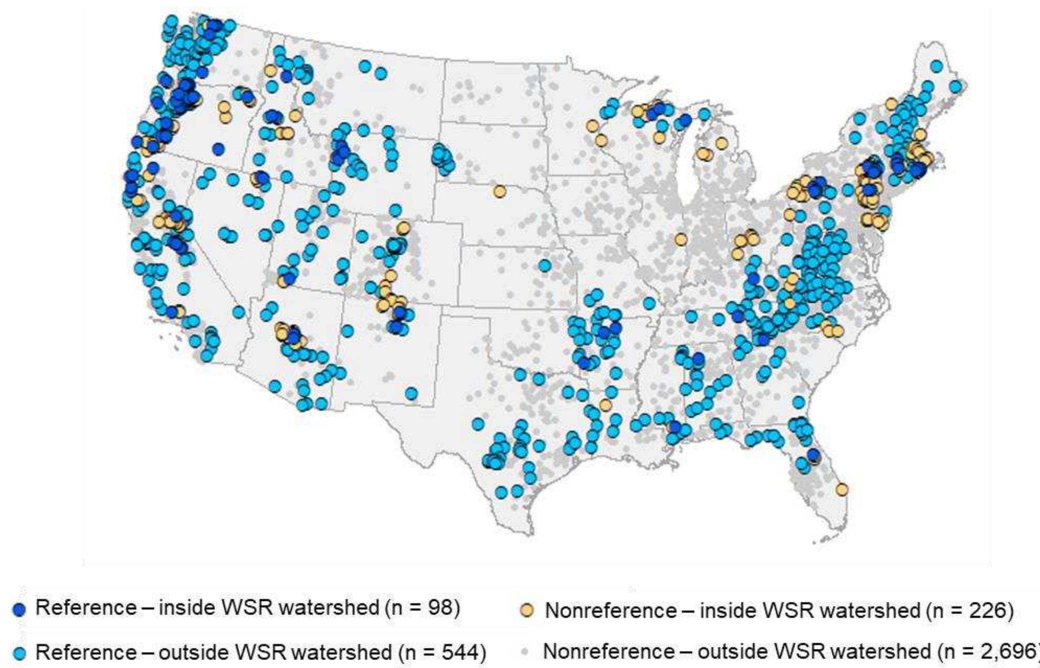

Figure 3. Map displaying all stream gages used in this study. Light blue and dark blue points represent reference watershed stream gages that were used in the development of the models. Orange points represent non-reference stream gages that were within a wild and scenic river (WSR) watershed but not used in model development. Grey points represent non-reference stream gages outside of a WSR watershed that were not used in model development.

4. RESULTS

Of the 104 models developed to predict natural flow regime metrics, 47 met our model performance requirements (Figure 4). There were no models of high flow duration (D_{high}), low flow duration (D_{low}), or low flow variability (CV_{min}) with satisfactory performance, so these flow statistics were entirely excluded from our analysis. More gages were analyzed for flow metrics associated with high flow occurrences (e.g., Q99, CV_{high}, S_{spring}, etc.) than low flow occurrences (e.g., Q1, CV_{low}, S_{winter}, etc.) because high flow models tended to perform better than low flow models. The region with the fewest metrics assessed was the W-D-NS region. See Figures A3-A7 in the Appendix for an in-depth review of each metric's performance based on NSE, RSR, PBIAS, model assumptions of normality and equal residuals, and PRESS.

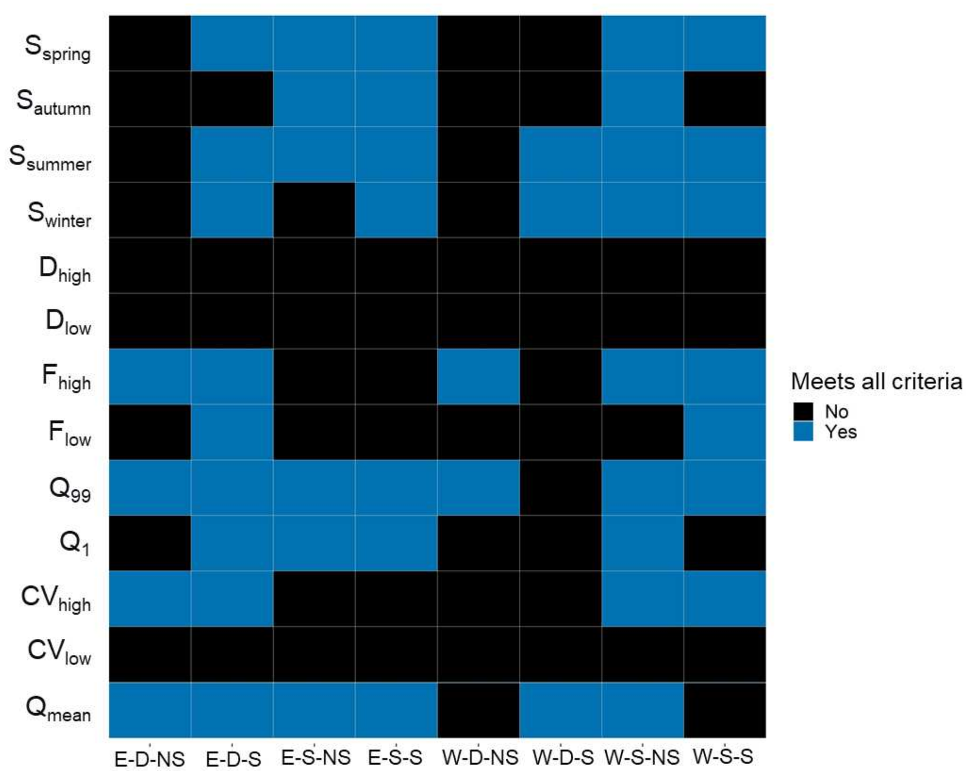

Figure 4. Matrix of models for each region and flow metric. Blue indicates that the model met our performance criteria and was used in the analyses.

Considering all flow metrics with sufficient models, nearly half (45.1%) of wild and scenic river gages had at least one flow metric that was altered relative to the predicted values (Figure 5). This was significantly less ($p < 0.05$) than the fraction of gages outside of wild and scenic river watersheds that had an altered flow metric (55.6%). Even though roughly half of all gages analyzed had at least one altered flow metric, each individual flow metric had a greater proportion of gages with unmodified flow. Wild and scenic river watersheds generally had a greater proportion of gages with unmodified flow than gages outside wild and scenic river watersheds. The two exceptions are the CV_{high} and the F_{low}, where the fraction of unmodified gages was significantly greater outside of wild and scenic river watersheds (Table 3).

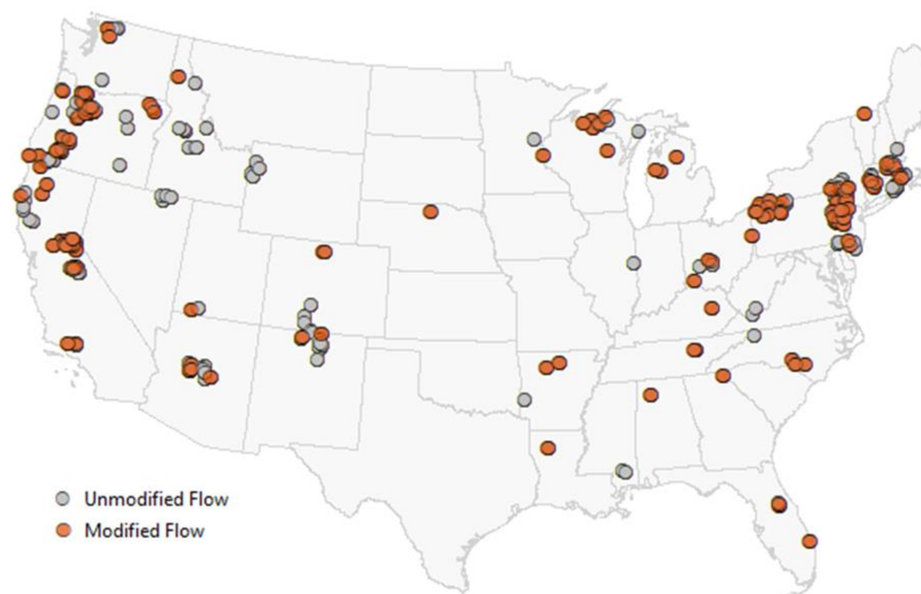

Figure 5. Gages in wild and scenic river watersheds that did or did not have streamflow within the 95% prediction intervals of the natural streamflow model predictions. Orange circles represent gages that had at least one flow metric that was considered modified (n=146). Grey circles represent gages that had no modified flow metrics (n=178).

Table 3. Percent of gages within or outside the 95% prediction interval of predicted flow for gages within wild and scenic river (WSR) watersheds and outside of WSR watersheds. Gages whose flow metrics were outside the range of the predicted flow's 95% prediction interval are categorized as having an increased or decreased observed value. Due to our model requirements, not all gages were used in the analysis of every flow metric. Asterisks indicate there was a significant ($p < 0.05$) difference in the proportion of unmodified gages within vs. outside of WSR watersheds.

Flow Metric	Within WSR Watershed				Outside WSR Watershed			
	Gage	Increased	Decreased	Unmodified	Gage	Increased	Decreased	Unmodified
Q_{mean}*	228	3.5	10.5	86.0	2675	14.5	8.5	77.0
S_{winter}*	176	5.1	1.7	93.2	1144	7.0	8.2	84.8
S_{spring}	231	1.7	13.9	84.4	1553	1.7	15.8	82.5
S_{summer}*	260	5.8	4.2	90.0	1882	14.5	3.9	81.6
S_{autumn}	175	12.0	2.3	85.7	1128	12.9	2.3	84.8
Q99	295	3.1	13.6	83.4	2911	8.5	11.0	80.5
Q1*	193	9.8	7.8	82.4	1440	16.7	20.7	62.6
CV_{high}*	106	10.4	11.3	78.3	1450	5.6	6.0	88.4
F_{high}*	164	8.5	1.2	90.2	1902	14.3	5.0	80.7
F_{low}*	56	19.6	5.4	75.0	425	10.4	3.5	86.1

Next we evaluated what types of flow modifications were present in wild and scenic river watersheds and found that 52.5% (n=170) had watershed characteristics known to alter flow. The top sources of basin alteration were dam storage densities ≥ 100 megaliters·km^{-2} (n=64) and wastewater treatment facilities discharging ≥ 100 MGD (n=63). Spatially, watersheds with high dam storage densities were prevalent in the Northeast and along the West Coast; none of the gages in the Southeast had dam storage densities ≥ 100 megaliters·km^{-2}. Crop land covers ≥ 2.5% were present primarily in watersheds of the Northeast and Midwest. Few wild and scenic river gages had impervious land covers ≥ 5% (n=27), and most of them were in the Northeast. Transbasin diversions and major wastewater treatment plants were found in wild and scenic river watersheds across the country (Figure 6). For the gages we flagged as having modified flow, 97 had a known basin alteration. Many of the gages with modified flow and no recognized watershed alteration were located in the West and Southeast (Figure 7). For more information related to the fraction of gages with watershed alterations by flow metric, see Figures A18-A27 in the Appendix.

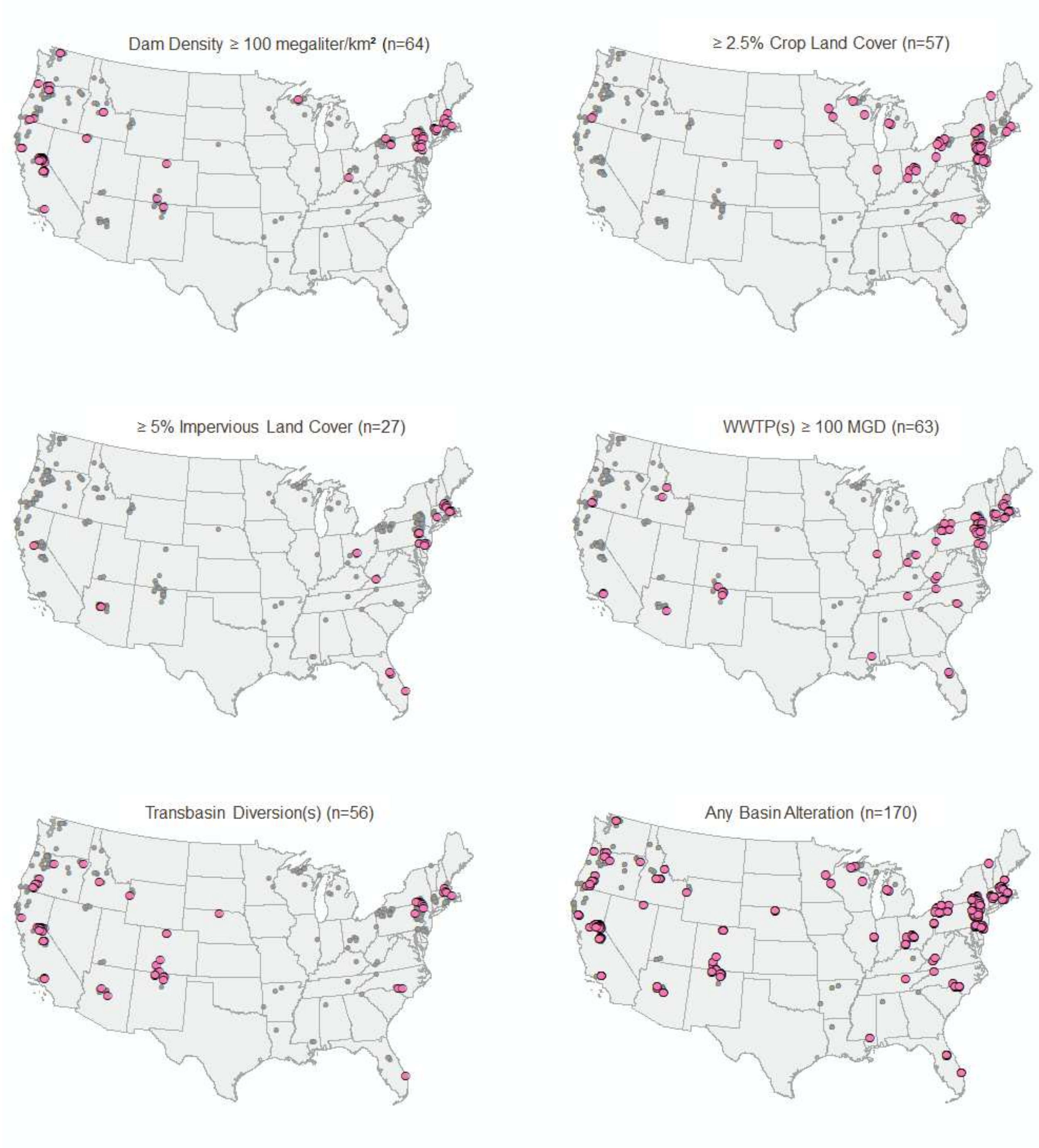

Figure 6. Wild and scenic river stream gages with and without watershed alterations. Pink circles represent gages whose watersheds have dam storage densities ≥ 100 megaliters·km^{-2}, ≥ 2.5% crop land cover, ≥ 5% impervious land cover, the presence of a wastewater treatment plant (WWTP) that discharges ≥ 100 MGD, or the presence of a transbasin diversion. Grey circles represent gages whose watersheds did not meet these thresholds for alteration. The final map displays gages that have any of these watershed alterations.

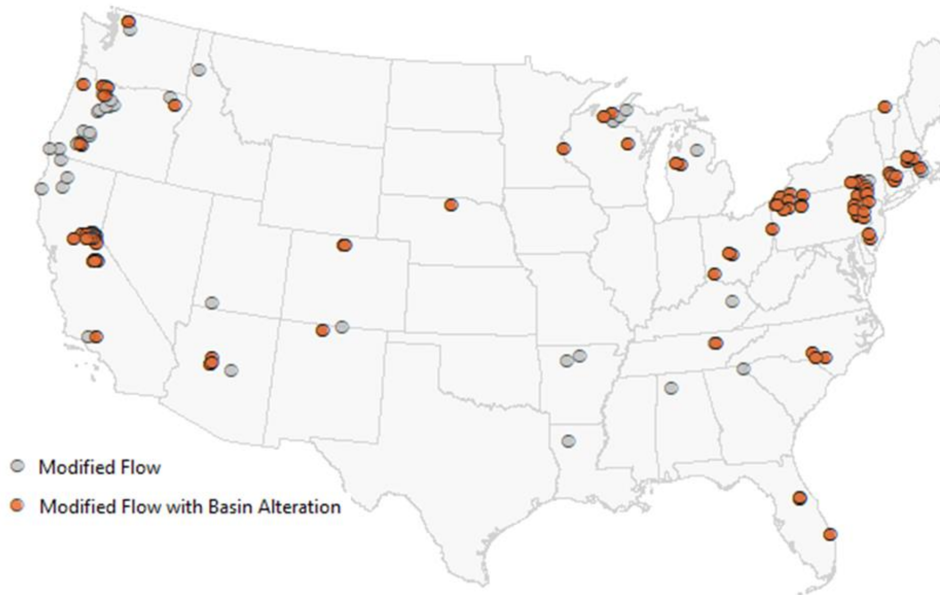

Figure 7. Gages in wild and scenic river watersheds with modified streamflow. Orange circles represent gages with modified streamflow as well as a known watershed alteration (n=97). Grey circles represent gages with modified flow without a known watershed alteration (n=49).

4.1 Seasonal Distribution of Flows

Mean annual flow and seasonal distributions of flow were some of the metrics best represented by the natural flow models (Figure 8). Wild and scenic river gages with Q_{mean} values outside the range of predicted values tended to have lower flow than expected and were found in the E-S-S, E-S-NS, and W-D-S regions. Of these watersheds, 83.3% had at least one type of basin alteration, with half having a transbasin diversion present, a dam storage density ≥ 100 megaliters·km^{-2}, or both. In contrast, Q_{mean} values tended to be higher for modified wild and scenic river gages in the E-D-S region. Of these seven watersheds, all but one had some type of hydrologic modification, with four of them having over 5% mean impervious land cover. Spring flows were modified across all regions, with most altered watersheds having decreased spring flow. Of these gages, over 90% had some type of hydrologic modification, with dam storage densities ≥ 100 megaliters·km^{-2} and the presence of wastewater treatment plants being the most common (46.9% and 37.9%, respectively). Summer flows that were greater than predicted were found in the Northeast and Florida, whereas summer flows in the West and Southeast regions were more often lower than predicted. All gages with increased summer flow had some sort of hydrologic modification, with the three most common sources being dam storage densities ≥ 100 megaliters·km^{-2} (66.7%), major wastewater treatment plant (53.3%), and the presence of a transbasin diversion (40.0%). Although 54.5% of gages with decreased summer flow had some sort of modification, no dominant watershed alteration type stood out. Fewer gages had sufficient models for assessing autumn flows, but the majority of watersheds with modified autumn flow had higher than expected values. Of these, 76.2% had some sort of basin modification, with dam storage densities ≥ 100 megaliters·km^{-2} the most common. Lastly, winter flows were the least modified of the flow metrics (6.8%), though this flow metric also had comparatively fewer gages with sufficient models to assess. Of the few gages that were modified, 41.7% of them had a transbasin diversion present (Figure 6).

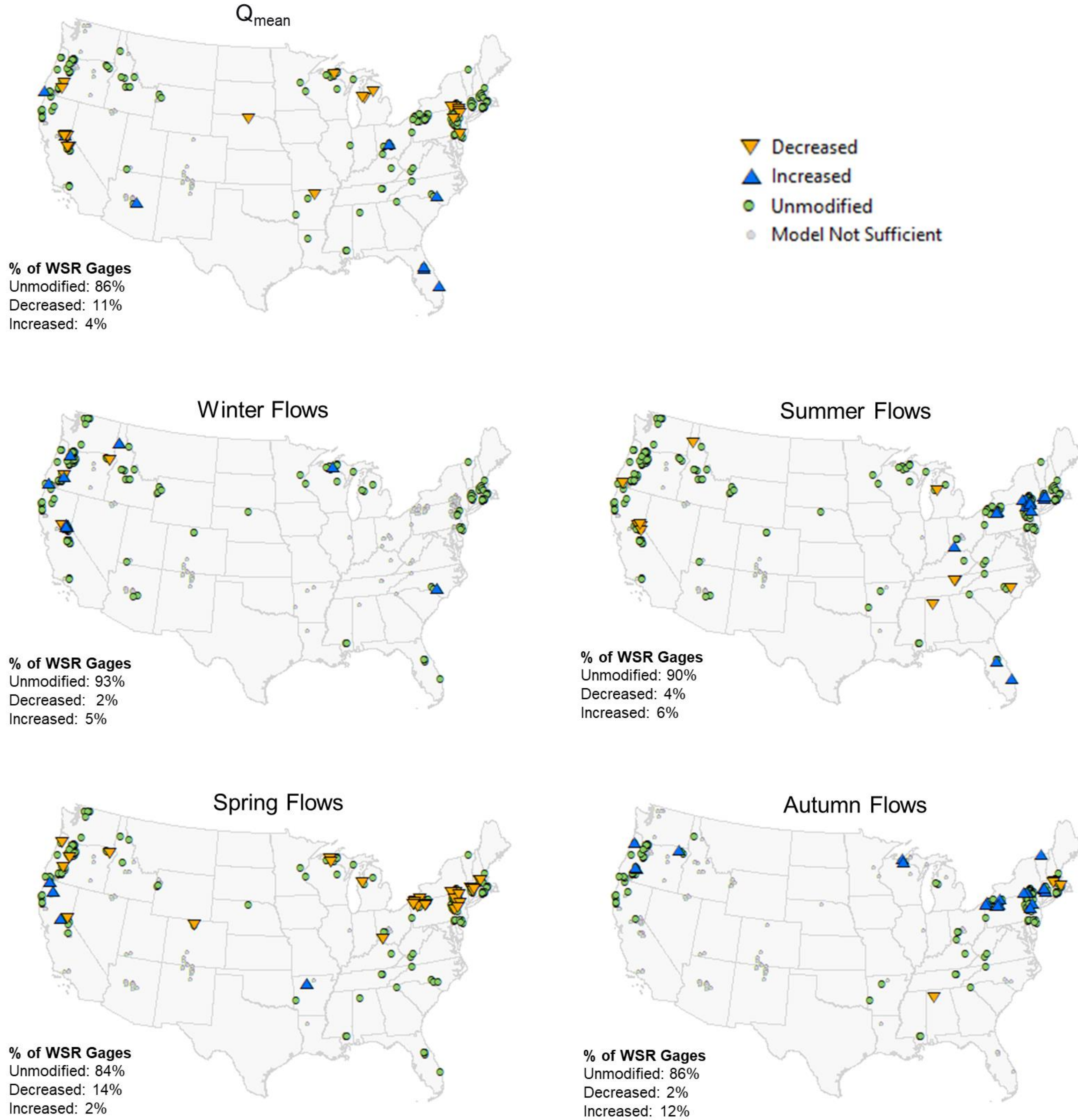

Figure 8. Wild and scenic river (WSR) watersheds with and without modified flow metrics for mean annual flow and the fraction of flow in each season. Blue upward-facing triangles represent stream gages whose observed value was greater than the predicted value. Orange downward-facing triangles represent stream gages whose observed value was less than the predicted value. Green circles represent gages whose observed value fell within the 95% prediction interval of the prediction.

4.2 High and Low Flows

The 99th percentile non-exceedance flow (Q99) had the greatest number of gages with sufficient models when compared to all other flow metrics. Gages with Q99 values outside the predicted range tended to be lower than predicted, with only nine gages having an increased Q99. Of those with a decreased Q99, dam storage densities ≥ 100 megaliters·km^{-2}, transbasin diversions, and cropland covers ≥ 2.5% were the most prevalent types of basin modification (47.5%, 35.0%, and 30%, respectively). Compared to the Q99, there were fewer gages with sufficient 1st percentile non-exceedance flow (Q1) models, with similar proportions of gages with increased and decreased Q1. Of those gages with an increased Q1, 89.5% had some sort of hydrologic modification, with the presence of major wastewater treatment facilities being the most prevalent (47.4%). Dam storage densities ≥ 100 megaliters·km^{-2}, ≥5% impervious cover, and ≥2.5% cropland cover were also common. For gages with decreased Q1, patterns in modification were less clear, with only five of the fifteen gages exhibiting any sort of watershed modification.

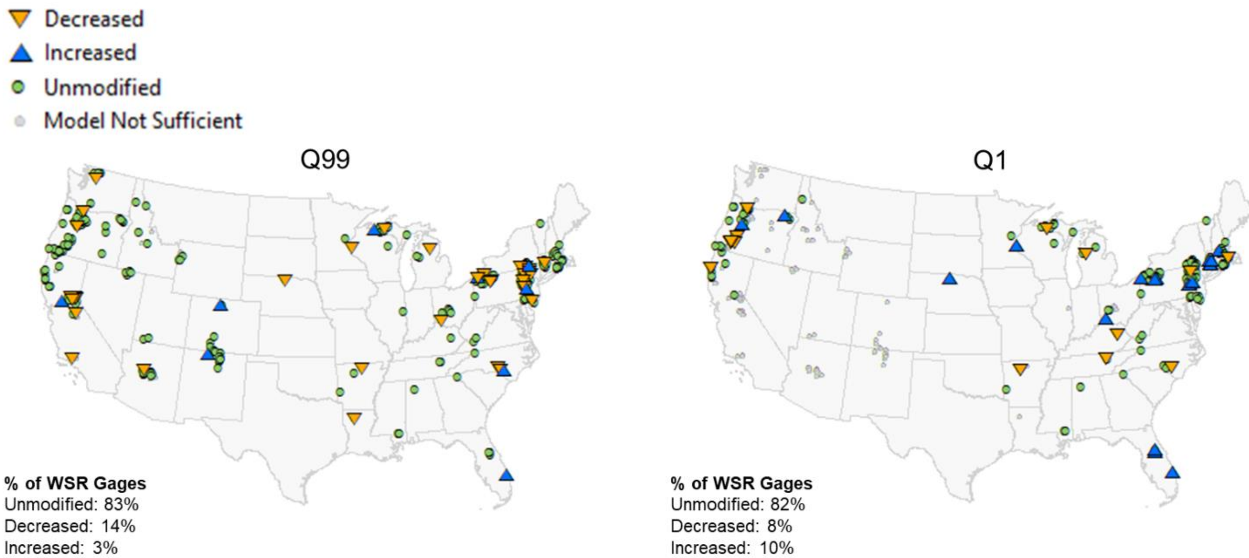

Figure 9. Wild and scenic river (WSR) watersheds with and without modified flow metrics for Q99 and Q1. Blue upward-facing triangles represent stream gages whose observed value was greater than the predicted value. Orange downward-facing triangles represent stream gages whose observed value was less than the predicted value. Green circles represent gages whose observed value fell within the 95% prediction interval of the prediction.

4.3 Variability, Frequency, and Duration of Flows

Models for the variability of low flows (CV_{low}) and the duration of the high and low flows (D_{high}, D_{low}) did not meet the performance criteria for any wild and scenic river gages. The variability of high flows (CV_{high}) had a high proportion of modified wild and scenic river gages, but it also had the second lowest number of gages assessed (n=106). CV_{high} values tended to be elevated and lower in equal proportions. There were no clear patterns in basin modification types for gages with diminished high flow variability, though 50.0% did have some sort of hydrologic modifier. On the other hand, 81.2% of gages with higher CV_{high} values had a watershed modification, with transbasin diversions and dam storage densities ≥ 100 megaliters·km^{-2} being the most common. Like the variability of high flows, the high and low flow frequencies (F_{high} and F_{low}) were analyzed at fewer wild and scenic river gages compared to other flow metrics. F_{high} values were higher than expected at most gages with modified flow. Roughly a third (28.6%) had

≥ 5% mean impervious land cover. F_{low} was the flow metric with the least number of gages analyzed (n=56). Of the gages assessed, F_{low} tended to be higher than expected, and 81.8% of the gages with higher F_{low} had some sort of basin modification; over half (54.5%) had dam storage densities ≥ 100 megaliters·km^{-2}.

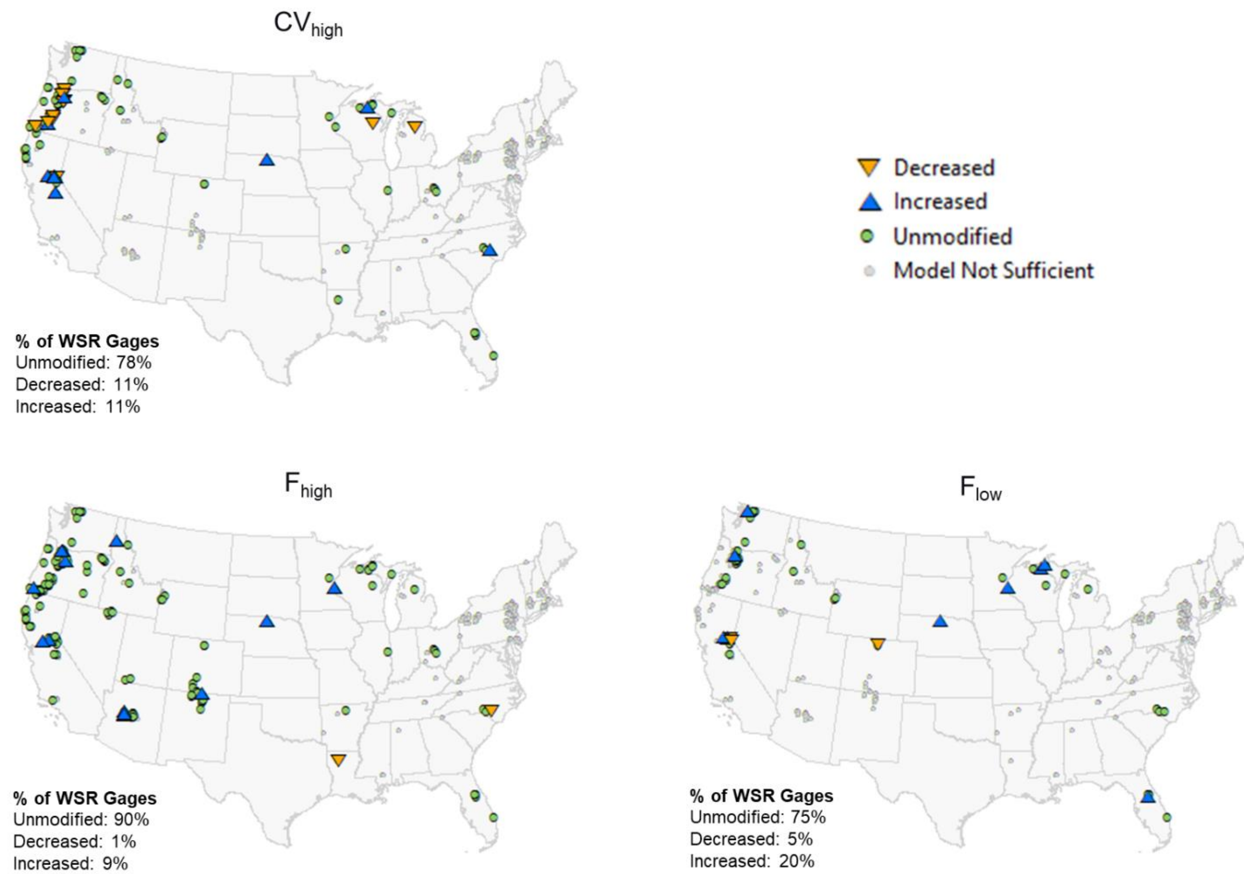

Figure 10. Wild and scenic river (WSR) watersheds with and without flow metrics associated with the variability, frequency, and duration of flows. Blue upward-facing triangles represent stream gages whose observed value was greater than the predicted value. Orange downward-facing triangles represent stream gages whose observed value was less than the predicted value. Green circles indicate gages whose observed value fell within the 95% prediction interval of the prediction.

5. DISCUSSION

5.1 Comparison to Other Stream Gage Assessments

Compared to the most analogous national-level flow assessment of the US (i.e., Eng et al., 2019), we found a greater fraction of gages with unaltered flow for gages both within and outside wild and scenic river watersheds. There are several explanations for this; in our analysis, not all flow metrics were assessed at each stream gage because not all of our models were of sufficient quality. Had all flow metrics been assessed at each stream gage, it is likely that it would have resulted in more gages with altered flow metrics. Second, our fraction of unmodified stream gages includes stream gages that were used in the development of the models. Eng et al. (2019)'s conclusion that 80% of assessed stream gages had altered flow does not include reference-quality gages in the assessment; removing reference gages from the proportion of stream gages assessed inherently raises the fraction of modified sites. For example, if we were to remove reference gages from our analysis, the percentage of modified gages within wild and scenic river watersheds would increase from 45.1% to 54.9%, and the fraction of modified gages outside wild and scenic river watersheds would increase from 55.6% to 62.9%. Third, our cutoff for defining unmodified flow was more generous than that of Eng et al. (2019). While Eng et al. (2019) deemed a flow status "indeterminant" if the O/P value was equal to or less than the model's local prediction error, we deemed a flow status to be within the range of natural variability if the prediction's 95% prediction interval contained the observed value. Eng et al. (2019) also developed flow models with a random forest algorithm, which is not well suited for extrapolation (Hengl et al., 2018). Specifically, if predictions were made at stream gages whose observed flow metric value was beyond the range of values used in the training dataset, it would automatically be considered modified even though this may not be the case. Lastly, Eng et al. (2019)'s training dataset used several stream gages whose watersheds had transbasin diversions and/or ≥ 2.5% cropland cover, which we identified as having the potential to modify streamflow from its "*natural condition*" and could therefore introduce model defects.

5.2 Wild and Scenic River Modifications and the "Free-Flowing" Paradox

We found that a smaller fraction of wild and scenic river watersheds have modified flow when compared to streams outside wild and scenic river watersheds. Our other finding that nearly one third (30.3%) of gages within wild and scenic river watersheds were of reference quality supports the notion that wild and scenic river watersheds have less hydrologic alteration than non-wild and scenic river watersheds. In part this is because of how and where wild and scenic rivers have historically been selected for designation. Concerns about wild and scenic river designations threatening economic growth and increasing land use restrictions, particularly on private property[2], have led to many wild and scenic rivers being in areas with less urbanization and/or private lands (Perry, 2017). The requirement that a wild and scenic river be free of "*impoundments, diversions, straightening, rip-rapping, and any other modifications of the waterway*" also limits where designations are viable. Our results demonstrate that wild and scenic river watersheds are less developed, as they have a significantly lower ($p < 0.05$) proportion of gages with ≥ 5% mean impervious cover, ≥ 2.5% cropland cover, or a major wastewater treatment plant compared to watersheds that are not associated with a wild and scenic river. Moreover, wild and scenic rivers are often in protected areas. 117 of the 196 wild and

---

[2] These concerns are misguided since land use regulations do not fall within the authority of the Wild and Scenic Rivers Act.

scenic rivers in our study are either located within a wilderness area or have a wilderness area in their watershed, whereas only 60 of the 196 wild and scenic river watersheds in our study contain an urban area (as delineated by the US Census Bureau, 2019). Lands protected for conservation purposes (i.e., GAP Statuses 1 and 2) make up 19.9% of the total wild and scenic river watershed area in our study. Because many wild and scenic rivers are located within or near protected lands, it is likely that their watersheds are more insulated from hydrologic modifications stemming from urban or agricultural development.

Yet, half of all wild and scenic river gages still have watershed alterations, with many of those gages also experiencing altered flow. This is seemingly at odds with the Wild and Scenic Rivers Act's goal of "*protecting and enhancing*" a free-flowing river's "*natural condition*". However, the consensus that a river's "*natural condition*" relies on the upstream sources that feed it and that a natural flow regime is critical in preserving a river's functionality had not been fully established until after the Wild and Scenic Rivers Act was ratified (Palmer, 2017). It wasn't until Bovee's Instream Flow Incremental Methodology (1982) that river managers began quantifying the link between streamflow and a river's capacity to support ecosystems, while the natural flow regime was only defined in 1997 (Poff et al., 1997). Therefore, the Wild and Scenic Rivers Act was not originally designed to account for this paradigm of evaluating a river's "*natural condition*" and instead focused on the discrete reaches of interest.

In response to the new perspective on "*free-flowing*" rivers, the Wild and Scenic Rivers Act has been used to designate tributaries to already-designated rivers (e.g., the Rogue Wild and Scenic River) and even entire watersheds (e.g., the Smith Wild and Scenic River, White Clay Creek Wild and Scenic River) that meet the requirements for inclusion into the system. Tangentially related to this concept is the recent suggestion by Perry (2017) that "*outstandingly* *remarkable values*” could be framed as ecosystem services; by recognizing a tributary’s role in supporting the ecosystem services of a wild and scenic river, it could better fulfill the *“outstandingly remarkable*” requirements for designation. Still, watershed-scale designations are less common.

Perhaps the most surprising finding from our analysis was that wild and scenic river watersheds had a significantly greater ($p < 0.05$) fraction of gages with dam storage densities $\geq$ 100 megaliters$\cdot$km^{-2} when compared to gages outside wild and scenic river watersheds. Even though dams are discouraged within designated wild and scenic river reaches, dams were the most common type of watershed alteration for wild and scenic river stream gages with modified flow. Specifically, high dam storage density was a top basin modifier for gages with decreased spring flows, increased summer and fall flows, decreased Q99, increased Q1, and increased frequency of low flows. This supports previous studies that found that dams homogenize flow regimes and cause more frequent low flow events (Poff et al., 2007; Carlisle et al., 2019; Eng et al., 2019). Dams used for water supply often refill during spring runoff, accounting for decreased spring flows downstream, then release water for use during the summer irrigation season. This release of water during low flow seasons can cause low flow magnitudes (Q1) to increase. Dams are also used for flood control, which can reduce downstream peak flows (Q99).

The fact that wild and scenic river watersheds have higher dam storage densities may seem contradictory given the Wild and Scenic Rivers Act’s clear objective being to support a river’s “*free-flowing*” character. In many ways, dams are the ultimate symbol of hydrologic alteration. Dams are largely responsible for the disproportionate decline of freshwater biodiversity, fragmentation of stream corridors, simplified geomorphologies and degraded floodplains, trapped sediment and debris, and recreational safety issues (Ricciardi and

Rasmussen, 2001; Marks et al., 2006; Graff, 2006; Schook et al., 2016; Ritchie et al., 2018). With some historical context, however, it is not so surprising. The Wild and Scenic Rivers Act was created at the zenith of dam construction in the US (Figure 11) as a tool to mitigate their spread: *"Congress declares that the established national policy of dams and other construction at appropriate sections of the rivers of the United States needs to be complemented by a policy that would preserve other selected rivers or sections thereof…"* (16 U.S. Code 1271-1278). In fact, many (if not most) wild and scenic rivers were designated in direct response to the threat of damming (e.g., the Wildcat Wild and Scenic River, the Merced Wild and Scenic River, the Middle Delaware Wild and Scenic River), or as a complement to damming occurring nearby (e.g., the Missouri Wild and Scenic River, the Rogue Wild and Scenic River). Early wild and scenic river ideologies also centered largely on the idea of protecting what was deemed beautiful, as exemplified by President Lyndon B. Johnson's goals of "beautifying" America (Johnson, 1965). This focus resulted in designating vast, grand riverscapes outside the radii of urbanization as opposed to more modest rivers with urban or agricultural influence (Palmer, 2004; Palmer, 2017; Perry, 2017). Though these dramatic rivers may be isolated from agriculture and urbanization, they are paradoxically the ideal setting for large dams (i.e., vast, sparsely colonized, steep canyon walls, etc.).

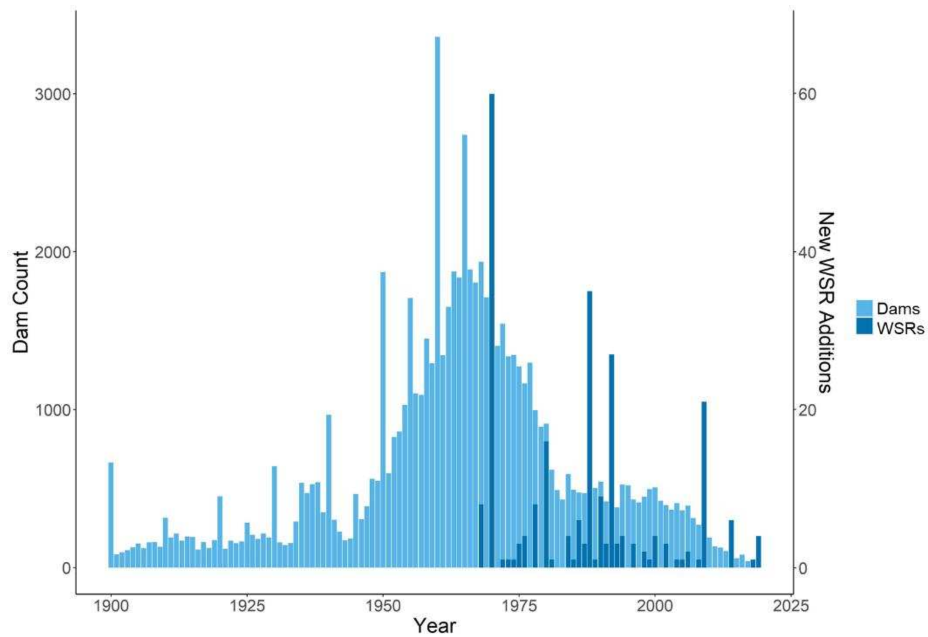

Figure 11. Number of dams constructed and the number of additions to the Wild and Scenic River (WSR) System through time. The WSR Act was created in part to mitigate the effects of increased dam construction in the US. Dam data was pulled from the National Inventory of Dams (USACE, 2019).

5.3 Partnership Wild and Scenic Rivers

Wild and scenic rivers of the Northeast (and the Wekiva Wild and Scenic River in Florida) are managed under a different framework than most other wild and scenic rivers, and in many ways, deviate from our preceding discussion. Partnership wild and scenic rivers are rivers that are managed in cooperation with the National Park Service but are predominantly cared for by local community groups. Officially developed in the 1990s, this wild and scenic river framework provided an opportunity for rivers to be designated that had previously been excluded: rivers that flowed within private lands, rivers whose watersheds supported dense populations, and rivers with urbanization along them (Fosburgh, DiBello, and Akers, 2008). Because partnership wild and scenic rivers are not typically associated with the remote settings of their traditional counterparts, they are faced with a more complex suite of watershed alterations (Figure 6). Alterations are also found at higher rates; 76.2% of gages within partnership wild and scenic river watersheds had some sort of watershed alteration, compared to 46.9% of gages in non-partnership wild and scenic river watersheds. However, by default partnership wild and scenic rivers have more direct community engagement than other wild and scenic rivers because a local management group is required for their existence. Though not evaluated at the system-wide scale, the US Forest Service (personal communication, February 2, 2021) reported that in most years, less than half of their wild and scenic rivers (i.e., no partnership rivers) receive assistance from stewardship groups. This suggests that traditional wild and scenic rivers have less community support than partnership rivers. So although partnership wild and scenic rivers experience higher rates of modification, they are fundamentally linked to community involvement that helps *"protect and enhance*" a wild and scenic river's "*natural condition*". In fact, many partnership rivers have used their wild and scenic river designation as leverage to make quantifiable improvements to their watersheds. Examples of such improvements include dam removals, riparian habitat restoration, implementing community-wide best management practices, and wastewater retrofits (National Park Service, 2016).

5. 4 Caveats to our Evaluation

Although this is the most comprehensive assessment of wild and scenic river flow conducted, it does not represent the full wild and scenic river system due to limitations in our methodology. Evaluating the status of gaged streams introduces biases associated with where stream gages are located. There are generally more gages located in the East than in the West, and within the West there are even fewer gages located in arid regions (Kiang et al., 2013). Moreover, stream gages are biased toward larger watersheds, at lower elevations, and along streams with confined control sections and limited ranges of flow (DeWeber et al., 2014; Kampf et al., 2020). By reducing our dataset to only stream gages whose watersheds are < 1,500 km^2

with a long period of record, we also limited our results to a fraction of the available gages deployed by the USGS. Beyond the general biases associated with stream gage locations, more than half (54.1%) of the wild and scenic rivers we sought to assess did not have a viable stream gage. We also did not evaluate any data associated with wild and scenic rivers within Alaska, Puerto Rico, and wild and scenic rivers whose watersheds exceed 100,000 km^2.

By defining watershed modification as one of only five potential basin characteristics (i.e., ≥ 2.5% cropland cover, ≥ 5% mean imperviousness, presence of a major wastewater treatment facility, a dam storage density ≥ 100 megaliters·km^{-2}, or the presence of a transbasin diversion), it is likely we are excluding other watershed features that can explain changes is streamflow. For example, watershed disturbances including timber harvesting, mining, forest fires, and bark beetle outbreaks could explain why many stream gages in the West without basin modifications still have altered streamflow (Clark et al, 2014; Nippgen et al., 2017; Hallema et al., 2018; Eurich, 2020). Moreover, the accumulation of small and often unreported stream barriers and diversions have been shown to significantly impact streamflow (Deitch, Kondolf, and Merenlender, 2009; Baker et al., 2011; Belletti et al., 2020). Though some of these characteristics are currently difficult to quantify across the country, work to address these data gaps is ongoing and could be used in future analyses (Maus et al., 2020; National Interagency Fire Center, 2020; Hicke et al., 2020; Whittemore et al., 2020; US Forest Service, 2021).

Climate change may also shed light on deviating streamflow patterns. Though Eng et al. (2019) concluded that climate variability did not frequently explain significant changes to streamflow, it is possible that these influences are obscured in the analyses of large datasets, and even large watersheds. The larger the watershed, the greater the heterogeneity of watershed characteristics within the area; this can reduce the connection between streamflow and watershed average features like climate (Hammond and Kampf, 2020). When measuring the effects of climate change at smaller scales, and particularly in snow-dominated hydrological settings, many studies have found that climate change has directly altered the variability, duration, and timing of streamflow (e.g. Kampf and Lefsky, 2016; Giles-Hansen et al., 2019). Therefore we believe a temporal assessment at the river-by-river scale could provide better insight into streamflow changes caused by climate change.

6. CONCLUSION

Even considering our study's limitations, it is clear that wild and scenic river watersheds are susceptible to conditions that degrade their "*natural condition",* and subsequently their "*free-flowing"* character. We found that nearly half (45.1%) of all wild and scenic river stream gages had at least one altered flow metric, and that they had significantly higher ($p < 0.05$) rates of dam storage densities $\geq$ 100 megaliters$\cdot$km^{-2} than gages outside of wild and scenic river watersheds. This finding may reflect the prevalence of wild and scenic river designations in stream systems well-suited for dam development.

A wild and scenic river designation is meant to "*protect and enhance*" the "*natural condition*" of rivers, but the Wild and Scenic Rivers Act was not designed with the intent to protect a river's flow regime, and instead focuses on the protection of select river reaches mostly regardless of their upstream status. However, the "*natural condition*" of a wild and scenic river cannot be determined one reach at a time, and is instead a function of many combined influences from upstream watersheds. Although the Wild and Scenic Rivers Act as written cannot fully regulate the degree of hydrologic alteration outside the designated reach, community organizations and managing agencies can identify where alterations pose a threat to the river-dependent values that got a river designated, and can seek strategies to make improvements to them. In fact, many wild and scenic rivers have used their designation to accomplish such improvements, while a recent re-envisioning of "*outstandingly remarkable values"* as ecosystem services could open up more tributaries to designation (National Park Service, 2016; Perry, 2017). Looking ahead, efforts to recover the free-flowing condition of wild and scenic river watersheds will be critical in protecting and enhancing the values that make these rivers so outstanding.

BIBLIOGRAPHY

Allen DM (1971). The prediction sum of squares as a criterion for selecting predictor variables. Technical Report 23, University of Kentucky, Department of Statistics.

Baker DW, Bledsoe BP, Albano SM, and NL Poff NL (2011). Downstream effects of diversion dams on sediment and hydraulic conditions of Rocky Mountain Streams. River Research and Applications, 27: 388-401.

Bartón K (2020). MuMIn: Multi-Model Inference. R package version 1.43.17.U https://CRAN.R-project.org/package=MuMIn.

Belletti B, Garcia de Leaniz C, J Jones, Bizzi S, Börger B, Segura G, Castelletti A, van de Bund A, Aarestrup K, Barry J, Belka K, Berkhuysen A, Birnie-Gauvin K, Bussettini M, Carolii M, Consuegra S, Dopico E, Feierfeil T, Fernández S, Garrido PF, Garcia-Vazquez E, Garrido S, Giannico G, Gough P, Jepsen N, Jones PE, Kemp P, Kerr J, King J, Łapińska M, Lázaro P, Lucas MC, Marcello L, Martin P, McGinnity P, O'Hanley J, Olivo del Amo R, Parasiewicz P, Pusch M, Rincon G, Rodriguez C, Royte J, Schneider CL, Tummers JS, Vallesi S, Vowles A, Verspoor E, Wanningen H, Wantzen KM, Wildman L, and M Zalewski (2020). More than one million barriers fragment Europe's rivers. Nature, 588: 436-441.

Bhaskar AS and C Welty (2012). Water balances along an urban-to-rural gradient of metropolitan Baltimore, 2001-2009. Environmental & Engineering Geoscience, 18(1): 37-50.

Blodgett D (2019). nhdplusTools: Tools for accessing and working with the NHDPlus, R package version 0.3.8. https://code.usgs.gov/water/nhdplusTools.

Bovee KD (1982). A Guide to stream habitat analysis using the instream flow incremental methodology: US Fish and Wildlife Service Report FWS/BS-82/26, 273 pp.

Brown TC, Taylor JG, and B Shelby (1992). Assessing the direct effects of streamflow on recreation: A literature review. Water Resources Bulletin of the American Water Resources Association, 27(6): 979-989.

Carlisle DM, Falcone J, Wolock DM, Meador MR, and RH Norris (2010). Predicting the natural flow regime: Models for assessing hydrological alteration in streams. River Research and Applications, 26: 118-136.

Carlisle DM, Wolock DM, Konrad CP, McCabe GJ, Eng K, Grantham TE, and B Mahler (2019). Flow modification in the Nation's streams and rivers: US Geological Survey Circular 1461: 75 pp.

De Cicco LA, Hirsch RM, Lorenz D, and WD Watkins (2018). dataRetrieval: R packages for discovering and retrieving water data available from the federal hydrologic web services, doi:10.5066/P9X4L3GE.

Deitch MJ, Kondolf GM, and AM Merenlender (2009). Hydrologic impacts of small-scale instream diversions for frost and heat protection in the California wine country. River Research and Applications, 25: 118-134.

DeWeber JT, Tsang YP, Krueger DM, Whittier JB, Wagner T, Infante DM, and G Whelan (2014). Importance of understanding landscape biases in USGS gage locations: Implications and solutions for managers. Fisheries, 39(4): 155-163.

Elmore A, Fries A, Nastase E, Kelsey H, Wilson M, Steele NLC, Schonek K, Pao S, and J Ford (2020). Verde River Watershed Report Card, 40 pp.

Eng K, Carlisle DM, Grantham TE, Wolock DM, and RL Eng (2019). Severity and extent of alterations to natural streamflow regimes based on hydrologic metrics in the conterminous United States, 1980-2014: US Geological Survey Scientific Investigations Report 2019-5001, 25 pp.

Environmental Systems Research Institute (ESRI) (2017). *ArcGIS Release 10.6.1.* Redlands, CA.

Eurich A: Effects of flow modification and forest disturbance on streamflow across Colorado, MS Thesis, Colorado State University, 2020.

Falcone JA (2011). GAGES-II: Geospatial attributes of gages for evaluating streamflow. US Geological Survey database, https://doi.org/10.3133/70046617.

Fantin-Cruz I, Pedrollo O, Girard P, Zeilhofer P, and SK Hamilton (2016). Changes is river water quality caused by a diversion hydropower dam bordering the Pantanal floodplain. Hydrobiologia, 768: 223-238.

Fosburgh J, DiBello J, and F Akers (2008). Partnership wild and scenic rivers. The George Wright Forum, 25(2): 37-44.

Gesch DB, Oimoen MJ, Greenlee SK, Nelson CA, Steuck M, and DJ Tyler (2002). The national elevation dataset. Photogrammetric Engineering and Remote Sensing, 68(1): 5-11.

Giles-Hansen K, Li Q, and X Wei (2019). Cumulative effects of forest disturbance and climate variability on streamflow in Deadman River watershed. Forests, 10(2): 9-11.

Gotvald AJ (2017). Methods for estimating selected low-flow frequency statistics and mean annual flow for ungagged locations on streams in North Georgia: US Geological Survey Scientific Investigations Report 2017-5001, 36 pp.

Graff WL (2006). Downstream hydrologic and geomorphic effects of large dams on American rivers. Geomorphology, 79(2006): 336-360.

Grill G, Lehner B, Thieme M, Geenen B, Tickner D, Antonelli F, Babu S, Borrelli P, Cheng L, Crochetiere H, Ehalt Macedo J, Filgueiras R, Goichot M, Higgins J, Hogan Z, Lip B, McCain ME, Meng J, Mulligan M, Nilsson C, Olden JD, Opperman JJ, Petry P, Reidy Liermann C, Sáenz L, Salinas-Rodríguez S, Schelle P, Schmitt RJP, Snider J, Tan F, Tockner K, Valdujo PH, can Soesbergen A, and C Zarfl (2019). Mapping the world's free-flowing rivers. Nature, 569: 215-221.

Hallema DW, Sun G, Caldwell PV, Norman SP, Cohon EC, Liu Y, Bladon KD, and SG McNulty (2018). Burned forests impact water supplies. Nature Communications, 9(1): 1-8.

Hammond JC, Saavedra FA, and SK Kampf (2020). MODIS MOD10A2 derived snow persistence and no data index for the conterminous US. Unpublished data.

Hammond JC and SK Kampf (2020). Sub-annual streamflow responses to rainfall and snowmelt inputs in snow-dominated watersheds of the western US. Water Resources Research, 56.

Harrison-Atlas D, Theobald DM, Dickson BG, Landau V, and I Leinwand (2017). Description of the approach, data, and analytical methods used to evaluate river systems in the western US. Conservation Science Partners, Truckee, CA.

Hart DD and CM Finelli (1999). Physical-biological coupling in streams: the pervasive effects of flow on benthic organisms. Annual Review of Ecology and Systematics, 30: 363-395.

Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, and B Graler (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6:e5518.

Hicke JE, Xu B, Meddens AJH, JM Egans (2020). Characterizing recent bark beetle-caused tree mortality in the western United States from aerial surveys. Forest Ecology and Management, 475: 118402.

Hortness JE and C Berenbrock. (2001). Estimating monthly and annual streamflow statistics at ungauged sites in Idaho: US Geological Survey Water-Resources Investigations Report 01-4093: 40 pp.

Johnson LB (1965). Special Message to the Congress on Conservation and Restoration of Natural Beauty, http://www.lbjlibrary.net/collections/selected-speeches/1965/02-08-1965.html

Junk W, Bayley PB, and RE Sparks (1989). The flood pulse concept in river-floodplain systems. Canadian Journal of Fisheries and Aquatic Sciences, 106: 110-127.

Kagawa H (1992). Effects of diversion on the chemistry of a stream in Japan. Regulated Rivers: Research & Management, 7(3): 291-302.

Kampf SK, Burges SJ, Hammond JC, Bhaskar A, Covino TP, Eurich A, Harrison H, Lefsky M, Martin C, McGrath D, Puntenney-Desmond K, and K Willi (2020). The case for an open water balance: Re-envisioning network design and data analysis for a complex, uncertain work. Water Resources Research, 56.

Kampf SK, and MA Lefsky (2016). Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range Floods. Water Resources Research, 52: 407-422.

Kiang JE, Stewart DW, Archfield SA, Osborne EB, and K Eng (2013). A national streamflow network gap analysis: US Geological Survey Investigations Report 2013-5013, 79 pp.

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, and R Core Team (2020). caret: Classification and Regression Training. R package version 6.0-86. https://cran.r-project.org/package=caret.

Marks JC, Parnell R, Carter C, Dinger EC, and GA Haden (2006). Interactions between geomorphology and ecosystem processes in travertine streams: Implications for decommissioning a dam on Fossil Creek, Arizona. Geomorphology, 77(2006): 299-307.

Maus V, Giljum S, Gutschlhofer J, da Silva DM, Probst M, Gass SLB, Luckeneder S, Lieber M, and I McCallum (2020). A global-scale data set of mining areas. Scientific Data: 7:289

McCabe GJ and DM Wolock (2011). Independent effects of temperature and precipitation on modeled runoff in the conterminous United States. Water Resources Research, 47(11): 1-11.

McKay L, Bondelid T, Dewald T, Johnston J, Moore R, and A Rhea (2012). NHDPlus Version 2: User Guide, 182 pp.

Moore RB, McKay LD, Rea AH, Bondelid TR, Price CV, Dewald TG, and CM Johnston (2019). User's guide for the national hydrography dataset plus (NHDPlus) high resolution: US Geological Survey Scientific Investigations Report 2019-1096, 80 pp.

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, and TL Veith (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the American Society of Agricultural and Biological Engineers, 50(3): 885-900.

Motion 009 – Protecting rivers and their associated ecosystems as corridors in a changing climate, International Union for the Conservation of Nature (IUCN) World Conservation Congress (2020), https://www.iucncongress2020.org/motion/009.

Nadeau AJ, Allen K, Hutchins H, and A Robertson (2018). Natural resource condition assessment: Bluestone National Scenic River. Natural Resource Report NPS/BLUE/NRR-2018/1810. National Park Service, Fort Collins, Colorado.

Narvaez M and A Homsey (2016). White Clay Creek state of the watershed report: an update on the health of the White Clay Creek Wild and Scenic Watershed in Delaware and Pennsylvania. University of Delaware, Newark DE.

National Interagency Fire Center (2020). Interagency Fire Perimeter History - All Years, https://data-nifc.opendata.arcgis.com/datasets/interagency-fire-perimeter-history-all-years

National Park Service (2016). Partnership wild & scenic rivers: 20 years of success protecting nationally significant river resources through locally based partnerships, http://npshistory.com/publications/nwsr/20-yr-success.pdf.

Nippgen F, Ross MRV, Bernhardt ES, and BL McGlynn (2017). Creating a more perennial problem? Mountaintop removal coal mining enhances and sustains saline baseflows of Appalachian watersheds. Environmental Science and Technology, 51: 8324-8334.

OARS for the Assabet, Sudbury, and Concord Rivers (2019). Sudbury, Assabet, and Concord River Report Card, https://ecoreportcard.org/report-cards/sudbury-assabet-concord-rivers/.

Palmer T (2004). *Endangered rivers and the conservation movement.* Rowman & Littlefield Publishers.

Palmer T (2017). *Wild and scenic rivers: An American legacy.* Oregon Press.

Perry, D (2017). Reframing the Wild and Scenic Rivers Act: Ecosystem-based resilience and adaptation. International Journal of Wilderness, 23(2): 41-48.

Poff NL, Allan D, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, and JC Stromberg (1997). The natural flow regime: A paradigm for river conservation and restoration. Bioscience, 47: 769-784.

Poff NL, Olden JD, Merritt DM, and DM Pepin (2007). Homogenization of regional river dynamics by dams and global biodiversity implications. Proceedings of the National Academy of Sciences of the United States of America, 104(14): 5732-5737.

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Bledsoe BP, Freeman MC, Henriksen J, Jacobson RB, Kennen JG, Merritt DM, O'Keeffe JH, Olden JD, Rogers K, Tharme RE, and A Warner (2010). The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshwater Biology, 55: 147-170.

Power ME, Sun A, Parker G, Dietrich WE, and JT Wootton (1995). Hydraulic food-chain models. BioScience, 45(3): 159-167.

Reisner M (1993). *Cadillac desert: the American West and its disappearing water.* Penguin Books.

Ricciardi A and JB Rasmussen (2001). Extinction rates of North American freshwater fauna. Conservation Biology, 13(5): 1220-1222.

Ritchie AC, Warrick JA, East AE, Magirl CS, Stevens AW, Bountry JA, Randle TJ, Curran CA, Hilldale RC, Duda JJ, Gelfenbaum GR, Miller IM, Pess GR, Foley MM, McCoy R, and AS Ogston (2018). Morphodynamic evolution following sediment release from the world's largest dam removal. Scientific Reports, 8(13279): 1-13.

Sanborn SC and BP Bledsoe (2006). Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. Journal of Hydrology, 325(1-4): 241-261.

Schook DM, Carlson EA, Sholtes JS, and DJ Cooper (2016). Effects of moderate and extreme flow regulation on *Populus* growth along the Green and Yampa Rivers, Colorado and Utah. River Research and Applications, 32: 1698-1708.

Snelder TH, Lamouroux N, Leathwick JR, Pella H, Sauquet Em and U Shankar (2009). Predictive mapping of the natural flow regimes of France. Journal of Hydrology, 373(1-2): 57-67.

Soil Survey Staff, Natural Resources Conservation Service, US Department of Agriculture (NRCS USDA). Web Soil Survey database, https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627

The Nature Conservancy (2009). Indicators of Hydrologic Modification Version 7.1 User's Manual.

Thornbrugh DJ, Leibowitz SG, Hill RA, Weber MH, Johnson ZC, Olsen AR, Flotemersch JE, Stoddard JL, and Peck DV (2018). Mapping watershed integrity for the conterminous United States. Ecological Indicators, 85: 1133-1148.

Tu J (2009). Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA. Journal of Hydrology, 379(3-4): 268-283.

Twarakavi NKC, Šimůnek J, and MG Schaap (2010). Can texture-based classification optimally classify soils with respect to soil hydraulics? Water Resources Research, 46: W01501.

US Army Corps of Engineers (USACE) (2019). National Inventory of Dams database, http://nid.usace.army.mil/cm_apex/f?p=838:12.

US Census Bureau (2019). 2019 Cartographic Boundary Files, https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html.

US Department of Interior. Snake River Headwaters Comprehensive River Management Plan/Environmental Assessment. Wyoming: Grand Teton National Park, Yellowstone National Park, John D. Rockefeller, Jr Memorial Parkway, and National Elk Refuge, 2013.

US Department of Interior. Virgin River Comprehensive Management Plan/Environmental Assessment. Utah: Zion National Park and Bureau of Land Management, St. George Field Office, 2013.

US Environmental Protection Agency (EPA), Facility Registry Service (2020). Wastewater Treatment Plants, https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=%7B86C00471-0D39-4352-B629-14E3C0E2D042%7D

US Environmental Protection Agency (EPA), Office of Water and Office of Research and Development (2016). National Rivers and Streams Assessment 2008-2009 Technical Report: EPA/841/R-16/008, 140 pp.

US Environmental Protection Agency (EPA), Office of Water Healthy Watersheds Program (2017). Overview of the preliminary healthy watersheds assessments project: Evaluating relative health and vulnerability of conterminous US watersheds, https://www.epa.gov/hwp/healthy-watersheds-integrated-assessment-reports

US Forest Service (2021). Timber Harvests, https://data.fs.usda.gov/geodata/edw/edw_resources/meta/S_USA.Activity_TimberHarvest.xml

US Geological Survey (USGS) (2020). National Water Information System: Water data for the Nation, http://waterdata.usgs.gov/nwis/.

Vanlear M (2019). Soil Texture Calculator, Natural Resource Conservation Service, US Department of Agriculture (NRCS USDA), https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167.

Wei G, Yang Z, Cui B, Li B, Chen H, Bai J, and S Dong (2009). Impact of dam construction on water quality and water self-purification capacity of the Lancang River, China. Water Resources Management, 23: 1763-1780.

Whittemore A, Ross MRV, Dolan W, Langhorst T, Yang X, Pawar S, Jorissen M, Lawton E, Januchowski-Hartley S, T Pavelsky (2020). A participatory science approach to expanding instream infrastructure inventories. Earth’s Future 8(11): e2020EF001558.

Wickham J, Homer C, Vogelmann J, McKerrow A, Mueller R, Herold N, and J Coulston (2014). The multi-resolution land characteristics (MRLC) consortium – 20 years of development and integration of USA national land cover data. Remote Sensing 6: 7424-7441.

Willi K, and J Back (2018). Evaluation of state water quality assessments and the National Wild and Scenic Rivers System. Technical Report of the Interagency Wild and Scenic Rivers Coordinating Council.

Wolock DM (2003). Base-flow index grid for the conterminous United States. US Geological Survey database, https://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml#stdorder.

APPENDIX

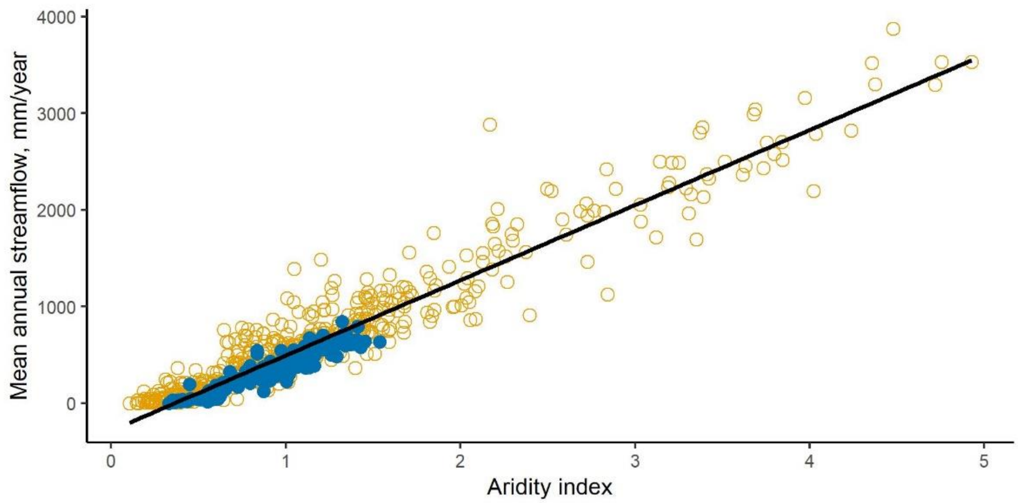

Figure A1. Graph depicting mean annual streamflow against watershed mean aridity index, for reference-quality gages only. Blue points depict stream gages whose watersheds have ≥ 2.5% cropland (NLCD, 2011).

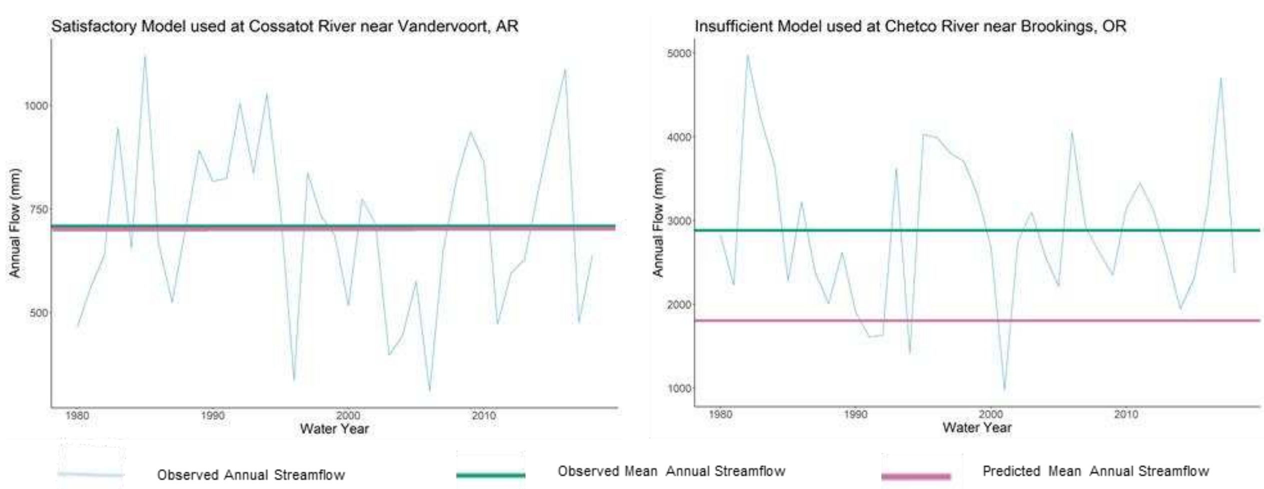

Figure A2. An example of a satisfactory model predicting mean annual flow at a reference stream gage (left), and an example of an insufficient model applied to a reference stream gage (right). For the satisfactory model, the predicted value is close to the true mean annual value. In the insufficient model, the predicted value has underestimated the true mean annual value.

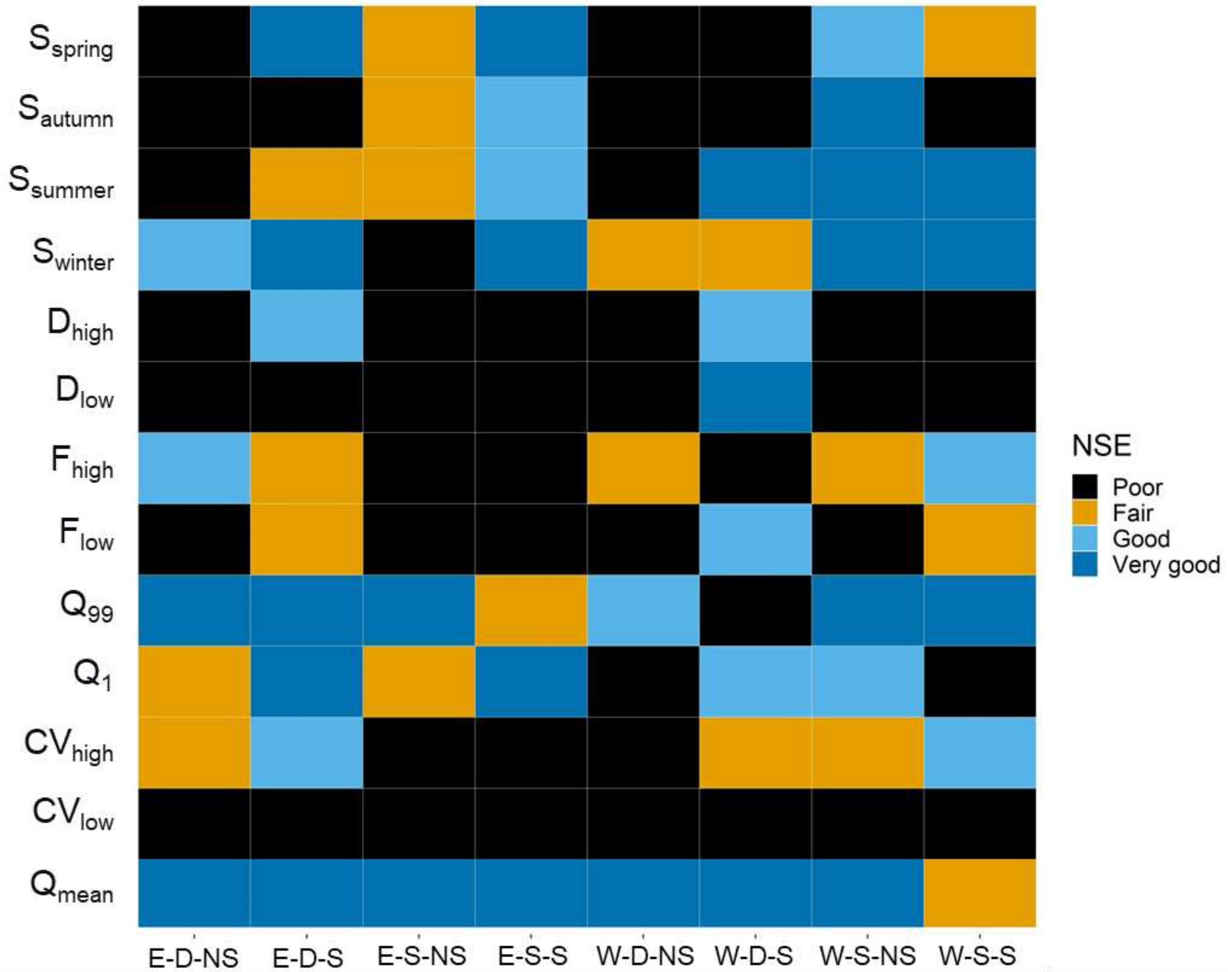

Figure A3. The ratio of the Nash-Sutcliffe efficiency (NSE) for each flow statistic model developed for each watershed group. NSE values > 0.75 are considered very good (dark blue), > 0.65 good (light blue), > 0.50 satisfactory (orange), and ≤ 0.50 are considered poor (Moriasi et al., 2007). Models with NSE values ≤ 0.50 (black) were removed from our analyses.

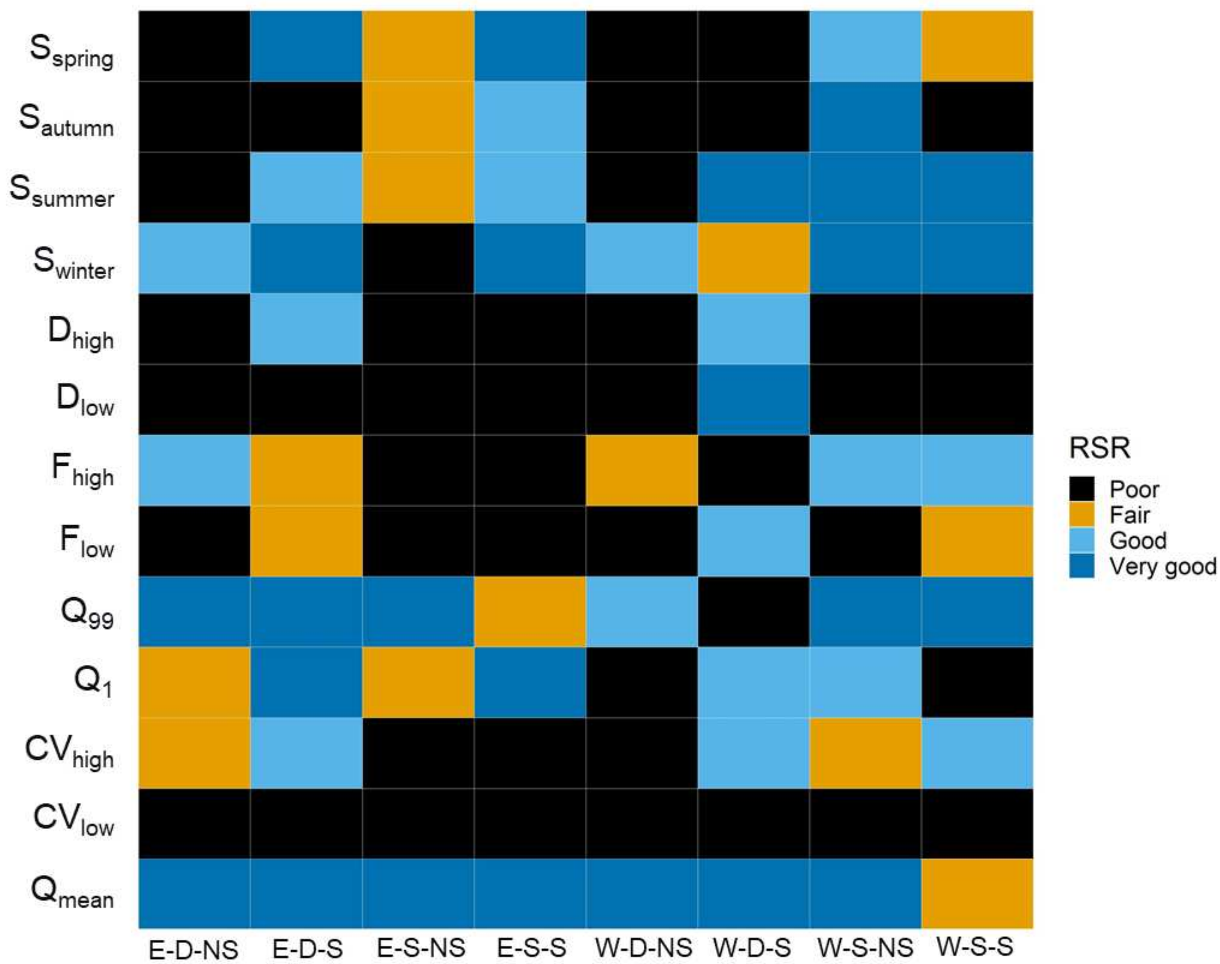

Figure A4. The ratio of the root mean square error to the standard (RSR) for each flow statistic model developed for each watershed group. RSR values ≤ 0.5 are considered very good (dark blue), ≤ 0.6 good (light blue), ≤ 0.7 satisfactory (orange), and > 0.7 are considered poor (Moriasi et al., 2007). Models with RSR values > 0.7 (black) were removed from our analyses.

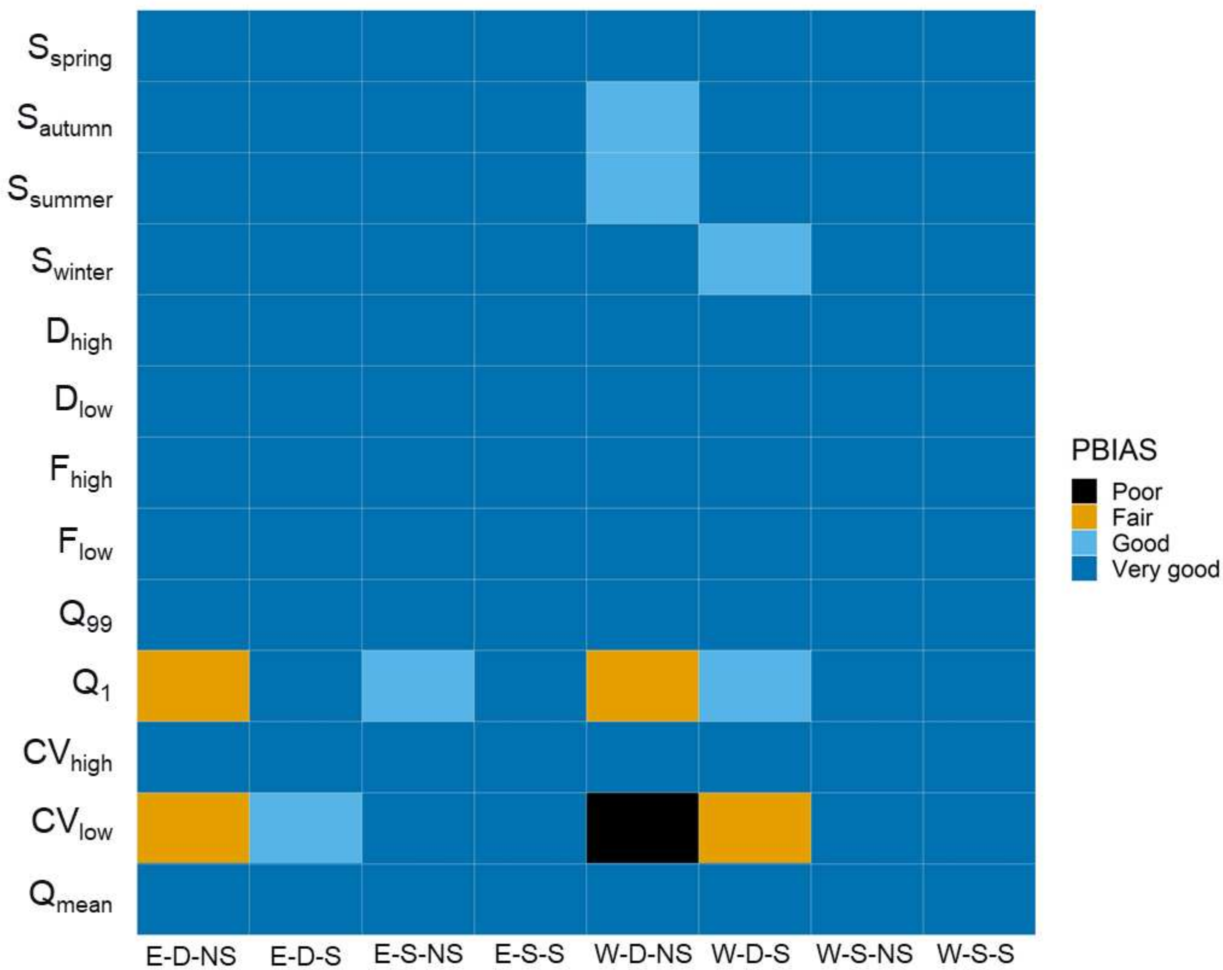

Figure A5. The percent bias (PBIAS) for each flow statistic model developed for each watershed group. PBIAS values < ±10 are considered very good (dark blue), < ±15 good (light blue), < ±25 satisfactory (orange), and ≥ ±25 are considered poor (Moriasi et al., 2007). Models with PBIAS values ≥ ±25 (black) were removed from our analyses.

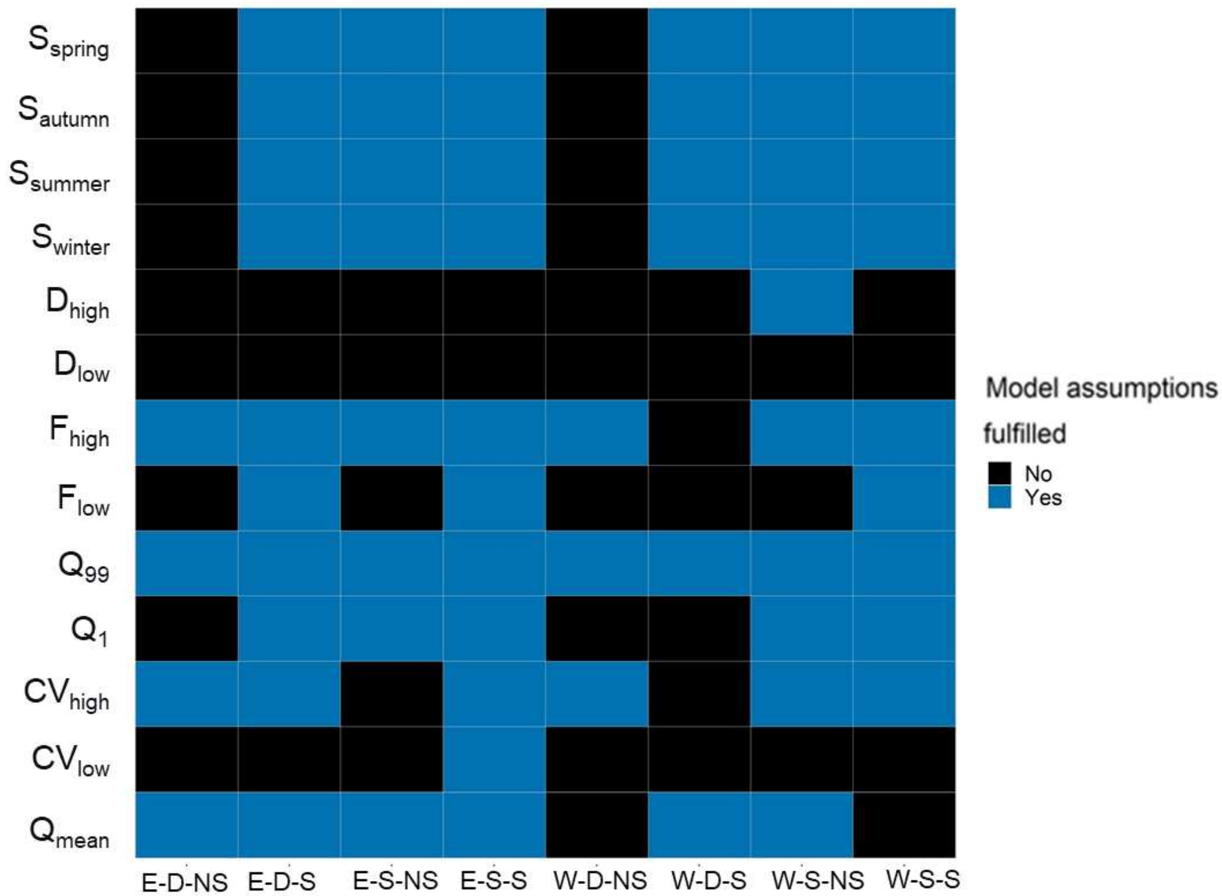

Figure A6. Whether the final model met assumptions of normality and equal variance of the residual. Models that did not meet these assumptions (black) were removed from our analyses.

	E-D-NS	E-D-S	E-S-NS	E-S-S	W-D-NS	W-D-S	W-S-NS	W-S-S
S_{spring}		0.115	0.107	0.022			0.107	0.376
S_{autumn}			0.17	0.025			0.125	
S_{summer}		0.049	0.136	0.063		1.495	0.541	0.349
S_{winter}		0.061		0.077		1.603	0.191	0.121
D_{high}								
D_{low}								
F_{high}	22.249	9.087			15.13		7.238	9.5
F_{low}		3.323						3.092
Q_{99}	11.83	2.21	15.33	9.031	41.253		31.109	21.971
Q_1		0.276	2.05	1.204			2.476	
CV_{high}	171.102	20.4					55.759	31.953
CV_{low}								
Q_{mean}	221.71	112.547	244.218	236.957		634.264	1291.127	

Figure A7. The predicted sum of squares (PRESS) statistic, which is a leave-one-out measure of cross validation that expresses how sensitive a model is to any single observation used in the analysis. Units of the PRESS statistic are reported in each flow metric's calculated units. Models that did not meet our model requirements (black) were removed from our analyses.

Table A1. Table of predictor variables used in the development of each flow metric model.

Flow Metric	Model Region	Selected Predictor Variables
Q_{mean}	E-S-NS	Mean Soil Clay Content, Mean Erodibility Factor, Dominant Aspect, Mean Fall Precipitation, Gage Elevation, Mean Aridity Index
	E-S-S	Mean Soil Organic Matter, Mean Fall Precipitation, Mean Baseflow Index, Percent Herbaceous Landcover, Percent Wetland Landcover
	E-D-NS	Mean Fall Precipitation, Gage Elevation, Gage Latitude, Mean Baseflow Index, Percent Watershed with > 30% Slope, Mean Winter Precipitation
	E-D-S	Dominant Aspect, Mean Aridity Index, Mean Slope, Percent Forest Landcover, Percent Herbaceous Landcover, Percent Wetland Landcover
	W-S-NS	Mean Soil Sand Content, Mean Soil Organic Matter, Mean Fall Precipitation, Gage Elevation, Gage Latitude, Mean Soil Permeability
	W-D-S	Mean Erodibility Factor, Mean Soil Sand Content, Mean Soil Silt Content, Dominant Lithology, Gage Longitude, Mean Annual Precipitation
S_{winter}	E-S-S	Dominant Lithology, Mean Fall Precipitation, Mean Baseflow Index, Mean Snow Persistence, Percent Herbaceous Landcover, Percent Wetland Landcover
	E-D-S	Mean Soil Clay Content, Mean Baseflow Index, Mean Snow Persistence, Mean Spring Precipitation, Mean Summer Precipitation
	W-S-NS	Mean Soil Sand Content, Dominant Lithology, Watershed Area, Gage Elevation, Gage Latitude, Gage Longitude
	W-S-S	Mean Soil Sand Content, Mean Soil Organic Matter, Dominant Lithology, Gage Longitude, Mean Baseflow Index, Mean Snow Persistence
	W-D-S	Dominant Lithology, Gage Elevation, Gage Latitude
S_{spring}	E-S-NS	Mean Soil Silt Content, Mean Soil Organic Matter, Watershed Area, Mean Fall Precipitation, Mean Baseflow Index, Mean Snow Persistence

S_{spring}	E-S-S	Mean Erodibility Factor, Mean Soil Organic Matter, Mean Fall Precipitation, Mean Snow Persistence, Percent Wetland Landcover, Mean Spring Precipitation
	E-D-S	Mean Soil Sand Content, Mean Soil Organic Matter, Watershed Area, Mean Fall Precipitation, Gage Elevation, Mean Baseflow Index
	W-S-NS	Mean Erodibility Factor, Dominant Lithology, Watershed Area, Mean Fall Precipitation, Mean Snow Persistence, Mean Slope
	W-S-S	Mean Soil Clay Content, Mean Erodibility Factor, Mean Baseflow Index, Mean Snow Persistence, Percent Forest Landcover, Percent Shrub Landcover
S_{summer}	E-S-NS	Mean Erodibility Factor, Mean Soil Organic Matter, Watershed Area, Mean Baseflow Index, Mean Aridity Index, Mean Soil Permeability
	E-S-S	Mean Soil Sand Content, Mean Soil Organic Matter, Gage Longitude, Mean Baseflow Index, Mean Aridity Index, Mean Soil Permeability
	E-D-S	Mean Fall Precipitation, Mean Baseflow Index, Mean Soil Permeability, Percent Shrub Landcover, Mean Summer Precipitation
	W-S-NS	Mean Soil Silt Content, Dominant Lithology, Dominant Aspect, Watershed Area, Gage Elevation, Gage Longitude
	W-S-S	Mean Erodibility Factor, Dominant Lithology, Dominant Aspect, Gage Longitude, Mean Snow Persistence, Percent Forest Landcover
	W-D-S	Dominant Lithology, Watershed Area, Gage Elevation, Gage Latitude, Mean Slope, Percent Forest Landcover
S_{autumn}	E-S-NS	Mean Soil Organic Matter, Watershed Area, Mean Fall Precipitation, Mean Baseflow Index, Mean Aridity Index, Mean Soil Permeability
	E-S-S	Mean Soil Sand Content, Mean Fall Precipitation, Gage Elevation, Mean Aridity Index, Percent Herbaceous Landcover, Percent Shrub Landcover
	W-S-NS	Mean Soil Clay Content, Mean Erodibility Factor, Dominant Aspect, Watershed Area, Mean Fall Precipitation, Gage Elevation
Q99	E-S-NS	Dominant Aspect, Mean Fall Precipitation, Mean Baseflow Index, Mean Aridity Index, Mean Snow Persistence, Mean Spring Precipitation

Q99	E-S-S	Mean Soil Organic Matter, Watershed Area, Mean Fall Precipitation, Gage Elevation, Mean Baseflow Index
	E-D-NS	Mean Soil Organic Matter, Mean Fall Precipitation, Gage Latitude, Mean Baseflow Index, Percent Wetland Landcover, Mean Winter Precipitation
	E-D-S	Mean Baseflow Index, Mean Aridity Index, Mean Slope, Percent Herbaceous Landcover, Mean Spring Precipitation
	W-S-NS	Dominant Lithology, Mean Fall Precipitation, Gage Elevation, Gage Latitude, Mean Baseflow Index, Mean Slope
	W-S-S	Mean Baseflow Index, Mean Snow Persistence, Percent Shrub Landcover, Percent Watershed with > 30% Slope, Mean Spring Precipitation, Mean Summer Air Temperature
	W-D-NS	Mean Soil Clay Content, Mean Baseflow Index, Percent Herbaceous Landcover, Mean Spring Precipitation, Mean Summer Air Temperature, Mean Winter Precipitation
Q1	E-S-NS	Mean Soil Clay Content, Mean Baseflow Index, Percent Shrub Landcover, Mean Spring Precipitation
	E-S-S	Mean Soil Clay Content, Mean Soil Organic Matter, Gage Longitude, Mean Baseflow Index, Mean Aridity Index, Percent Shrub Landcover
	E-D-S	Mean Baseflow Index, Percent Forest Landcover, Percent Shrub Landcover
	W-S-NS	Dominant Aspect, Mean Fall Precipitation, Gage Elevation, Mean Soil Permeability, Mean Snow Persistence, Mean Slope
CV_{high}	E-D-NS	Mean Fall Precipitation, Gage Latitude, Mean Snow Persistence, Mean Winter Precipitation
	E-D-S	Mean Baseflow Index, Percent Shrub Landcover, Mean Spring Precipitation
	W-S-NS	Mean Soil Sand Content, Dominant Lithology, Gage Elevation, Mean Snow Persistence, Mean Summer Air Temperature
	W-S-S	Dominant Aspect, Watershed Area, Gage Latitude, Gage Longitude, Mean Soil Permeability, Percent Herbaceous Landcover

F_{high}	E-D-NS	Mean Soil Silt Content, Gage Latitude, Mean Snow Persistence, Percent Shrub Landcover, Percent Watershed with > 30% Slope, Mean Winter Precipitation
	E-D-S	Mean Soil Clay Content, Mean Soil Sand Content, Watershed Area, Gage Elevation, Gage Longitude, Mean Snow Persistence
	W-S-NS	Mean Erodibility Factor, Dominant Lithology, Watershed Area, Gage Elevation, Gage Latitude, Percent Wetland Landcover
	W-S-S	Mean Soil Clay Content, Mean Soil Organic Matter, Gage Elevation, Gage Latitude, Gage Longitude, Mean Baseflow Index
	W-D-NS	Watershed Area, Mean Soil Permeability, Mean Annual Precipitation, Median Elevation, Mean Spring Precipitation
F_{low}	E-D-S	Mean Soil Clay Content, Watershed Area, Mean Fall Precipitation, Gage Elevation, Gage Longitude, Mean Snow Persistence
	W-S-S	Mean Erodibility Factor, Mean Soil Organic Matter, Dominant Aspect, Gage Latitude, Percent Forest Landcover, Mean Summer Precipitation

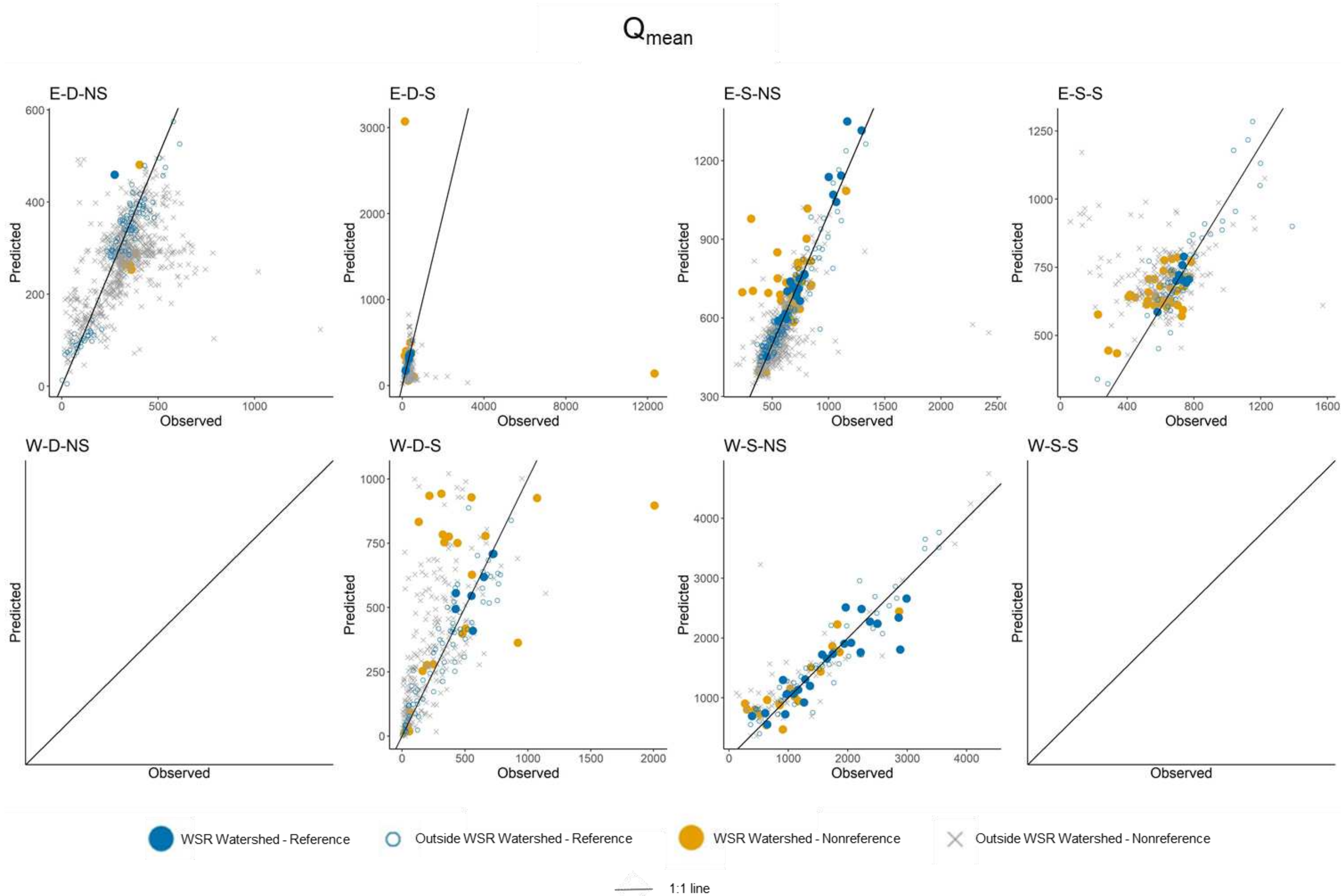

Figure A8. Predicted vs. observed mean annual flow (mm/year) at all gages used in the analysis. Blank graphs represent regions with insufficient models.

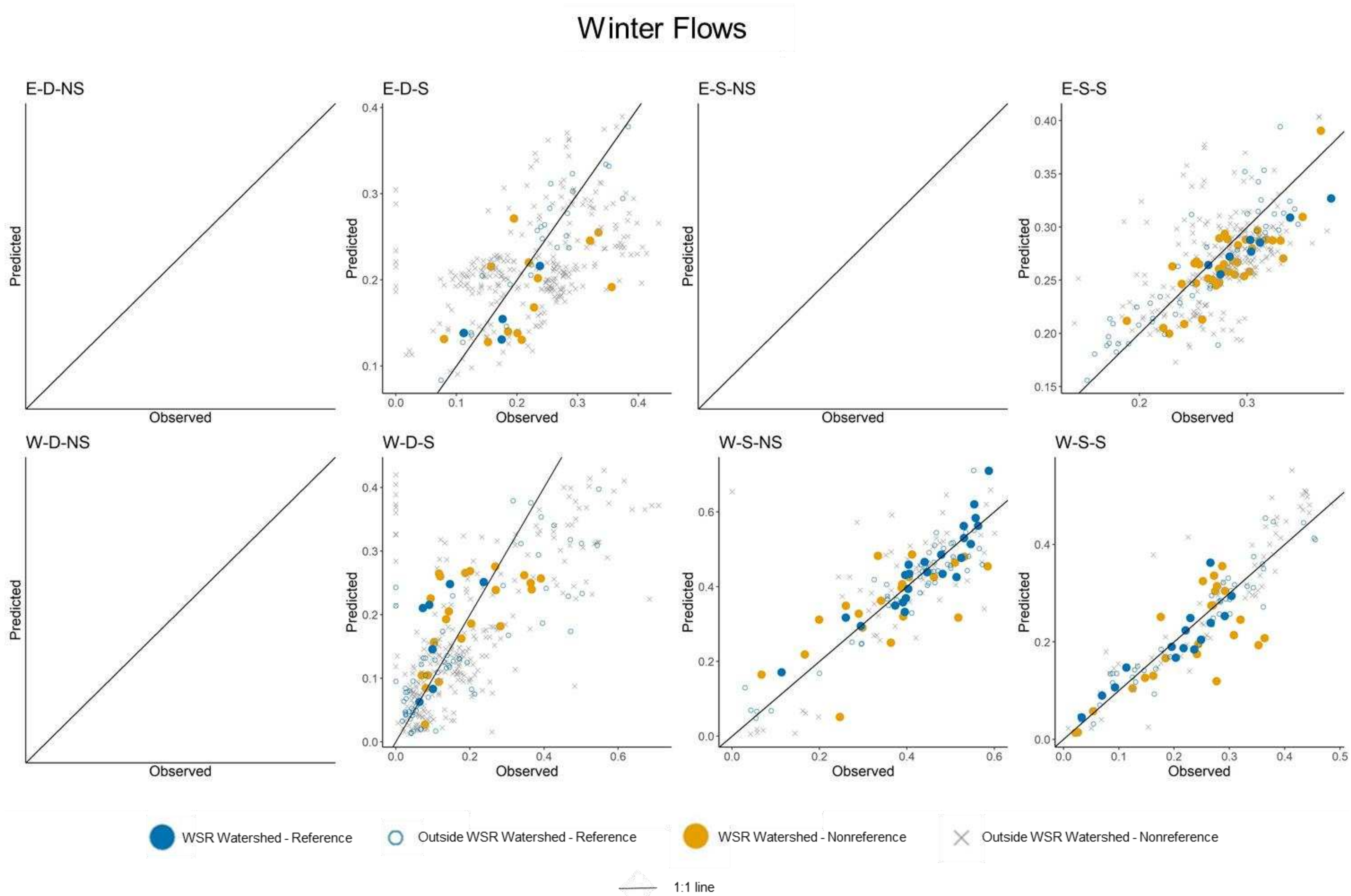

Figure A9. Predicted vs. observed values for the fraction of the mean annual flow in winter at all gages used in the analysis. Blank graphs represent regions with insufficient models.

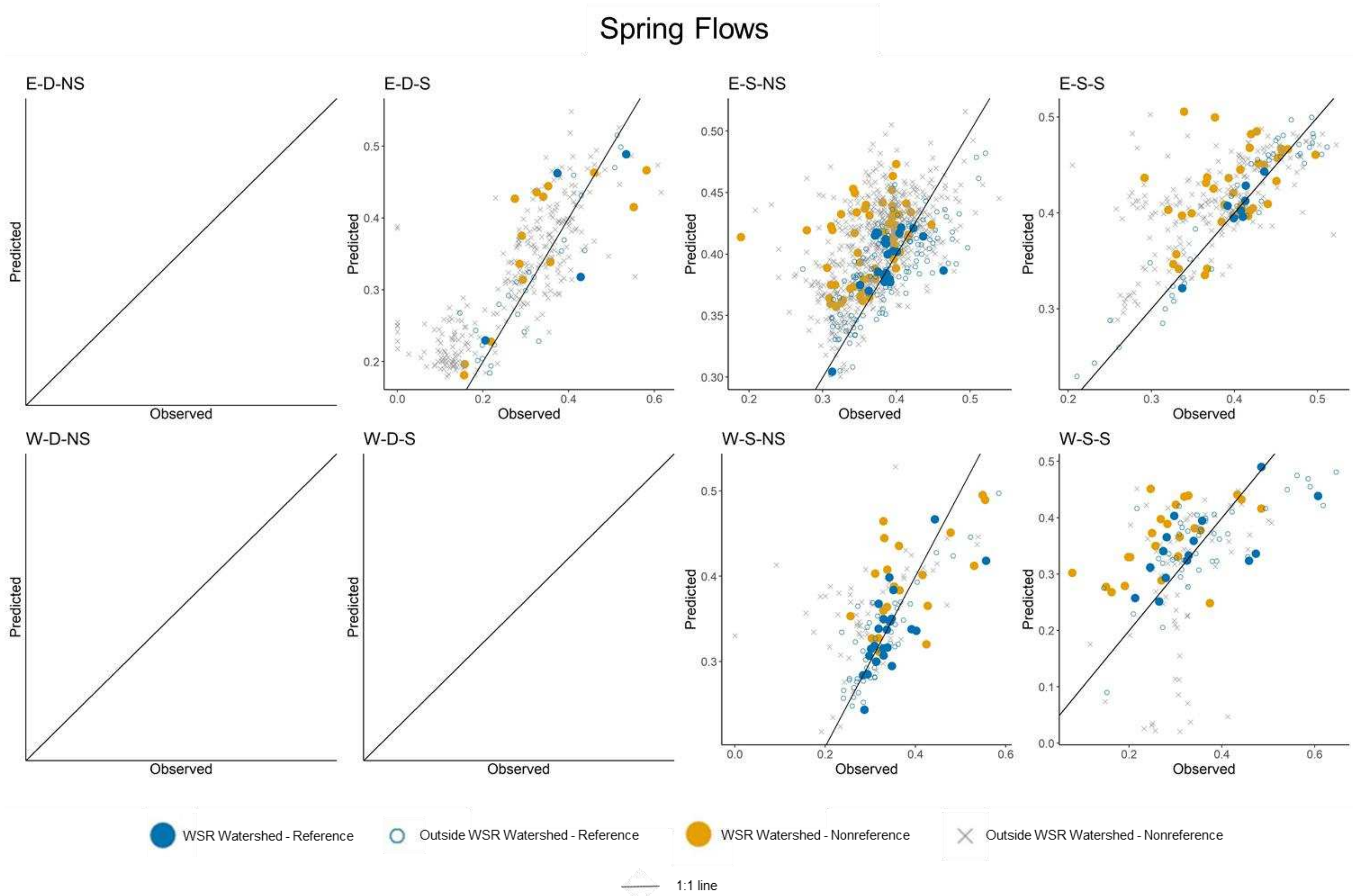

Figure A10. Predicted vs. observed values for the fraction of the mean annual flow in spring at all gages used in the analysis. Blank graphs represent regions with insufficient models.

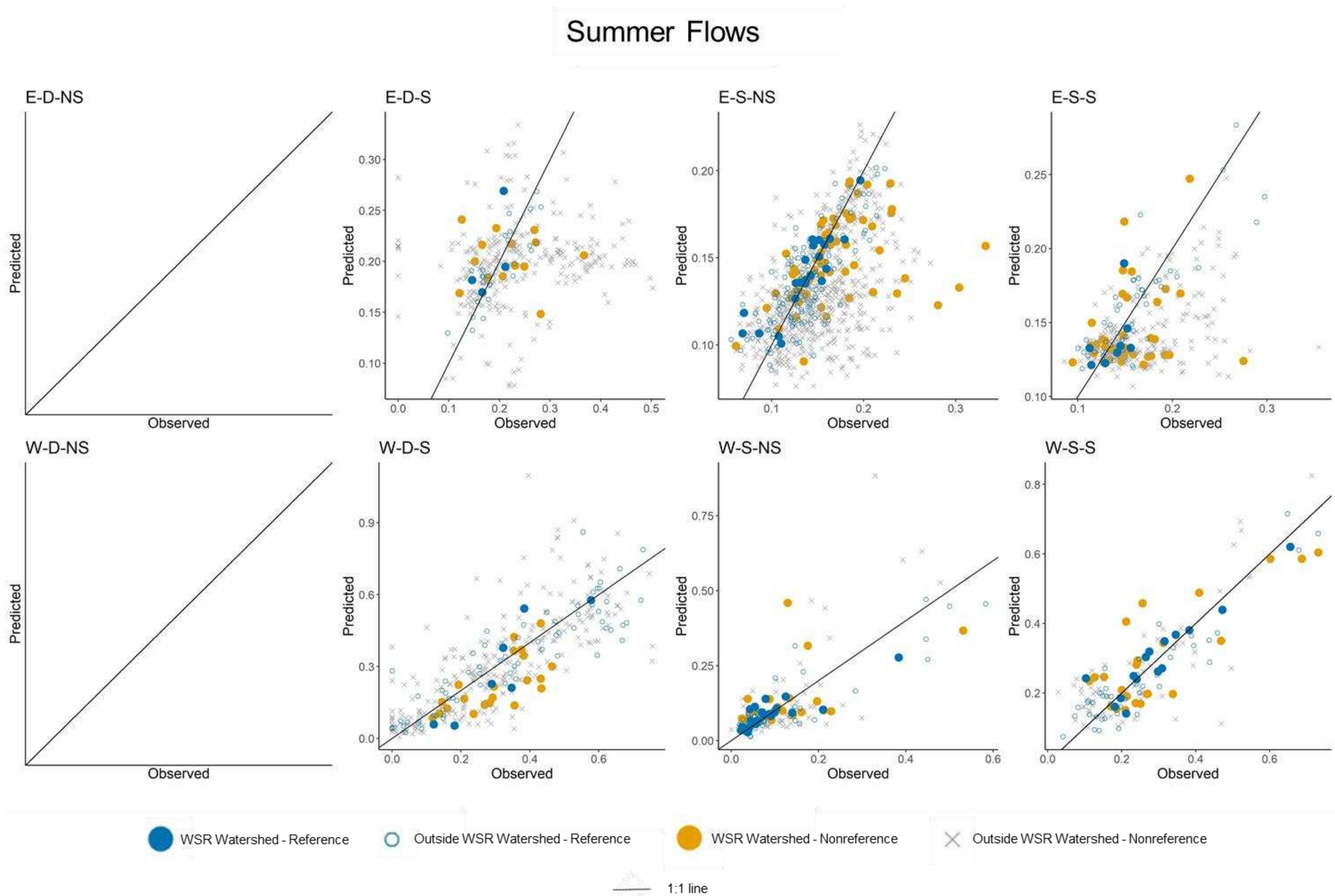

Figure A11. Predicted vs. observed values for the fraction of the mean annual flow in summer at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

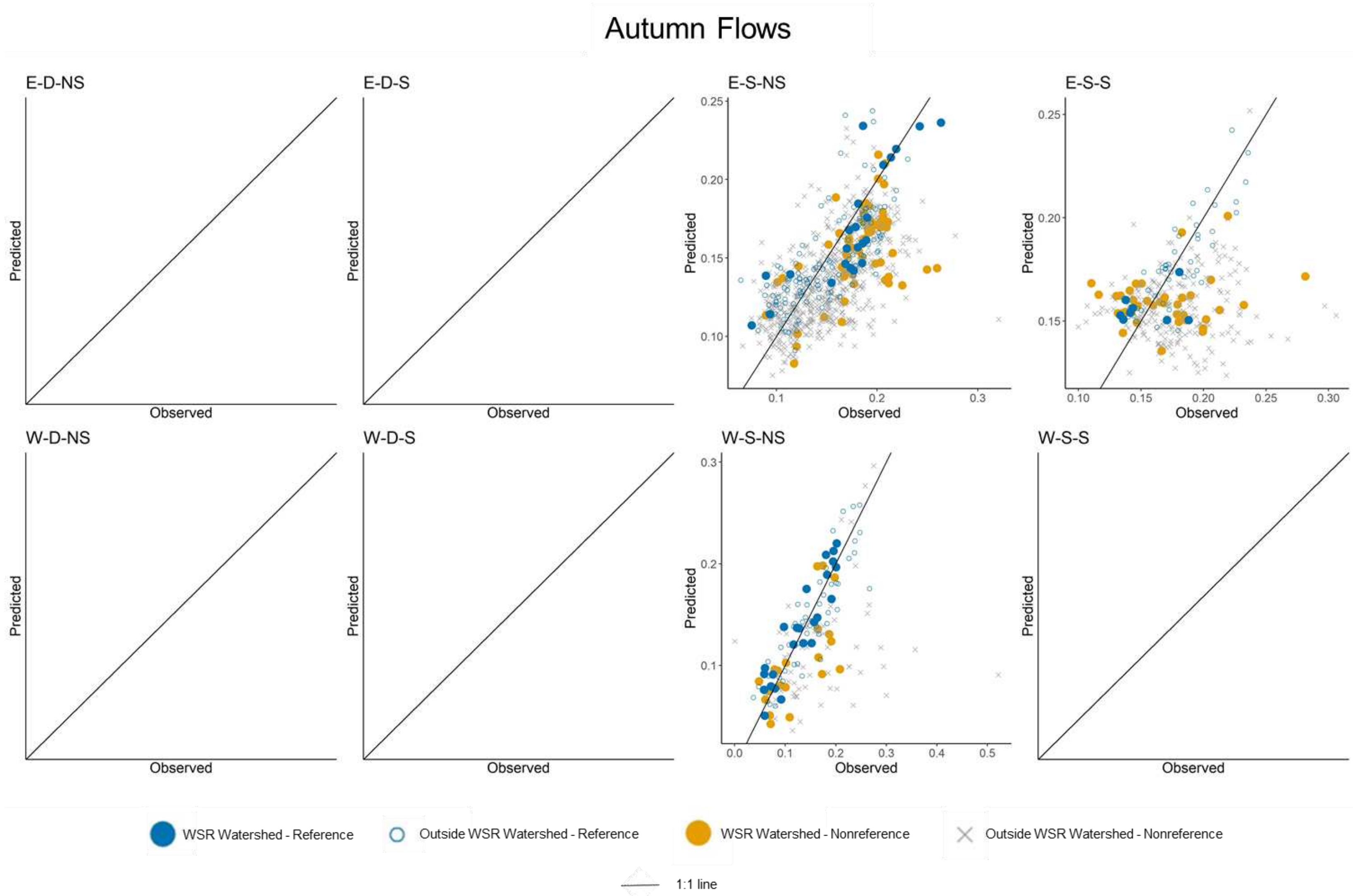

Figure A12. Predicted vs. observed values for the fraction of the mean annual flow in autumn at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

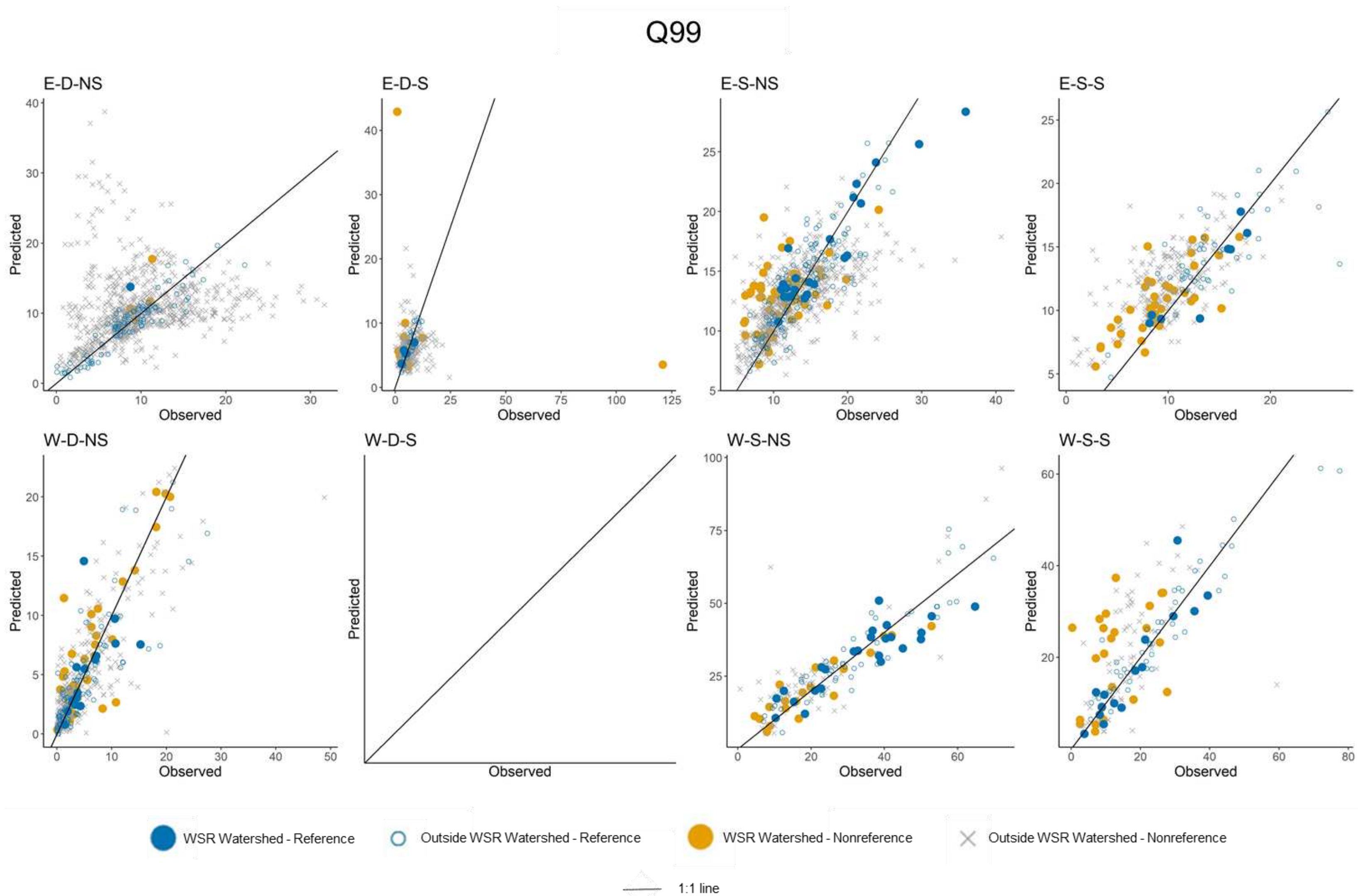

Figure A13. The predicted vs. observed 99th percentile non-exceedance flow (Q99) at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

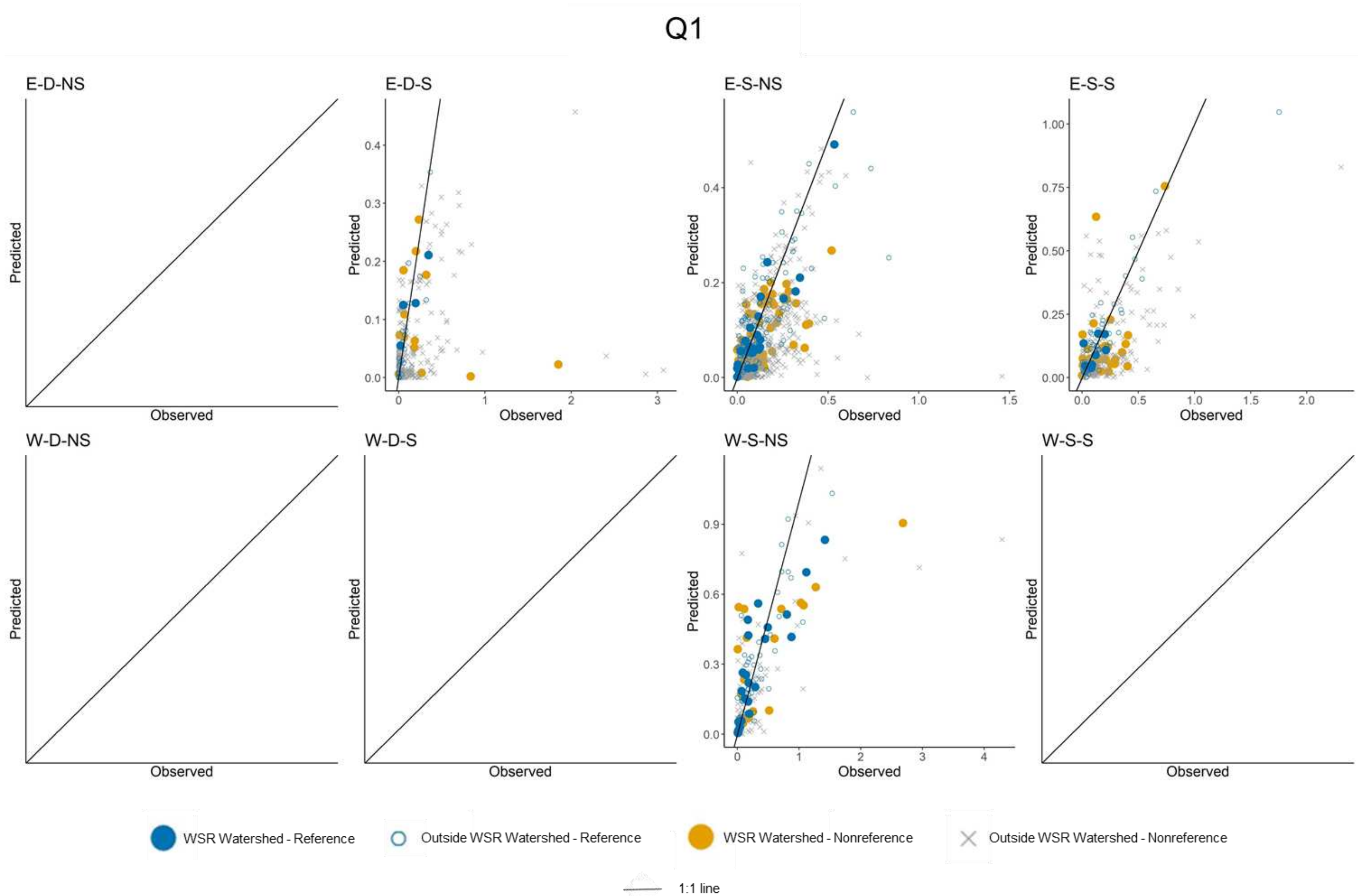

Figure A14. The predicted vs. observed 1st percentile non-exceedance flow (Q1) at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

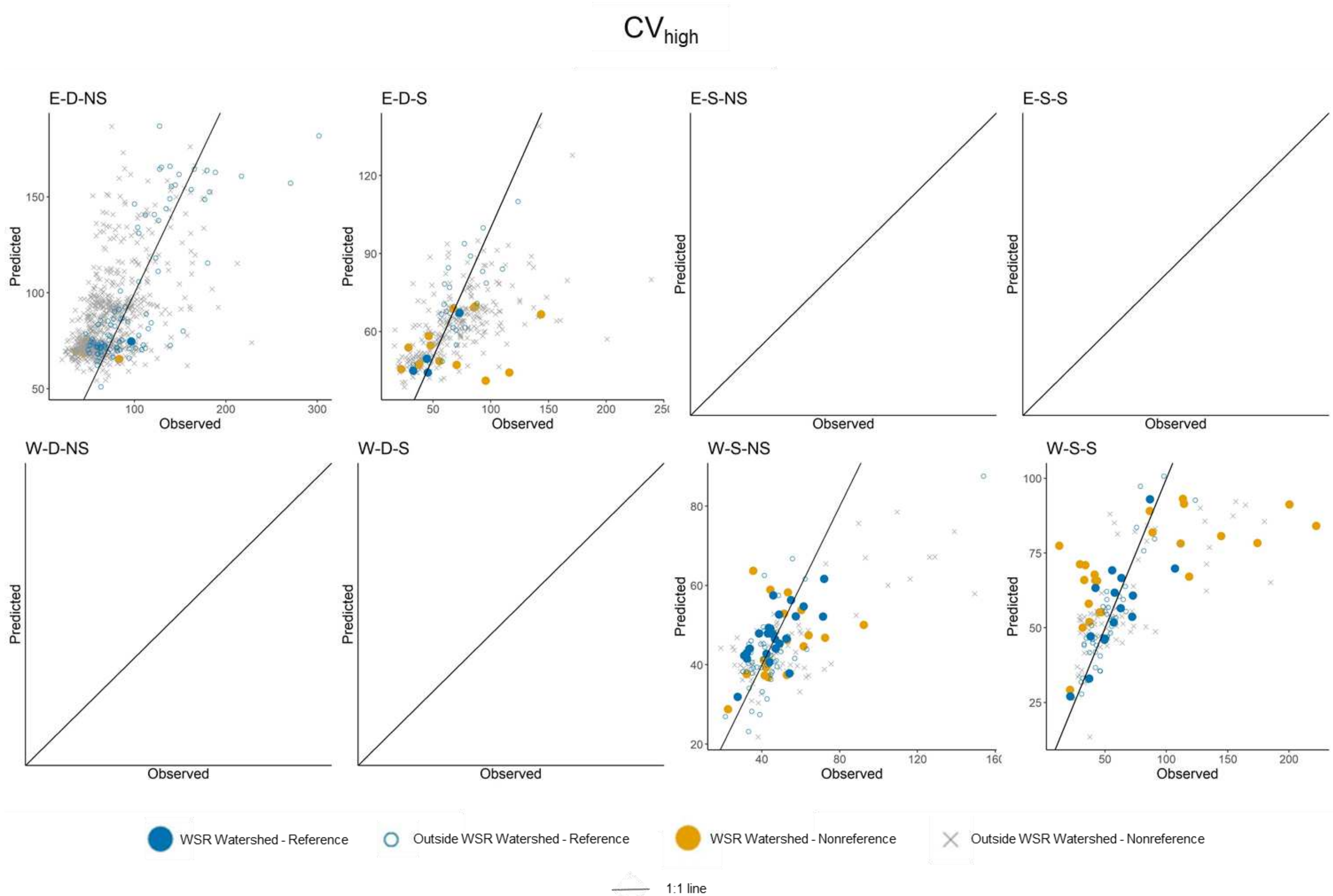

Figure A15. The predicted vs. observed high-flow coefficient of variation (CV_{high}) at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

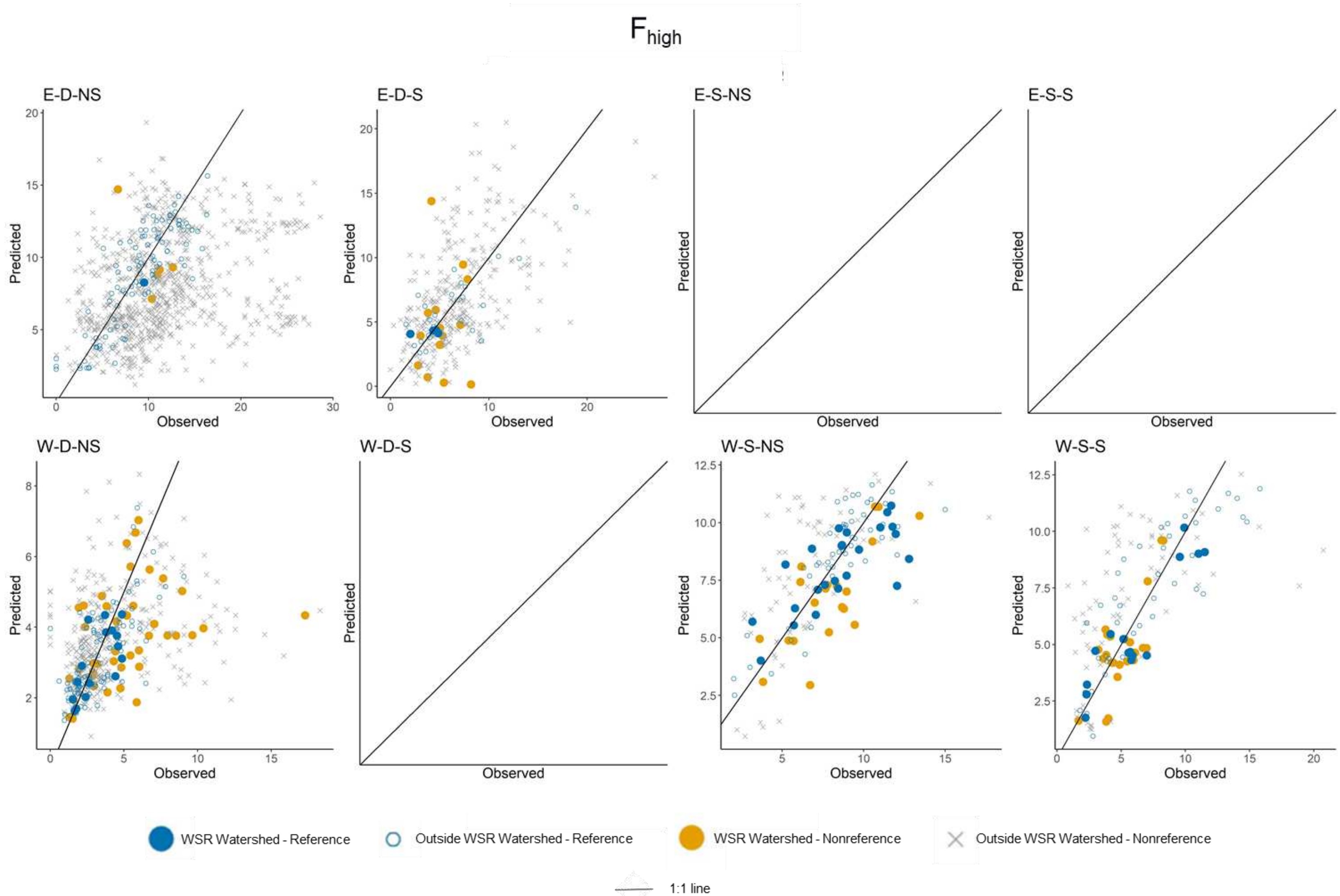

Figure A16. The predicted vs. observed frequency of high flows (F_{high}, number of flow pulses) at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

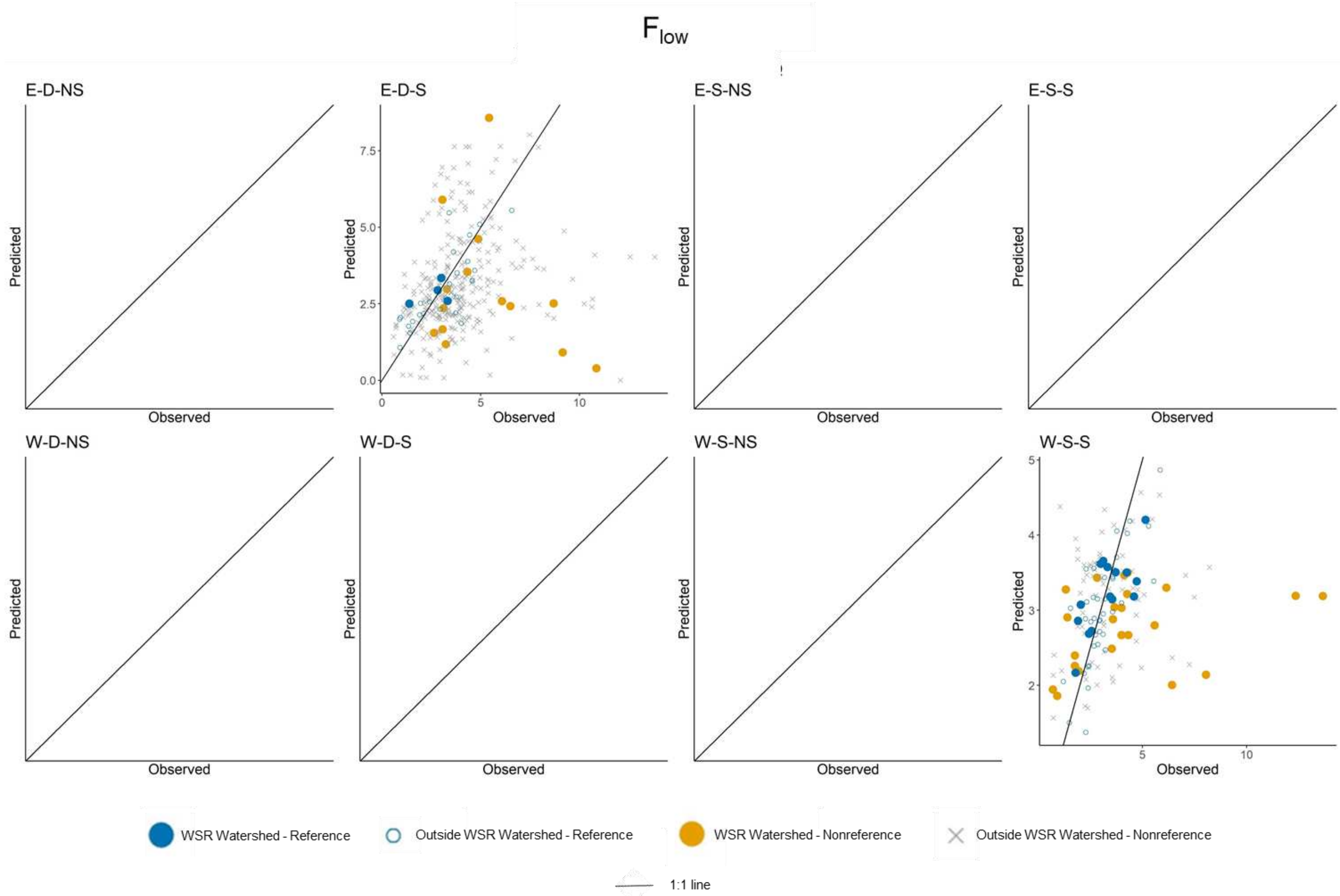

Figure A17. The predicted vs. observed frequency of low flows (F_{low}, number of flow pulses) at all gages used in the analysis, by region. Blank graphs represent regions with insufficient models.

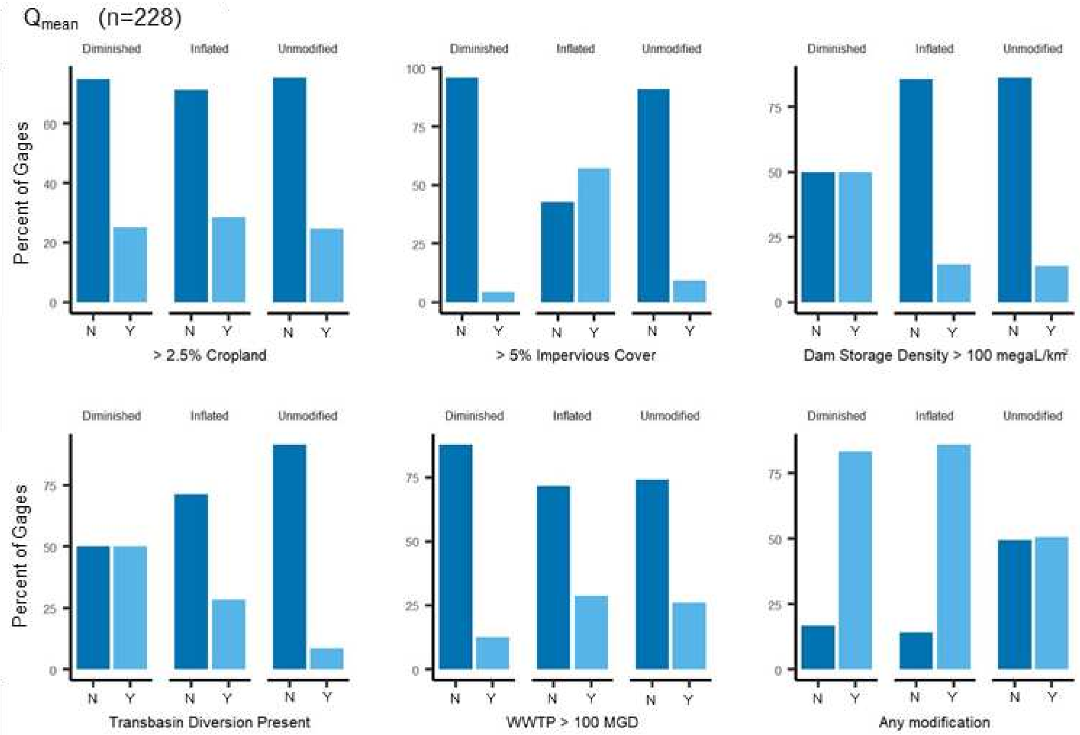

Figure A18. Percentage of WSR gages used in analyzing mean annual flow (Q_{mean}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

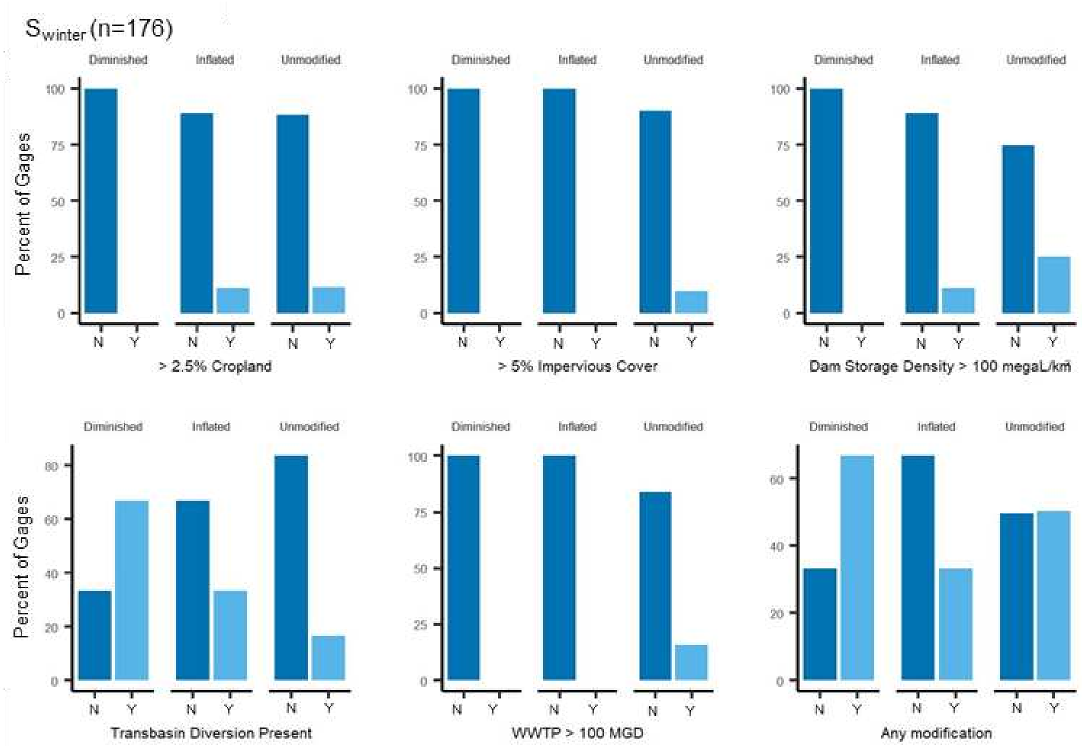

Figure A19. Percentage of WSR gages used in analyzing winter flow (S_{winter}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

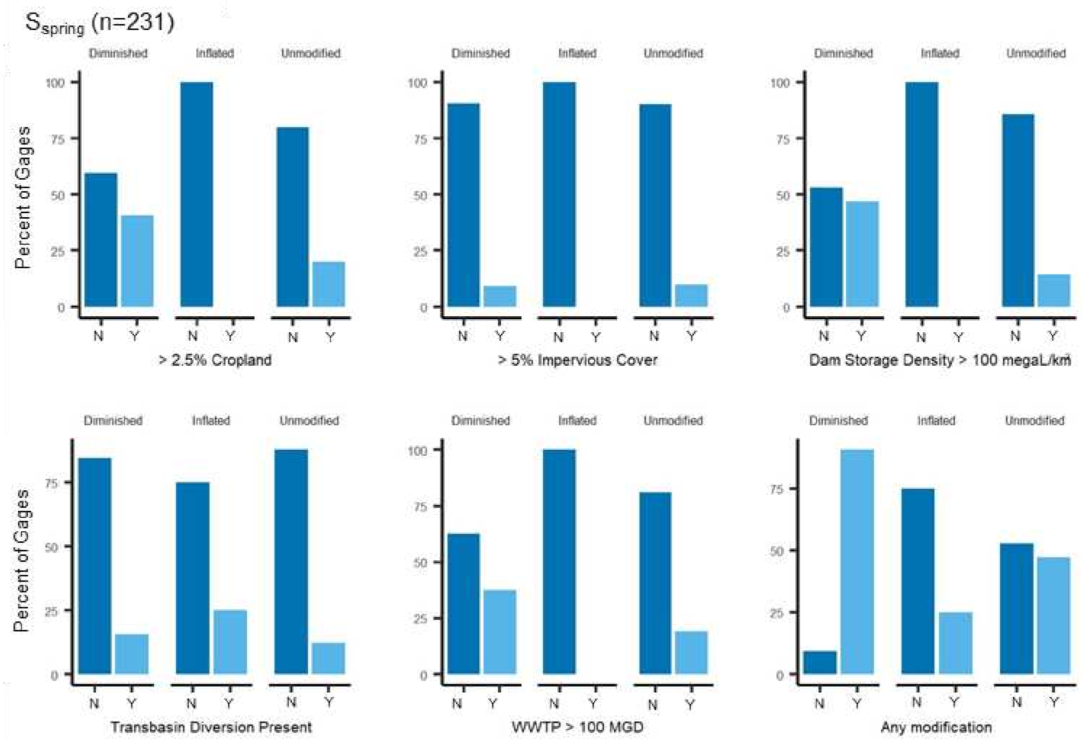

Figure A20. Percentage of WSR gages used in analyzing spring flow (S_{spring}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

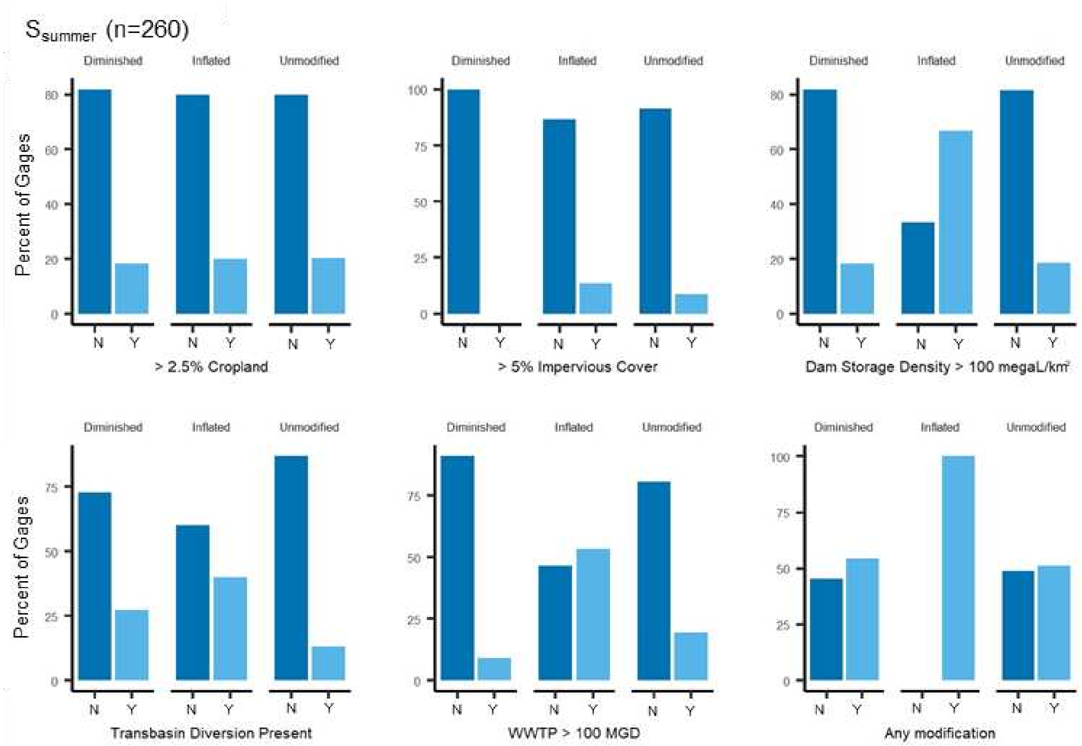

Figure A21. Percentage of WSR gages used in analyzing summer flow (S_{summer}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

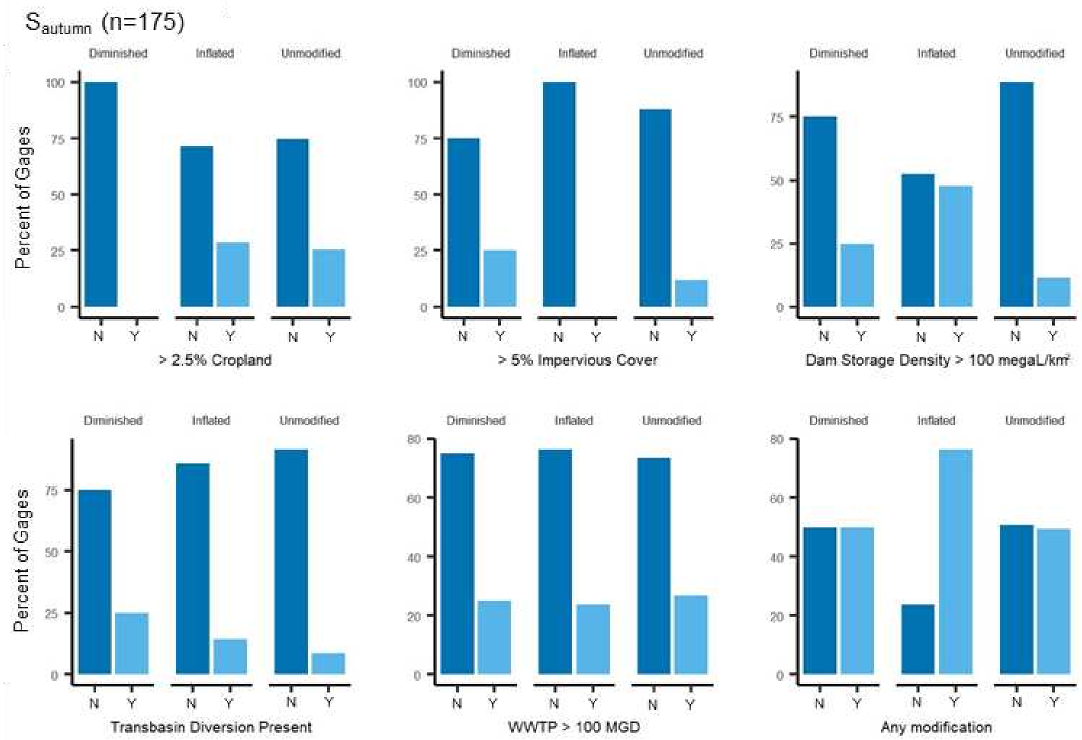

Figure A22. Percentage of WSR gages used in analyzing autumn flow (S_{autumn}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

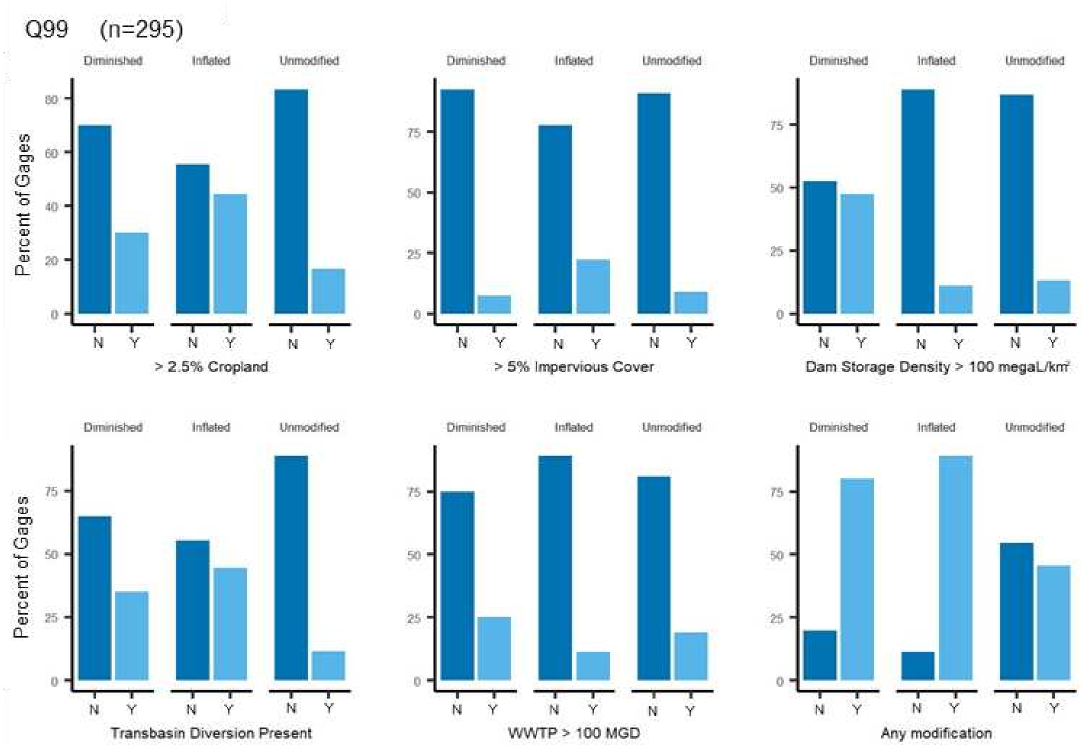

Figure A23. Percentage of WSR gages used in analyzing the 99th percentile non-exceedance flow (Q99) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

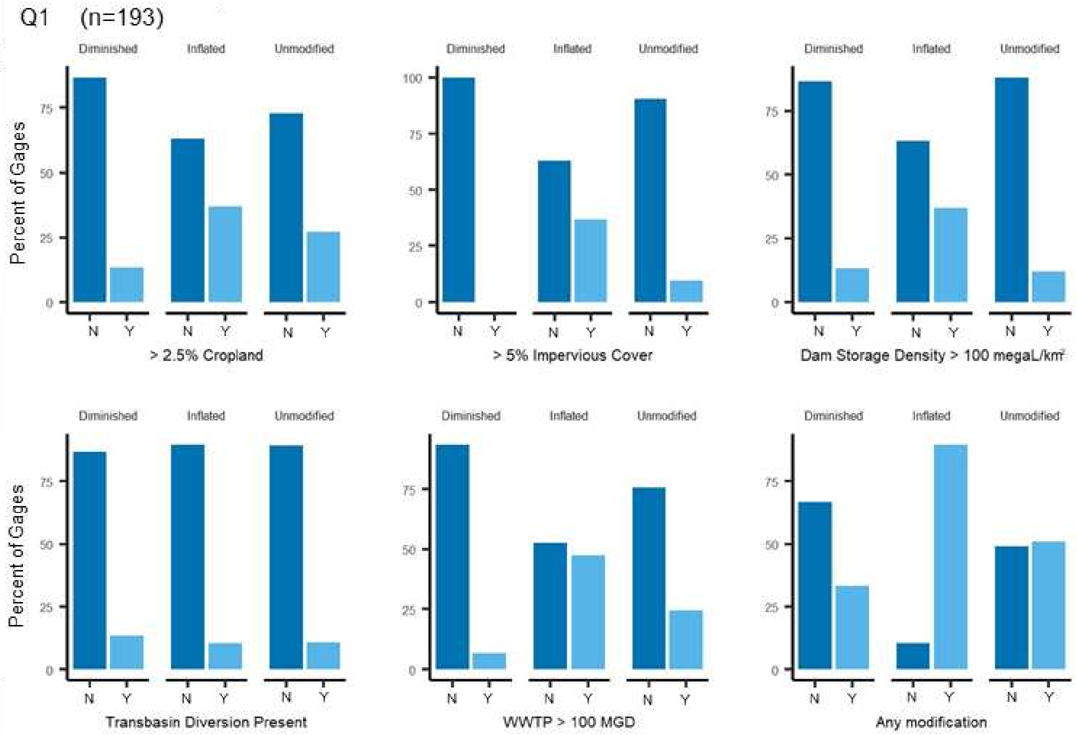

Figure A24. Percentage of WSR gages used in analyzing the 1st percentile non-exceedance flow (Q1) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

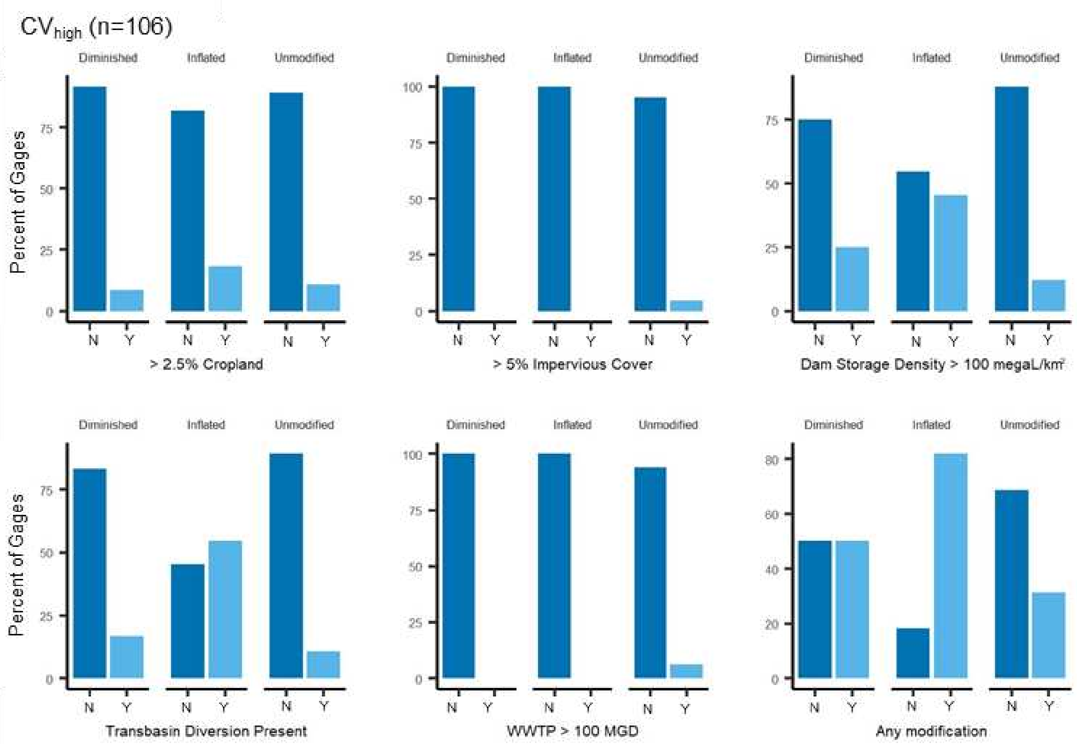

Figure A25. Percentage of WSR gages used in analyzing the coefficient of variation of annual maximum daily flows (CV_{high}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

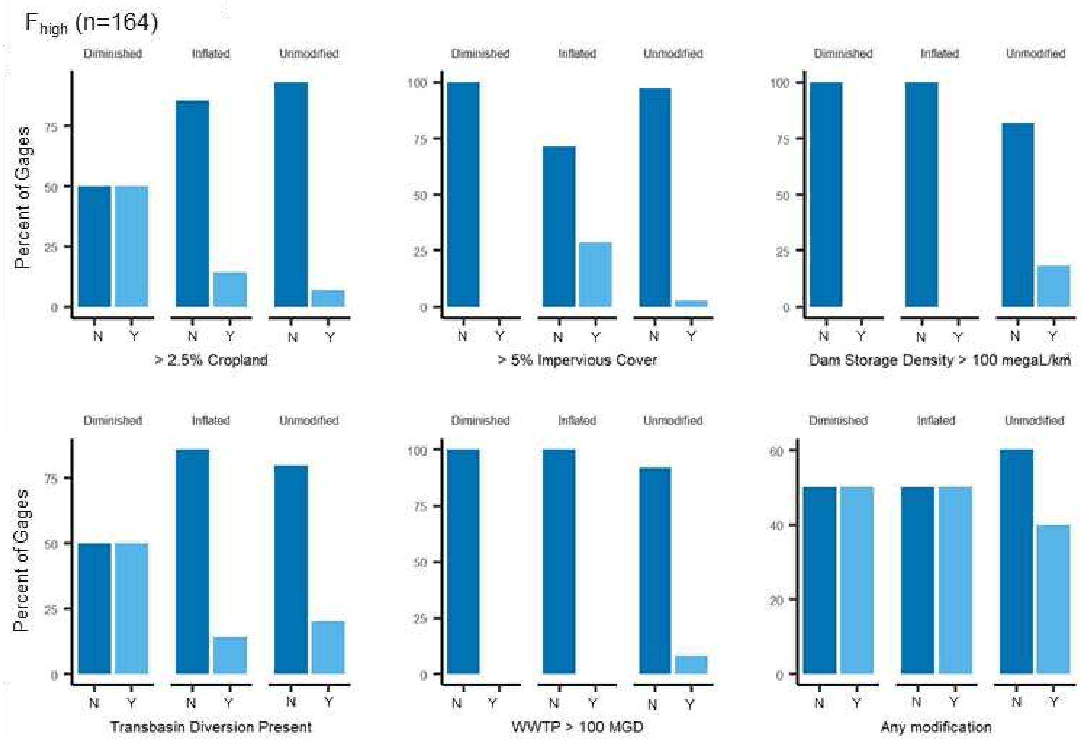

Figure A26. Percentage of WSR gages used in analyzing the frequency of high flows (F_{high}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.

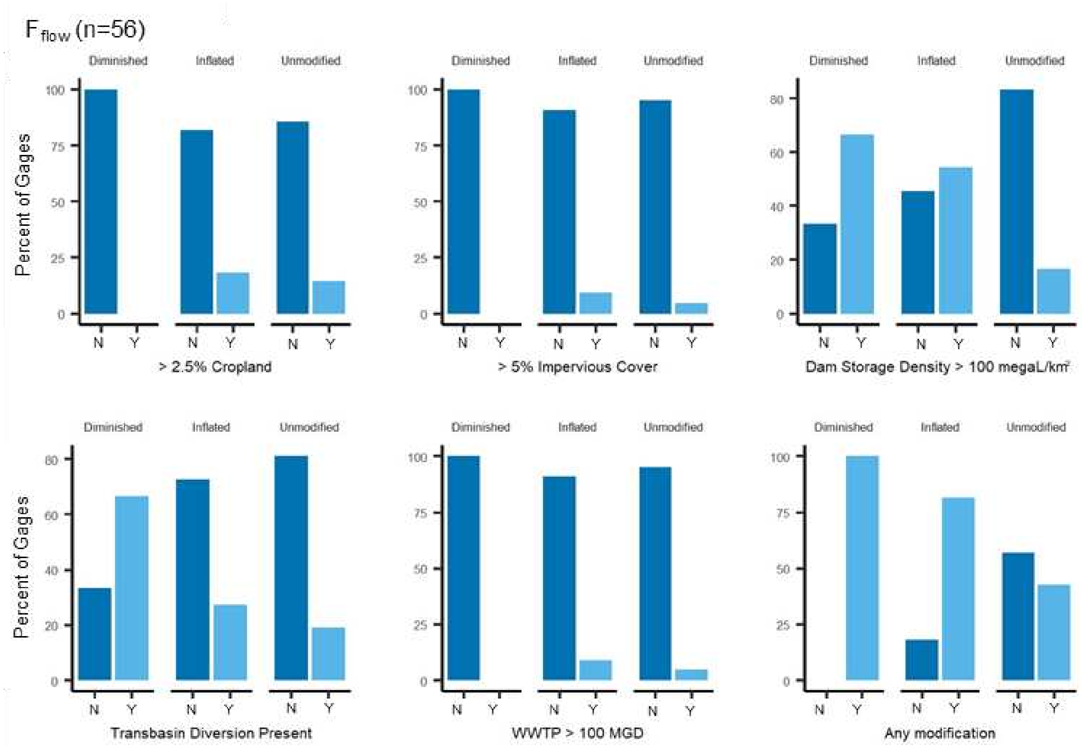

Figure A27. Percentage of WSR gages used in analyzing the frequency of low flows (F_{low}) with basin alterations, based on our cutoffs for selecting reference gages and grouped by alteration type.