

# **Dynamic Stream Permanence Estimates at Regional and Local Extents**

A Dissertation

Presented in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy

with a

Major in Water Resources Science and Management

in the

College of Graduate Studies

University of Idaho

by

Konrad C. Hafen

Major Professor: Paul E. Gessler, Ph.D.

Committee Members: Kyle W. Blasch, Ph.D.; Erin S. Brooks, Ph.D.; Timothy E. Link, Ph.D.

Department Administrator: Timothy E. Link, Ph.D.

May 2021

### Authorization to Submit Dissertation

This dissertation of Konrad C. Hafen, submitted for the degree of Doctor of Philosophy with a Major in Water Resources Science and Management and titled "Dynamic Stream Permanence Estimates at Regional and Local Extents," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: \_\_\_\_\_ Date: \_\_\_\_\_  
Paul E. Gessler, Ph.D.

Committee Members: \_\_\_\_\_ Date: \_\_\_\_\_  
Kyle W. Blasch, Ph.D.

\_\_\_\_\_ Date: \_\_\_\_\_  
Erin S. Brooks, Ph.D.

\_\_\_\_\_ Date: \_\_\_\_\_  
Timothy E. Link, Ph.D.

Department  
Administrator: \_\_\_\_\_ Date: \_\_\_\_\_  
Timothy E. Link, Ph.D.

## Abstract

In the United States (US), the frequency and duration of surface water in a stream channel (i.e. stream permanence) determines if a stream is subject to regulation under the Clean Water Act. While stream permanence is important for policy implementation, quantifying streamflow and water quantity through observation and modeling has been the primary focus of water resource managers. The most comprehensive dataset of stream permanence classifications for the US is the National Hydrography Dataset (NHD), which gives classifications of perennial, intermittent, or ephemeral for most stream reaches. NHD stream permanence classifications were made during topographic map field surveys that occurred from approximately 1920-2000 and have been shown to exhibit high rates of disagreement with more recent stream permanence observations. Thus, there is currently not an available stream permanence dataset with sufficient accuracy for regulatory determinations. I present three studies to assess the influence of climate in NHD stream permanence disagreements, implement a monthly water balance model (MWBM) to create dynamic stream permanence estimates for headwater streams in the NHD network, and apply the Watershed Erosion Prediction Project (WEPP) hydrologic model to simulate stream permanence in gauged and ungauged watersheds.

Analysis of NHD stream permanence disagreements with other stream permanence observations indicates that differences in climate conditions between the years observations were made contributes to the probability of disagreement. This finding supports the need for dynamic simulation of stream permanence based on climate conditions. Dynamic stream permanence estimates for NHD headwater reaches are then calculated using the MWBM, which incorporates climate inputs and hydrological processes to simulate runoff and streamflow permanence. Results from application of the MWBM indicate that more stream permanence data are needed to properly assess process-based models at regional extents. On 40% of headwater streams no MWBM parameter sets were greater than 65% accurate when compared to observations. However, on other headwaters streams MWBM simulations did show some promise for simulating stream permanence. Additional data would encourage and inform further model development. WEPP simulations produced daily stream permanence accuracies up to 93% and annual stream permanence accuracies up to 87% but different parameter sets performed better for daily and annual time steps. These results indicate that, when implemented for stream permanence simulation, assessment of physically-based models should include both daily and annual accuracies. Additionally, future stream permanence data collection methods and methodologies should be strategically planned to capture the spatiotemporal dynamics of stream permanence that are required to effectively develop and evaluate models.

## **Acknowledgements**

This work would not have been possible without support from a number of individuals and organizations. My dissertation committee of Paul Gessler, Kyle Blasch, Erin Brooks, Tim Link provided exquisite mentorship, training, and support through the entirety of my degree-seeking activities. USGS collaborators Al Rea, Jason Dunham, Roy Sando, Kris Jaeger willingly provided their data, time, and expertise to connect with important resources, projects, groups, funding opportunities, and helped integrate me into new scientific communities. Results from these studies would not have been possible without data acquisition, technical support, and local information supplied by Andy Gendaszek, Justin Thorston, Nate Chelgren, Emily Heaston. Sherri Johnson, Steve Wondzell, Chris Konrad, Keven Roth, and Mariana Dobre. Adrienne Marshall, Abby Lute, Matt Dunkle and other IGERT students contributed many hours to brainstorming, troubleshooting code, providing constructive feedback, and generating lasting friendships. Thank you all!

This research was funded by the National Science Foundation Integrative Graduate Education and Research Training (IGERT) Program (Award Number 1249400), the USGS eco drought project, and the USGS Northwest Climate Adaptation Science Center. Stream discharge data were provided by the HJ Andrews Experimental Forest research program funded by the US Forest Service Pacific Northwest Research Station, the National Science Foundation's Long-Term Ecological Research Program (DEB 2025755), and Oregon State University.

### **Dedication**

Dedicated to my family - Chalese, Hans, and Kaci – for sticking it out with me.

## Table of Contents

Authorization to Submit Dissertation .....	ii
Abstract .....	iii
Acknowledgements .....	iv
Dedication .....	v
Table of Contents .....	vi
List of Tables .....	viii
List of Figures .....	x
Statement of Contribution .....	xiv
Chapter 1: Introduction .....	1
Literature Cited.....	4
Chapter 2: The Influence of Climate Variability on the Accuracy of NHD Perennial and Non-perennial Stream Classifications .....	6
Abstract .....	6
Introduction .....	6
Methods and Materials .....	9
Results .....	17
Discussion .....	24
Literature Cited.....	28
Chapter 3: Precision of headwater stream permanence estimates from a monthly water balance model in the Pacific Northwest, USA .....	31
Abstract .....	31
Introduction .....	31
Methods .....	34
Results .....	42
Discussion .....	52
Conclusion.....	55

Literature Cited.....	56
Chapter 4: Estimating Stream Permanence with the Watershed Erosion Prediction Project Model: Implications for Surface Water Presence Modeling and Data Collection.....	60
Abstract .....	60
Introduction .....	60
Methods .....	63
Results .....	72
Discussion .....	86
Literature Cited.....	89
Chapter 5: Conclusion .....	92
Literature Cited.....	95

## List of Tables

Table 2.1. Number of ‘wet’ and ‘dry’ non-NHD in situ streamflow observations (NNO) for each high-resolution National Hydrography Dataset (NHD-HR) stream permanence classification (SPC). .....	12
Table 2.2. Parameter estimates for a logistic regression model to estimate the probability that a streamflow observation disagreed with the High Resolution National Hydrography Dataset (NHD-HR) stream permanence classification (perennial or non-perennial) for a given reach. Model parameters are observation type (OT; wet or dry), difference in Palmer Drought Severity Index (dPDSI) between the year of an observation and the year of the NHD topographic field survey for a reach separated by positive ( $dPDSI < 0$ ) and negative ( $dPDSI \geq 0$ ) values, and Strahler stream order (SO; 1-7).....	23
Table 3.1. Parameters assessed in sensitivity analysis and calibration of the monthly water balance model. ....	36
Table 3.2. Number of wet and dry observations used to calibrate the monthly water balance model for each calibration group. ....	44
Table 3.3. Number and accuracy range of parameter sets for each calibration region that met accuracy constraints for annual and monthly simulations. ....	48
Table 4.1. WEPP parameters and their samples ranges for HJA and WW watersheds in comparison to the default WEPPCloud parameters. Parameter abbreviations are as follows: baseflow coefficient=KB, deep seepage coefficient=KS, vertical conductivity of the restrictive layer=KS, and crop coefficient=KC. ....	70
Table 4.2. WEPP parameters determined by calibration to observed streamflow for ach HJA watershed. The default WEPP parameters are also shown for reference. KC=crop coefficient, KR=vertical conductivity of the restrictive layer, KS=deep seepage coefficient, and KB=baseflow recession coefficient. ....	74
Table 4.3. Daily and annual accuracy of the WEPP parameter sets with the highest adjusted accuracy value for each WW watershed. WAcc and DAcc describe the modeled accuracy when compared to wet and dry surface water presence observations, respectively. PAcc and NPAcc describe annual modeled accuracy when compared to permanent and nonpermanent locations, respectively. KC is the crop coefficient parameter and KB the baseflow recession coefficient parameter.....	79



Table 4.4. Annual WEPP accuracy (Acc) in each WW watershed (WS) with and without a dry-day threshold, and the corresponding values for crop coefficient (KC) and baseflow recession coefficient (KB).....	82
---	----

## List of Figures

Figure 2.1. The spatial extent of the Pacific Northwest study area showing A) stream permanence classifications for the High Resolution version of the National Hydrography Dataset (NHD-HR) and B) non-NHD streamflow observations made between 1977 and 2016. Gray, internal borders show state boundaries. ....	13
Figure 2.2. Non-National Hydrography Dataset streamflow observations (NNO) in the Pacific Northwest by month. A wet classification indicates surface water was present in the channel, a dry classification indicates surface water was not present in the stream channel.....	14
Figure 2.3. A) Topographic field survey year, B) mean self-calibrated Palmer Drought Severity Index (PDSI), and C) precipitation percentile for 1:24,000-scale (7.5-minute quadrangle) maps in the conterminous United States. Topographic field survey year is the year in which stream permanence classifications ( <i>e.g.</i> perennial, intermittent, ephemeral) were made by survey crews.....	18
Figure 2.4. Agreement and disagreement between the high resolution and medium resolution versions of the National Hydrography Dataset (NHD-HR and NHD-MR, respectively) and non-NHD streamflow observations (NNO) in the Pacific Northwest during the years 1977-2016. ....	19
Figure 2.5. Disagreement between non-NHD streamflow observations (NNO) and High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classifications in the Pacific Northwest grouped by NHD-HR stream order. ....	20
Figure 2.6. A) Disagreement between non-NHD streamflow observations (NNO) and High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classifications and B) the difference in the self-calibrated Palmer Drought Severity Index (PDSI) between the year the NHD stream permanence classification was made ( <i>i.e.</i> the year of the topographic field survey for a 7.5-minute USGS quadrangle map) and the year the NNO was collected. ....	21
Figure 2.7. Probability that a non-NHD streamflow observation (NNO) disagreed with a High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classification in the Pacific Northwest. Results shown are from a logistic regression model that includes covariates of NNO type (wet or dry) and the difference in Palmer Drought Severity Index (PDSI) between the year the NHD-HR stream permanence classification was made ( <i>i.e.</i> the year of the topographic field survey for a 7.5-minute USGS quadrangle map) and the year the NNO was made. Red and blue open	

circles show disagreement/agreement of individual NNO with NHD-HR. Red and blue lines show the modeled (logistic regression) probability of disagreement for dry and wet NNO, respectively. .... 24

Figure 3.1. Diagram of the monthly water balance model. Bold parameter abbreviations correspond to the parameter descriptions and ranges in Table 1. Adapted from McCabe and Markstrom, 2007. .... 36

Figure 3.2. Location of *in situ* observations of surface water presence or absence on headwater streams in the Pacific Northwest from 1977-2019. .... 37

Figure 3.3. Catchments and streams that show the geospatial fabric to which the monthly water balance model was applied. The grid shows the extent of individual PRISM grid cells (~4 km)..... 39

Figure 3.4. Spatial distribution of stream permanence classification (permanent or nonpermanent) to the monthly water balance model parameters presented in Table 1..... 43

Figure 3.5. Sensitivity of streamflow estimates to each parameter in Table 1 for each calibration group..... 45

Figure 3.6. Spatial occurrence of stream reaches in each calibration group as determined by the parameter regionalization methods..... 46

Figure 3.7. Distributions of drainage area, maximum elevation, and total annual precipitation for headwater catchments in each calibration group..... 47

Figure 3.8. Distributions of parameter values for parameter sets that met benchmark constraints. The x-axis represents the sample range of parameter values (Table 1), except for FT, which is subset to 0.0-2.0 L/s to represent the full range of suitable values determined during calibration. .... 49

Figure 3.9. Model precision estimates for headwater streams in the Pacific Northwest during a drier than normal year (1987), approximately normal year (1990), wetter than normal year (1997), and the average for the modeled period (1977-2019). .... 50

Figure 3.10. Estimated model precision for each calibration group during each year of the study period, where n is in the number of suitable parameter sets tested for each calibration group. .... 51

Figure 3.11. Estimated model precision summed for the entire study area (PNW) for each year of the study. The black line is the total percentage of headwater streams with greater than 90% precision. That is, the sum of the two darkest red and darkest blue areas on the plot. .... 52

Figure 4.1. Watersheds in the H. J. Andrews Experimental forest where data collection and modeling occurred. ....	64
Figure 4.2. Watersheds in the Willow-Whitehorse basin where thermistors were deployed and modeling occurred. ....	65
Figure 4.3. Comparison of observed (blue) and simulated (dashed orange) streamflow in HJA watersheds. Percent bias (PB), Nash Sutcliffe Efficiency (NSE), and NSE calculated on the natural log of discharge values (log Q) represent model fits over the 2000-2019 water years. ....	73
Figure 4.4. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each HJA watershed with default WEPP parameters where $n$ is the number of modeled stream reaches in each watershed. ....	75
Figure 4.5. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each HJA watershed with WEPP parameters calibrated to streamflow where $n$ is the number of stream reaches modeled in each watershed. ....	76
Figure 4.6. Daily accuracy of modeled (WEPP) surface water presence with different parameter sets when compared with observed surface water presence at each WW watershed where $n$ is the sum of observed days at all thermistor sites in a watershed. ....	78
Figure 4.7. Annual analysis of WEPP accuracy for all tested parameter sets in WW watersheds where $n$ is the number of years with observed data summed across all thermistor locations in each watershed. ....	80
Figure 4.8. Annual accuracy values for the best WEPP parameter set for each dry-day threshold value in each WW watershed. ....	81
Figure 4.9. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW01. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map). ....	83
Figure 4.10. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW02. Numbers inside the grid show the	

corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map). ..... 84

Figure 4.11. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW03. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map). ..... 85

Figure 4.12. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW04. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map). ..... 86

## Statement of Contribution

### Chapter 2 Author Contributions

Chapter 2 has already been published with the following coauthors contributing as designated below.

**Konrad Hafen:** Conceptualization; data curation; formal analysis; methodology; project administration; writing-original draft. **Kyle Blasch:** Data curation; funding acquisition; resources; supervision; writing-review & editing. **Roy Sando:** Data curation; methodology; resources; supervision; writing-review & editing. **Al Rea:** Conceptualization; data curation; funding acquisition; project administration; resources; supervision; writing-review & editing. **Paul Gessler:** Methodology; resources; supervision; writing-review & editing.

## Chapter 1: Introduction

Stream permanence, the likelihood a stream will maintain surface water throughout a given year, has long been identified as an important hydrological and ecological characteristic (Datry et al., 2014, 2016; Vander Vorste et al., 2020). As the United States Geological Survey (USGS) began a campaign to comprehensively map the landscapes and features of the contiguous United States (CONUS) in the early 1900s, identifying and mapping permanent (maintain surface water throughout a year) and nonpermanent (periodically cease to maintain surface water) streams was one of the explicit goals (Beaman, 1928). While much research has focused on quantifying and simulating the quantity of water in streams, less attention has been granted to identifying the presence and duration of surface water in streams. And while water quantity studies provide important information for identifying and responding to water surpluses and shortages, policy and regulation of streams is based on the presence of surface water. Indeed, streams subject to regulation under US Clean Water Act (1972) are determined based on the duration of surface water in a stream channel during a normal year (US Army Corps of Engineers and US Environmental Protection Agency, 2020).

The most comprehensive resource for identifying stream permanence in the CONUS is the National Hydrography Dataset (NHD), a line network describing the location and stream permanence classification (perennial, intermittent, or ephemeral, where perennial streams are permanent and intermittent and ephemeral are nonpermanent) for most streams in the US. The NHD is derived from the USGS map surveys that began in the early 1900s (Johnston et al., 2009). Geographic data describing stream location in the NHD network are frequently updated as new topographic and aerial data become available but many NHD stream permanence classifications maintain the values assigned during the initial cartographic survey (Hafen et al., 2020). Unsurprisingly, multiple studies have documented substantial disagreement between NHD stream permanence classifications and more recent stream permanence observations, especially on low-order and headwater streams (Fritz et al., 2013; Nadeau et al., 2015). Disagreements in NHD classifications are expected to arise because the static NHD classifications do not accurately capture surface water extents and duration, which can change year to year based on annual climate conditions (Godsey and Kirchner, 2014). Because of these disagreements the US Environmental Protection Agency and the US Army Corps of Engineers have deemed that NHD stream permanence classifications are not adequate to determine regulator status under the US Clean Water Act ([https://www.epa.gov/sites/production/files/2020-01/documents/nwpr\\_fact\\_sheet\\_-\\_mapping.pdf](https://www.epa.gov/sites/production/files/2020-01/documents/nwpr_fact_sheet_-_mapping.pdf)).

To represent the dynamic nature of stream permanence in space and time (Godsey and Kirchner, 2014), recent efforts have utilized observations of surface water presence and absence to

develop statistical and physically-based models that simulate stream permanence (Gendaszek et al., 2020; Jaeger et al., 2019; Jensen et al., 2018; Sando and Blasch, 2015; Ward et al., 2018; Williamson et al., 2015). Models have been developed at both regional (e.g. Jaeger et al., 2019) and local extents (e.g. Williamson et al., 2015) and model accuracies range from approximately 70% to 90%. However, in comparison to data available from stream gages, there are very few data describing surface water presence (e.g. Jaeger et al., 2019). Most modeling studies implement sparse data over large (103 km<sup>2</sup>) spatial extents (e.g. Jaeger et al., 2019) or record higher resolution data (both temporally and spatially) at small (1 km<sup>2</sup>) extents (Jensen et al., 2018; e.g. Ward et al., 2018; Williamson et al., 2015). Gendaszek et al., (2020) did develop statistical models, with good success, at the 100 km<sup>2</sup> extent based on relatively dense array of sensors, and Ward et al., (2020) a model that was validated at the 1 km<sup>2</sup> extent to a 600 km<sup>2</sup> watershed.

Physically-based models have been widely implements across spatial extents to simulate streamflow magnitude and the initial success of similar models to estimate stream permanence is promising. However, current implementation of physically-based models for stream permanence simulation have only been validated over small spatial extents (e.g. Ward et al., 2018; Williamson et al., 2015). Therefore, performance of these, and similar, models at larger extents is largely unknown. Recent development of stream permanence sampling protocols (Jaeger et al., 2020) and methods to record surface water presence from automated sensors (Arismendi et al., 2017; Blasch et al., 2002) are providing the necessary underlying data to evaluate performance of additional physically-based models at multiple spatial extents. Furthermore, streamflow permanence data are often collected in locations that lack streamflow records (Gendaszek et al., 2020; Jaeger et al., 2019; Sando and Blasch, 2015), which enriches current hydrologic understanding across landscapes. Surface water presence data present new opportunities for model assessment and calibration that may allow model validation and development in regions that previously had too few data to evaluate model results.

The overarching objective of this work is to advance stream permanence modeling by incorporating surface water presence observations into physically-modeling workflows and evaluations. Subsequent chapters in this document meet this objective by establishing the accuracy of current stream permanence mapping products, calibrating and evaluating the precision of a monthly water balance to simulate stream permanence at a regional extent, and evaluating the accuracy of the Watershed Erosion Prediction Project (WEPP) hydrological model to simulate stream permanence when calibrated with traditional streamflow data and surface water presence observations. Each of these studies provides new insight to methods for classifying and simulating stream permanence.



Chapter 2, herein, identifies the effect of annual climate differences on disagreements between NHD stream permanence classifications and surface water presence observations. These findings help establish the importance of modeling efforts to account for the dynamic effects of climate on stream permanence. Additionally, Chapter 2 establishes the accuracy of NHD stream permanence observations in the Pacific Northwest Region (PNW) of the United States when compared to *in situ* surface water presence observations. This provides an important benchmark value to which model results can be compared.

Chapter 3 applies a widely implemented monthly water balance model (MWBM) to the PNW. Instead of calibrating the MWBM after the traditional manner using streamflow observations, parameter sets are identified based on their accuracy when compared to surface water presence observations. This study demonstrates how surface water presence observations may be used to inform process-based models at regional extents.

Chapter 4 presents two case studies of simulating stream permanence with the WEPP model in arid and humid watersheds using both streamflow observations and surface water presence observations for calibration and validation. These case studies provide examples of how validation with different data types over different time periods can influence accuracy of simulated stream permanence classifications at daily and annual time steps. Additionally, this study demonstrates how a surface water presence time series may be used to calibrate and evaluate performance of a physically based model. Results also point to the spatial and temporal resolution that are needed for surface water presence observations to be most useful in stream permanence modeling.

### Literature Cited

- Arismendi, I., J.B. Dunham, M.P. Heck, L.D. Schultz, and D. Hockman-Wert, 2017. A Statistical Method to Predict Flow Permanence in Dryland Streams from Time Series of Stream Temperature. *Water* 9:946.
- Beaman, W.M., 1928. Topographic Instructions of the United States Geological Survey. Topographic Mapping. U.S. Geological Survey. doi:10.3133/b788E.
- Blasch, K.W., T.P. Ferré, A.H. Christensen, and J.P. Hoffmann, 2002. New Field Method to Determine Streamflow Timing Using Electrical Resistance Sensors. *Vadose Zone Journal* 1:289–299.
- Datry, T., K. Fritz, and C. Leigh, 2016. Challenges, Developments and Perspectives in Intermittent River Ecology. *Freshwater Biology* 61:1171–1180.
- Datry, T., S.T. Larned, and K. Tockner, 2014. Intermittent Rivers: A Challenge for Freshwater Ecology. *BioScience* 64:229–235.
- Fritz, K.M., E. Hagenbuch, E. D’Amico, M. Reif, P.J. Wigington, S.G. Leibowitz, R.L. Comeleo, J.L. Ebersole, and T.-L. Nadeau, 2013. Comparing the Extent and Permanence of Headwater Streams From Two Field Surveys to Values From Hydrographic Databases and Maps. *JAWRA Journal of the American Water Resources Association* 49:867–882.
- Gendaszek, A.S., J.B. Dunham, C.E. Torgersen, D.P. Hockman-Wert, M.P. Heck, J. Thorson, J. Mintz, and T. Allai, 2020. Land-Cover and Climatic Controls on Water Temperature, Flow Permanence, and Fragmentation of Great Basin Stream Networks. *Water* 12:1962.
- Godsey, S.E. and J.W. Kirchner, 2014. Dynamic, Discontinuous Stream Networks: Hydrologically Driven Variations in Active Drainage Density, Flowing Channels and Stream Order. *Hydrological Processes* 28:5791–5803.
- Hafen, K.C., K.W. Blasch, A. Rea, R. Sando, and P.E. Gessler, 2020. The Influence of Climate Variability on the Accuracy of NHD Perennial and Nonperennial Stream Classifications. *JAWRA Journal of the American Water Resources Association* 56:903–916.
- Jaeger, K.L., J. Burnett, E.D. Heaston, S.M. Wondzell, N. Chelgren, J.B. Dunham, S. Johnson, and M. Brown, 2020. FLOWPER User Guide—For Collection of FLOW PERmanence Field Observations. Open-File Report, US Geological Survey. <https://pubs.er.usgs.gov/publication/ofr20201075>.
- Jaeger, K.L., R. Sando, R.R. McShane, J.B. Dunham, D.P. Hockman-Wert, K.E. Kaiser, K. Hafen, J.C. Risley, and K.W. Blasch, 2019. Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest. *Journal of Hydrology* X 2:100005.
- Jensen, C.K., K.J. McGuire, Y. Shao, and C. Andrew Dolloff, 2018. Modeling Wet Headwater Stream Networks across Multiple Flow Conditions in the Appalachian Highlands. *Earth Surface Processes and Landforms* 43:2762–2778.
- Johnston, C.M., T.G. Dewald, T.R. Bondelid, B.B. Worstell, L.D. McKay, A. Rea, R.B. Moore, and J.L. Goodall, 2009. Evaluation of Catchment Delineation Methods for the Medium-Resolution National Hydrography Dataset. USGS Numbered Series, U.S. Geological Survey, Reston, VA. <http://pubs.er.usgs.gov/publication/sir20095233>. Accessed 31 Oct 2018.

- Nadeau, T.-L., G.S. Leibowitz, J.P. Wigington, L.J. Ebersole, M.K. Fritz, A.R. Coulombe, L.R. Comeleo, and A.K. Blocksom, 2015. Validation of Rapid Assessment Methods to Determine Streamflow Duration Classes in the Pacific Northwest, USA. *Environmental Management* 56:34–53.
- Sando, R. and W.K. Blasch, 2015. Predicting Alpine Headwater Stream Intermittency: A Case Study in the Northern Rocky Mountains. *Ecohydrology & Hydrobiology* 15. <https://linkinghub.elsevier.com/retrieve/pii/S1642359315000154>.
- US Army Corps of Engineers and US Environmental Protection Agency, 2020. The Navigable Waters Protection Rule: Definition of “Waters of the United States.” *Federal Register* 85 FR 22250:22250–22342.
- Vander Vorste, R., R. Sarremejane, and T. Datry, 2020. Intermittent Rivers and Ephemeral Streams: A Unique Biome With Important Contributions to Biodiversity and Ecosystem Services.
- Ward, A.S., N.M. Schmadel, and S.M. Wondzell, 2018. Simulation of Dynamic Expansion, Contraction, and Connectivity in a Mountain Stream Network. *Advances in Water Resources* 114:64–82.
- Ward, A.S., S.M. Wondzell, N.M. Schmadel, and S.P. Herzog, 2020. Climate Change Causes River Network Contraction and Disconnection in the HJ Andrews Experimental Forest, Oregon, USA. *Front. Water* 2: 7. Doi: 10.3389/Frwa.
- Williamson, T.N., C.T. Agouridis, C.D. Barton, J.A. Villines, and J.G. Lant, 2015. Classification of Ephemeral, Intermittent, and Perennial Stream Reaches Using a TOPMODEL-Based Approach. *JAWRA Journal of the American Water Resources Association* 51:1739–1759.

## **Chapter 2: The Influence of Climate Variability on the Accuracy of NHD Perennial and Non-perennial Stream Classifications**

Hafen, KC, KW Blasch, A Rea, R Sando, and PE Gessler. 2020. The Influence of Climate Variability on the Accuracy of NHD Perennial and Nonperennial Stream Classifications. *Journal of the American Water Resources Association (JAWRA)* 56:903-916. <https://doi.org/10.1111/1752-1688.12871>

### **Abstract**

National Hydrography Dataset (NHD) stream permanence classifications (SPC; perennial, intermittent, and ephemeral) are widely used for data visualization and applied science, and have implications for resource policy and management. NHD SPC were assigned using a combination of topographic field surveys and interviews with local residents. However, previous studies indicate that field-based, non-NHD streamflow observations (NNO) frequently disagree with NHD SPC. We hypothesized that differences in annual climate conditions between map creation years and the years NNO were collected contributed to disagreement between NNO and NHD SPC. We compared NHD SPC to 10,055 NNO (classified as ‘wet’ or ‘dry’) collected in the Pacific Northwest (PNW) between 1977 and 2015. Annual climate conditions were described with the Palmer Drought Severity Index (PDSI). Stream order was added as a covariate to account for different effects along the stream network. NHD SPC agreed with 80.5% of NNO. ‘Dry’ NNO were 5 times more likely to disagree with NHD than ‘wet’ NNO ( $p < 0.0001$ ). Disagreement was greatest on first order streams. When NHD SPC were collected during a wetter period than NNO the probability of disagreement increased by a factor of 1.17 ( $p < 0.0001$ ) per unit difference in PDSI. The influence of climate on disagreements between NNO and NHD SPC provides support for the continued development of dynamic models representing SPC as opposed to static NHD classifications.

### **Introduction**

Distinction between streams that flow continuously (perennial), and those that periodically cease to flow (intermittent and ephemeral), is important to understand the cumulative ecology and hydrology of stream networks and the potential impacts of climate fluctuations and anomalies on stream reaches and watersheds. Stream permanence (i.e. the likelihood a stream will remain flowing throughout a given year) is often spatially (e.g. streams may have both permanent and non-permanent sections) and temporally variable, yet it is most often represented by static maps that classify stream reaches as either perennial or non-perennial (i.e. intermittent or ephemeral). For the United States (US), the National Hydrography Dataset (NHD; U.S. Geological Survey, National Hydrography

Dataset. Accessed October 1, 2019, <https://www.usgs.gov/nathydro>) provides the most comprehensive classification of perennial, intermittent, and ephemeral streams (Nadeau and Rains, 2007). NHD is used extensively for cartography, study design, monitoring, and modeling at local, regional, and national extents (Hill et al., 2016; Isaak et al., 2016; Macfarlane et al., 2017; National Rivers and Streams Assessment, 2008-2009; Sheng et al., 2007; Stanislawski, 2009; Stoddard et al., 2005), and such applications frequently use the NHD stream permanence classifications (i.e. NHD “hydrographic categories” of perennial, intermittent, and ephemeral) (Benstead and Leigh, 2012; Roy et al., 2009). Comparison of NHD stream permanence classifications (hereafter referred to as SPC) to non-NHD, in-situ streamflow observations (hereafter referred to as NNO) indicate that misclassification of stream class by the NHD can be an issue, particularly for headwater streams (Caruso and Haynes, 2011; Fritz et al., 2013; Nadeau et al., 2015; Nadeau and Rains, 2007). This classification inaccuracy can place limitations on studies and decisions that leverage the NHD. Inaccuracies deriving from cartographic methodologies associated with the NHD have been well documented (Caruso and Haynes, 2011; Chorley and Dale, 1972; Leopold, 1994; Nadeau and Rains, 2007), but the influence of annual climate fluctuation on NHD SPC has received less attention (Nadeau et al., 2015). We assess how differences in annual climate conditions when NHD SPC were developed and annual climate conditions when NNO were collected may influence the accuracy of NHD SPC.

NHD is a vector line network of streams in the contiguous United States (CONUS), Alaska, Hawaii, Guam, American Samoa, Puerto Rico, the US Virgin Islands, and the Northern Mariana Islands. Medium- and high-resolution versions of the NHD (NHD-MR and NHD-HR, respectively) are available. NHD-HR stream segments were derived from streamlines on US Geological Survey (USGS) 7.5-minute quadrangle (1:24,000-scale) maps (7.5-min maps), which cover most of the CONUS. Seven and one half minute maps were surveyed by field crews between 1881 and 2000 with the majority of the surveys conducted between 1955 and 1990. Streamlines on 7.5-min maps were created from topographic field surveys, aerial imagery, or a combination of both (Beaman, 1928; Drummond, 1974; Fritz et al., 2013; Mark, 1983, p. 83). NHD-MR stream segments were derived from USGS 1:100,000 scale maps, most of which were developed from the 7.5-min maps (Johnston et al., 2009). However, at the time the 1:100,000-scale maps were developed, 7.5-min maps did not exist for the entire CONUS. To fill in these potential data gaps, 1:62,500-scale maps that predated the 7.5-min maps were also used to create NHD-MR (Johnston et al., 2009). Each NHD-HR streamline was assigned a SPC of perennial, intermittent, ephemeral, or designated as an artificial path (flowlines through features represented by polygons such as large rivers, reservoirs, and wetlands).

SPC for 7.5-min maps were assigned during topographic field surveys (usually conducted between April 15 and November 15), or interviews with local residents (Personal communication from Keven Roth, USGS National Mapping Division, to K. Hafen, November 14, 2018), between 1881 and 2000 (Beaman, 1928). Perennial streams were defined as containing “water throughout the year, except for infrequent periods of severe drought”, intermittent streams as containing “water for only part of the year, but more than just after rainstorms and at snowmelt” and ephemeral streams as containing “water only in direct response to precipitation” (Guptill, 1990; NHD User’s Guide. Accessed February 13, 2020, <https://nhd.usgs.gov/userguide.html> ). SPC designations made by survey crews were intended to represent stream permanence conditions across a normal range of climatic conditions, not just the conditions observed during the survey year (Personal communication from Keven Roth, USGS National Mapping Division, to K. Hafen, November 14, 2018 and from Stephen Aichele, USGS National Geospatial Program, to K. Hafen, December 6, 2020). Locations of NHD streamlines were sometimes revised where aerial imagery was available, but these updates could only be applied to streams visible from above, which prevented updates for many headwater streams where riparian vegetation masked the channel or where the image resolution was too coarse to identify the presence or absence of water. When updates were made, it was rare to change SPC because the temporal resolution of aerial images was generally not adequate to provide more information than the original field survey (Personal communication from Keven Roth, USGS National Mapping Division, to A. Rea, March 30, 2020). Thus, for most small, headwater streams, NHD SPC represent conditions during the year (or recent past in the case of resident interviews) of the topographic field survey, or resident interview, and are a static classification based on observation during a specific time period. Topographic field surveys of 7.5-min maps spanned several decades; therefore, the annual climate conditions during the field survey may have affected the NHD SPC accuracy.

Multiple studies have observed that stream permanence is a dynamic process attributed to changes in land use, wildfire, and climate fluctuations, especially in headwater streams (Fritz et al., 2013; Godsey and Kirchner, 2014; Sando and Blasch, 2015). Static hydrology data products, such as NHD, do not always accurately represent observed variability (Fritz et al., 2013). The accuracy of NHD SPC can be highly variable, especially for headwater streams (Caruso, 2014; Colson et al., 2008; Nadeau et al., 2015; O’Connor et al., 2014; Stoddard et al., 2005; Wood et al., 2009). Specifically, NHD-MR is likely to overestimate stream permanence in arid areas (Nadeau et al., 2015) and NHD-HR SPC disagreed with NNO for ~50% of observed headwater streams in select areas of the Pacific Northwest (PNW), USA (Fritz et al., 2013; Nadeau et al., 2015). SPC accuracy differs considerably between some adjacent 7.5-min maps (Colson et al., 2008) with differences attributed to subjective cartographic methods, which resulted in variability between cartographic

analysts (Caruso and Haynes, 2011; Fritz et al., 2013; Nadeau et al., 2015). Differences could also be attributed to climatic fluctuations including variable precipitation and temperature conditions (i.e. drought conditions) (Cooper et al., 2018) during the years adjacent 7.5-min maps were field surveyed because of the broad time period over which topographic field surveys were conducted (e.g. for two maps that share a boundary one may have been surveyed during a wet year and the other during a dry year). In the PNW, NHD-HR accuracy differs spatially based on climate characteristics (Nadeau et al., 2015). However, the effect of conducting 7.5-min map topographic field surveys (i.e. making SPC) and collecting NNO under different annual climate conditions, potentially decades apart, has not been considered. Given the dynamic nature of stream permanence, differences in annual precipitation and temperature at the time of these observations may explain some of the observed differences between NHD SPC and NNO.

Our objective is to demonstrate that the broad time frame over which 7.5-min maps were field surveyed has effects on the accuracy of NHD SPC, and that effects are related to the climatic conditions under which topographic field surveys were conducted. Specifically, we hypothesize that disagreements between NHD-HR SPC and NNO are related to differences in annual climatic conditions between the years when maps were surveyed and the years NNO were made. We quantify the effect of annual climatic differences between 7.5-min map validation years and a large (10,055 observations) database of wet/dry NNO in the PNW (McShane et al., 2016). Along with other studies (Caruso, 2014; Caruso and Haynes, 2011; Chorley and Dale, 1972; Colson et al., 2008; Leopold, 1994; Nadeau et al., 2015; Nadeau and Rains, 2007), our results help establish the limitations of static SPC and support inclusion of climatic variables to develop statistical and process-based models that dynamically estimate stream permanence (Jaeger et al., 2019; Williamson et al., 2015).

## **Methods and Materials**

### *Study Area*

We examine two spatial extents in this study. First, at the extent of the CONUS, we show patterns in the spatial distribution of annual climate conditions and years during which topographic field surveys were conducted for 7.5-min maps. Second, at the extent of the PNW, an existing dataset of NNO (McShane et al., 2016) is used to assess the effect of annual climate conditions on the accuracy of NHD SPC.

The PNW region, as defined by this study, is concurrent with NHD Hydrologic Unit Code (HUC) 17, and includes the Columbia River, its major tributaries, and coastal watersheds in Oregon and Washington (Figure 2.1). This region spans the majority of Idaho, Oregon, and Washington and includes smaller portions of California, Montana, Nevada, Utah, and Wyoming near the boundary

perimeter. Elevations range from sea level along coastlines of the Pacific Ocean to over 4000 m (in the Cascade Range). Annual precipitation ranges from 200 mm in arid regions east of the Cascade Mountains to over 5000 mm in the Oregon Coast Range and Olympic Peninsula (Jaeger et al., 2019). The region includes arid and semi-arid uplands, snow-dominated alpine zones, and temperate rainforests (Leibowitz et al., 2016). Physiographic and ecosystem diversity makes the PNW a good area to test the accuracy of NHD and the effect of annual climate on NHD classifications over a wide range of environments.

## *Data*

### 7.5 Minute Quadrangle Maps

Extents of each 7.5-min map were retrieved for the CONUS from the US Topo and Historical Map Collection (“USGS US Topo and Historical Topographic Map Collection”). In this dataset, the topographic field survey year was associated with the extent of each 7.5-min map. Topographic field survey year was used to identify the year NHD-HR SPC were made for stream segments within the boundaries of each 7.5-min map. Maps that did not have a value for topographic survey year were excluded from analyses.

### National Hydrography Dataset

Both NHD-MR and NHD-HR were used for different analyses at the spatial extent of the PNW. NHD SPC were divided into two groups: perennial (perennial and artificial path classifications) and non-perennial (intermittent and ephemeral) classifications (Figure 2.1). Artificial paths represent flowlines through large rivers, reservoirs, ponds, wetlands, etc., and such features may be either perennial or non-perennial. Artificial path features were joined to the NHDArea attribute in the NHD database, which provides additional information about the features represented by artificial paths, and artificial paths associated with intermittent or ephemeral features were changed to non-perennial.

### Non-NHD in-situ Streamflow Observations (NNO)

NNO were collected and recorded by multiple agencies throughout headwater streams in the PNW between 1977 and 2016 and a large number of these observations (n=24,316) have been consolidated and merged into a single dataset (McShane et al., 2017). The NNO dataset was obtained by consolidating observations from 11 sources that included data from aquatic habitat surveys, channel mapping, reconnaissance surveys, and modeling studies (Jaeger et al., 2019; McShane et al., 2017). From information provided by the data collectors, McShane et al. (2017) classified each observation as ‘wet’ or ‘dry’ and aligned observations to NHD-MR stream grid pixels with drainage areas greater than 0.09 km<sup>2</sup>. ‘Dry’ points include observations of “dry channel, no flow, intermittent,



and ephemeral” and ‘wet’ points include observations “such as a wet or flowing stream” (McShane et al., 2017). Jaeger et al. (2019) used the results of McShane et al. (2017) to statistically model stream permanence.

We joined each NNO point to the appropriate NHD-MR flowline with the grid identification number (which represents unique, rasterized stream segments) from the NHD-MR stream grids. Points that were not associated with a NHD-MR grid identification number (i.e. points that occurred on stream segments that were not mapped during topographic field surveys) were excluded from the analysis. NNO points were then snapped to the nearest NHD-HR flowline. We manually inspected all NNO points where the nearest NHD-HR flowline was greater than 20 m from the point’s original coordinates (prior to snapping) to ensure points were associated with correct flowlines. Where available, NHD stream names were used for clarification. If the correct flowline for a given point could not be determined the point was excluded from analyses. Snapping points to the NHD flowlines assumes that the NNO coordinates accurately represent the NNO location and that the stream reach represented by the NNO is also represented by the NHD (i.e. the stream reach was excluded or omitted from the NHD). Any NNO that was farther than 100 m from the nearest NHD-HR flowline was excluded from the analysis.

Wet NNO did not necessarily represent perennial streams because wet observations could have been made on a non-perennial stream during periods of precipitation, snowmelt, or an elevated water table. To help identify wet observations that may not accurately represent stream permanence for a given year, we plotted the number of wet and dry observations by month. The number of dry observations increased considerably from July to August, indicating August may be the month when most intermittent streams in the PNW become seasonally dry (Figure 2.2). Across most of the PNW precipitation is generally the lowest during the months of July through September reducing the likelihood of ephemeral and intermittent streamflow events. Thus, we excluded all wet NNO that did not occur in August or September (last two months of a water year), or for which no month was recorded, from our analyses because they may not be accurate indicators of annual stream permanence. All dry observations were included. NNO within the extent of 7.5-min maps for which a topographic survey year was not recorded were also excluded from the analyses. No NHD reach coincided with multiple NNO.

Metadata records for each NHD reach where a NNO occurred were checked to ensure the NHD SPC had not been updated after development of the topographic maps. The Strahler stream order (Strahler, 1952, 1957) of the reach on which each NNO was located was obtained from the NHD-HR attributes. Strahler stream order assigns stream reaches a value 1 to n based on position in

the drainage network. Segments that have no tributaries receive a stream order value of one (first order). A second-order stream is formed at the confluence of two first-order streams, and so on. NNO occurred on stream orders ranging from one to nine. All NNO on stream orders greater than seven were classified as ‘wet’ and did not add any information to determining stream permanence, so they were excluded from the analysis. In all, 10,055 NNO points (2,701 ‘dry’ and 7,354 ‘wet’) were used for analyses in this study (Table 1). Of the NNO, 8220 were associated with perennial NHD-HR reaches and 2240 were associated with non-perennial NHD-HR reaches.

Table 2.1. Number of ‘wet’ and ‘dry’ non-NHD in situ streamflow observations (NNO) for each high-resolution National Hydrography Dataset (NHD-HR) stream permanence classification (SPC).

NNO	NHD SPC	
	Perennial	Non-perennial
Wet	6682	672
Dry	1360	1341

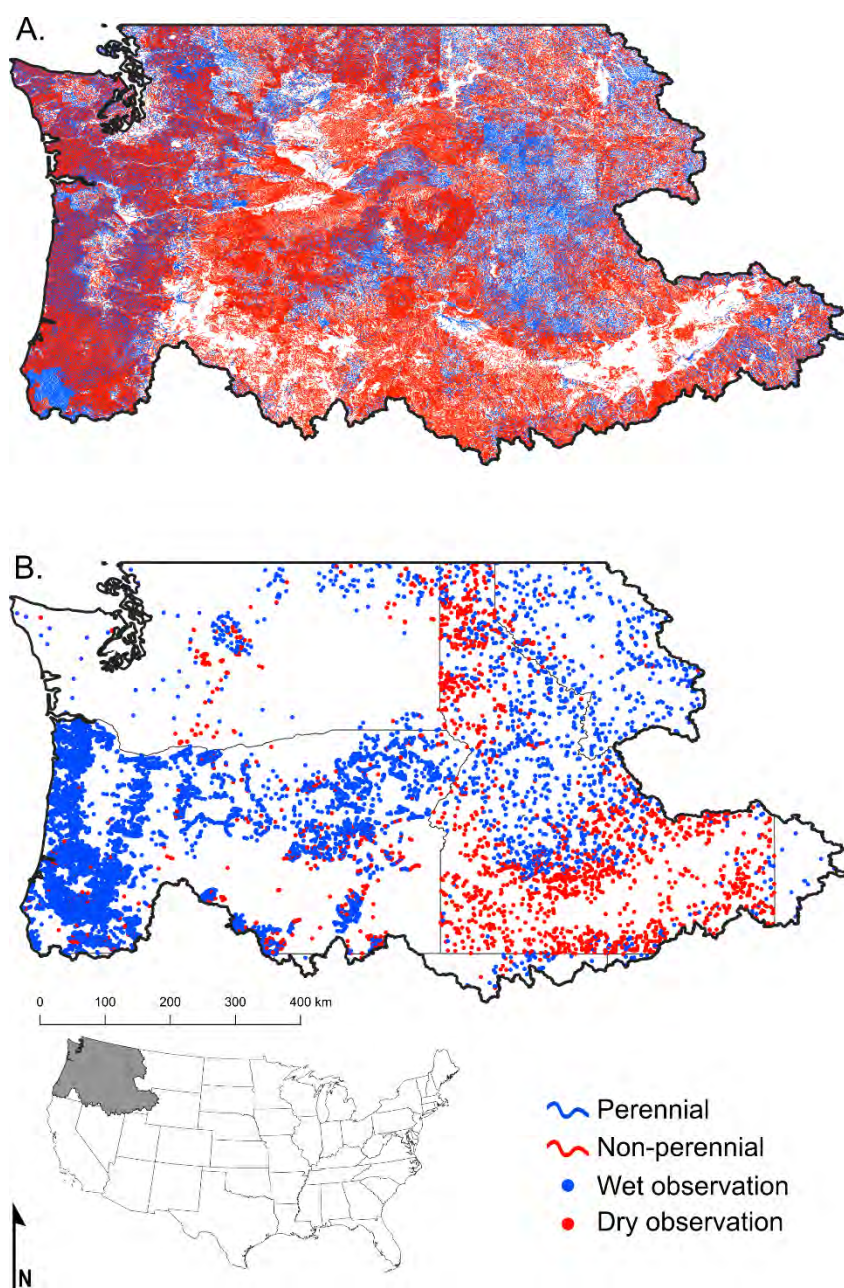


Figure 2.1. The spatial extent of the Pacific Northwest study area showing A) stream permanence classifications for the High Resolution version of the National Hydrography Dataset (NHD-HR) and B) non-NHD streamflow observations made between 1977 and 2016. Gray, internal borders show state boundaries.

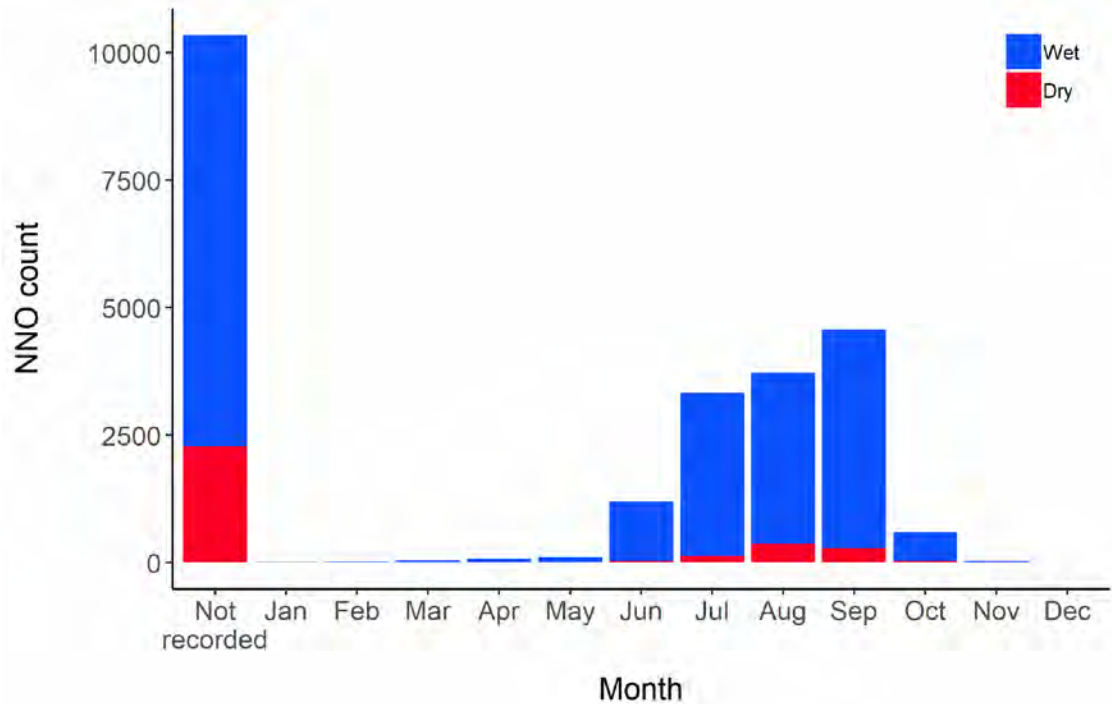


Figure 2.2. Non-National Hydrography Dataset streamflow observations (NNO) in the Pacific Northwest by month. A wet classification indicates surface water was present in the channel, a dry classification indicates surface water was not present in the stream channel.

### Climate Indices

Palmer Drought Severity Index (PDSI) (Wells et al., 2004) and precipitation percentile were used to describe the annual climate conditions for each 7.5-min extent and each NNO.

PDSI is a convenient drought metric because it incorporates precipitation, temperature and evapotranspiration to assess water availability against water demand (i.e. drought) (Palmer, 1965). The self-calibrated version of PDSI (used herein) overcomes issues of spatial standardization with local calibration to make values more comparable (Alley, 1984) and can be implemented at the extent of the CONUS (Wells et al., 2004). Monthly PDSI values were obtained from the Western Regional Climate Center (Accessed October 1, 2019, <https://wrcc.dri.edu/>) as 4 km grids and averaged for each water year (October 1 of the preceding year through September 30 of the current year) from 1896-2016. Averages were calculated by water year because much of the precipitation in the PNW (and many other portions of the CONUS) accumulates as snow in October, November, and December and influences streamflow as it melts the following spring. The PDSI value for each 7.5-min map was obtained by averaging the PDSI grid cells in the map extent for the water year in which the topographic field survey occurred. The PDSI for each NNO was calculated as the PDSI at the observation location for the water year in which it was recorded.

Precipitation was also used as an indicator of climate conditions for 7.5-min map field survey years. Because precipitation varies greatly across the PNW and CONUS we standardized precipitation values as percentiles over the 1896-2016 water years, where a higher percentile indicates a wetter year (e.g. 100% indicates the water year with the most precipitation for the period of record). Precipitation values were acquired from the PRISM Climate Group (“PRISM Climate Group, Oregon State University”). Precipitation percentile for each 7.5-min map was determined by summing total precipitation at each grid cell for each water year (1896-2016) and calculating the percentile at each grid cell for each year from the mean and standard deviation of total water year precipitation. The percentile for each map extent was obtained by averaging the percentile of grid cells within the map extent for the map survey year.

### *Data Analysis*

Data were analyzed at two different spatial extents, the CONUS and the PNW. At the CONUS extent, only data describing 7.5-min maps were used for visual analysis of the climate conditions during topographic field surveys. At the PNW extent, disagreements between NNO and NHD SPC were assessed and evaluated based on climate conditions.

#### Climate during 7.5-minute Quadrangle Map Surveys in the CONUS

We calculated and mapped PDSI and precipitation percentile for the field survey year of each 7.5-min map in the CONUS. Five 7.5-min map extents, surveyed before 1896, were excluded because PDSI and precipitation data were not retrieved prior to the 1896 water year. All five of these maps represented regions of the State of Maine. Spatial and temporal variation in annual climate conditions between map validation years were visually assessed with maps of validation year, PDSI, and precipitation percentile (Figure 2.3). The overall distributions of validation years and annual climate conditions were assessed with histograms (Figure 2.3).

#### Comparison of NHD and NNO in the PNW

We identified disagreements between NHD-MR and NHD-HR SPC and NNO in the Pacific Northwest. NNO and NHD were determined to disagree when a wet NNO collected in August or September occurred on a stream reach classified as non-perennial in NHD, or a dry NNO occurred on a stream reach classified as perennial in NHD. We report the overall percentage of agreement between NNO and NHD-MR and NHD-HR SPC and the percentage of disagreements that occur on perennial and non-perennial NHD reaches. It is important to note that a wet NNO on a non-perennial NHD reach does not guarantee a disagreement, because the NNO represents a single point in time, thus, the stream reach may have been dry before or after the observation was made. With the same reasoning, a wet observation on a perennial reach does not guarantee agreement. Identifying

agreements and disagreements is also complicated by the NHD definition of perennial streams, which allows an exception for, but does not define, ‘years of extreme drought’. For analyses herein, we classify disagreements as defined above, with no exceptions.

#### Effects of Climate Differences on Disagreement Between NHD SPC and NNO in the PNW

The hypothesis that differences in climate conditions between topographic field survey years of 7.5-min maps and NNO years influence disagreement between NNO and NHD SPC was tested with logistic regression using the Generalized Linear Model (glm) function in R statistical software version 3.6.0 (R Core Team, 2019). The probability that a field observation disagreed with the associated NHD SPC was modeled as a function of the observation type (OT; wet or dry), PDSI difference (dPDSI) between the 7.5-min map field survey year and the NNO year, and Strahler stream order (SO). This analysis was only considered for NHD-HR stream classifications because topographic field survey years were only obtained for 7.5-min maps, and some regions of the NHD-MR were based on a different set of maps (as mentioned above). Negative dPDSI values indicate a 7.5-min map was surveyed in a drier year than a NNO and positive dPDSI values indicate the 7.5-min map was surveyed in a wetter year than a NNO. dPDSI values near zero indicate a small difference in climate conditions between 7.5-min map survey year and NNO year. As dPDSI values increase and decrease from zero, the magnitude of PDSI difference between map survey year and NNO year increase. The relation between OT and dPDSI was modeled as an interaction because it was possible for the effect of dPDSI to depend upon whether a NNO was wet or dry. For example, there may be a lower probability of disagreement for wet NNO with positive dPDSI values than for dry NNO with similar dPDSI values.

To allow for potential, nonlinear effects of dPDSI (i.e. different effects when NNO were collected in a wetter year than NHD versus a drier year than NHD), we represented dPDSI as a spline (e.g. Ruprecht et al., 2016) with a knot at zero. A spline model treats dPDSI values less than zero ( $dPDSI < 0$ ) and dPDSI values greater than or equal to zero ( $dPDSI \geq 0$ ) as separate variables, allowing different behavior (i.e. parameter estimates) on either side of zero. For example, this model parameterization allows the probability of disagreement to decrease (negative slope) as dPDSI increases, for values less than zero, and the probability of disagreement to increase as dPDSI increases (positive slope), for values greater than zero. The final logistic regression model took the form

$$O_d = OT + dPDSI_{<0} + dPDSI_{\geq 0} + SO + (OT * dPDSI_{<0}) + (OT * dPDSI_{\geq 0}) + (OT * SO)$$

where  $OT*dPDSI_{<0}$  and  $OT*dPDSI_{\geq 0}$  are the interaction between OT and dPDSI when dPDSI is less than and greater than or equal to zero, respectively, and  $O_d$  is the log (natural) odds of disagreement between NNO and NHD SCP. Using logistic regression, a parameter value was estimated for the effect of each term on the right side of the equation (Table 2). The probability of disagreement ( $P_d$ ) between NNO and NHD SCP can then be calculated as

$$P_d = \frac{\exp(O_d)}{1 + \exp(O_d)}$$

Logistic regression model parameters were assessed for significance by p-value using the Wald test, where  $p < 0.05$  (95% confidence level) indicated a statistically significant effect. Analyses described herein consider if disagreement between streamflow permanence observations and NHD SPC was affected by OT, SO and/or dPDSI, or interactions between these variables.

#### *Data Availability*

Data used for the analyses herein are publicly available from the ScienceBase data catalog at the URL <https://doi.org/10.5066/P9Z6XZP0>.

## **Results**

### *Climate During 7.5-min Map Surveys in the Contiguous United States*

Seven and one half minute map field surveys in the CONUS occurred between 1881 and 2000 and exhibited spatial and temporal climatic trends (Figure 2.3a). In general, 7.5-min maps east of the Mississippi River (except for Maine and New Hampshire) were surveyed in earlier years than western states, and the majority of surveys occurred between 1925 and 1975. Most 7.5-min maps in western states were surveyed between 1965 and 2000 (except for California). The temporal distribution of 7.5-min map topographic field survey years was approximated by a normal distribution, with very few surveys occurring before 1930 and the majority occurring between 1960 and 1980. It is important to note that 7.5-min maps located relatively close together (or even adjacent to each other) may have been surveyed as many as 40 years apart (Figure 2.3). Thus, the spatial proximity of topographic maps does not equate to temporal proximity. PDSI in the CONUS during 7.5-min map field validation year ranged from -6.2 to 7.2 and was also approximated by a normal distribution (Figure 2.3b). In general, topographic field surveys of 7.5-min maps in the southwestern, south-central, east-northeast states were conducted during periods of drought while areas of the northwest, upper-midwest, and southeast states were surveyed during wetter periods. Precipitation percentile ranged from 0.3 to 99.9 and was uniformly distributed (i.e. not normal) across all percentiles (Figure 2.3c). Similar to PDSI, precipitation shows wetter conditions during topographic



field surveys in the northwest and southeast and drier conditions in the east-northeast. Overall, topographic field surveys occurred across a broad spectrum of climatic conditions and multiple 7.5-min maps adjacent or near to each other were surveyed under different annual climate conditions.

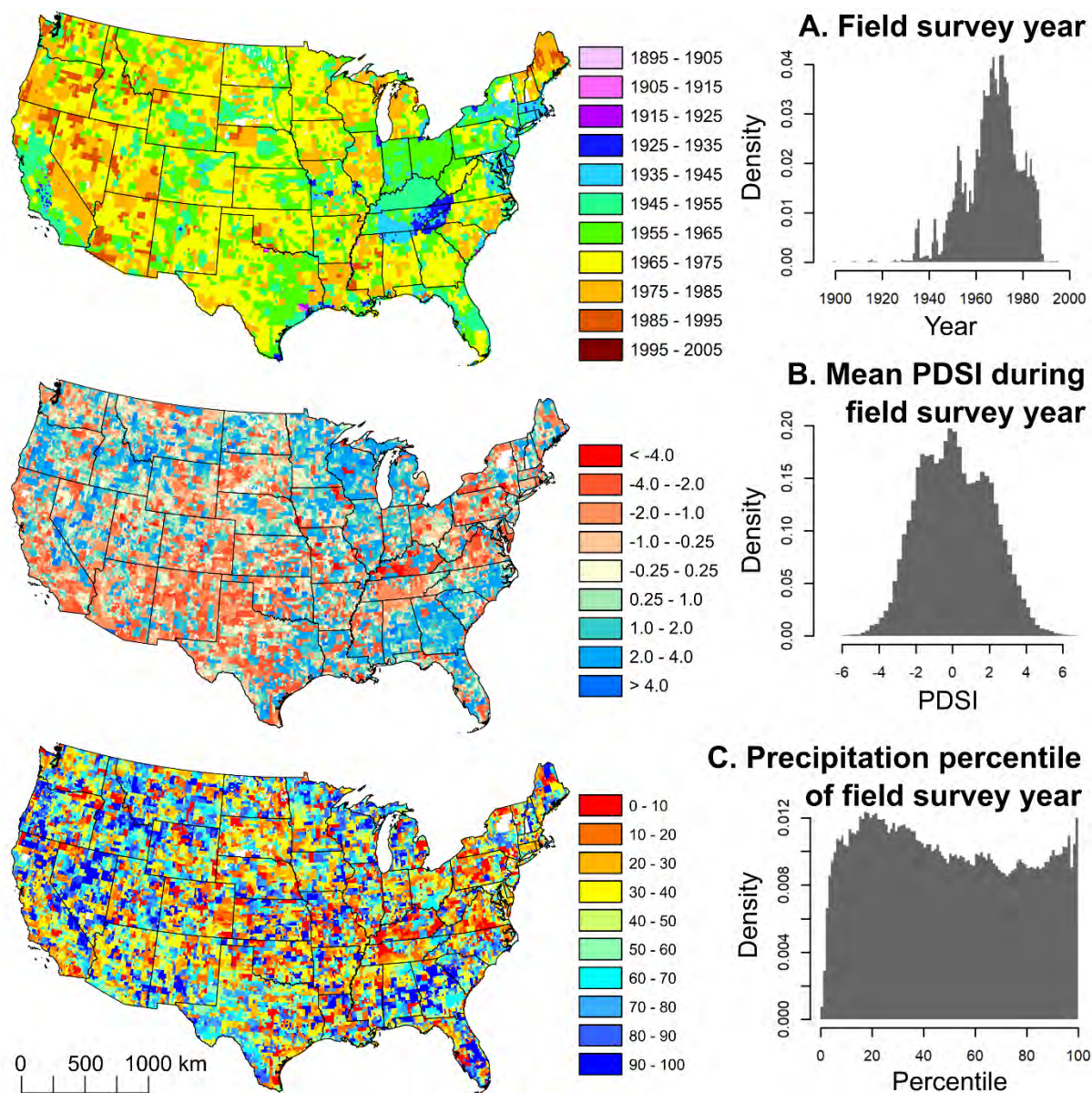


Figure 2.3. A) Topographic field survey year, B) mean self-calibrated Palmer Drought Severity Index (PDSI), and C) precipitation percentile for 1:24,000-scale (7.5-minute quadrangle) maps in the conterminous United States. Topographic field survey year is the year in which stream permanence classifications (*e.g.* perennial, intermittent, ephemeral) were made by survey crews.

#### *Comparison of NHD and Streamflow Observations in the Pacific Northwest*

In the PNW, 80.8% of NHD-HR, and 80.2% of NHD-MR, SPC agreed with NNO. Of the total NNO, 13.8 and 16.2% occurred when a ‘dry’ observation fell on a perennially classified NHD



reach for NHD-HR and NHD-MR, respectively. A ‘wet’ (August or September) observation fell on a non-perennially classified NHD reach for 5.3 and 3.6% of NNO, for NHD-HR and NHD-MR (Figure 2.4). Overall, 33% of NNO on non-perennial streams resulted in disagreements and 17% of NNO on perennial streams resulted in disagreements (Table 1). Disagreements between NHD-HR and NNO decreased as stream order increased (Figure 2.5). Disagreements were spatially distributed throughout the PNW, but some areas exhibited a greater density of disagreements (Figure 2.6). When disagreements were mapped with dPDSI, there was qualitative evidence that many disagreements, occurred in areas with relatively high or low dPDSI values. For example, a cluster of disagreements in northern Idaho appear to correlate with negative dPDSI values and disagreements along the Idaho-Montana border appear correlated with positive dPDSI values (Figure 2.6).

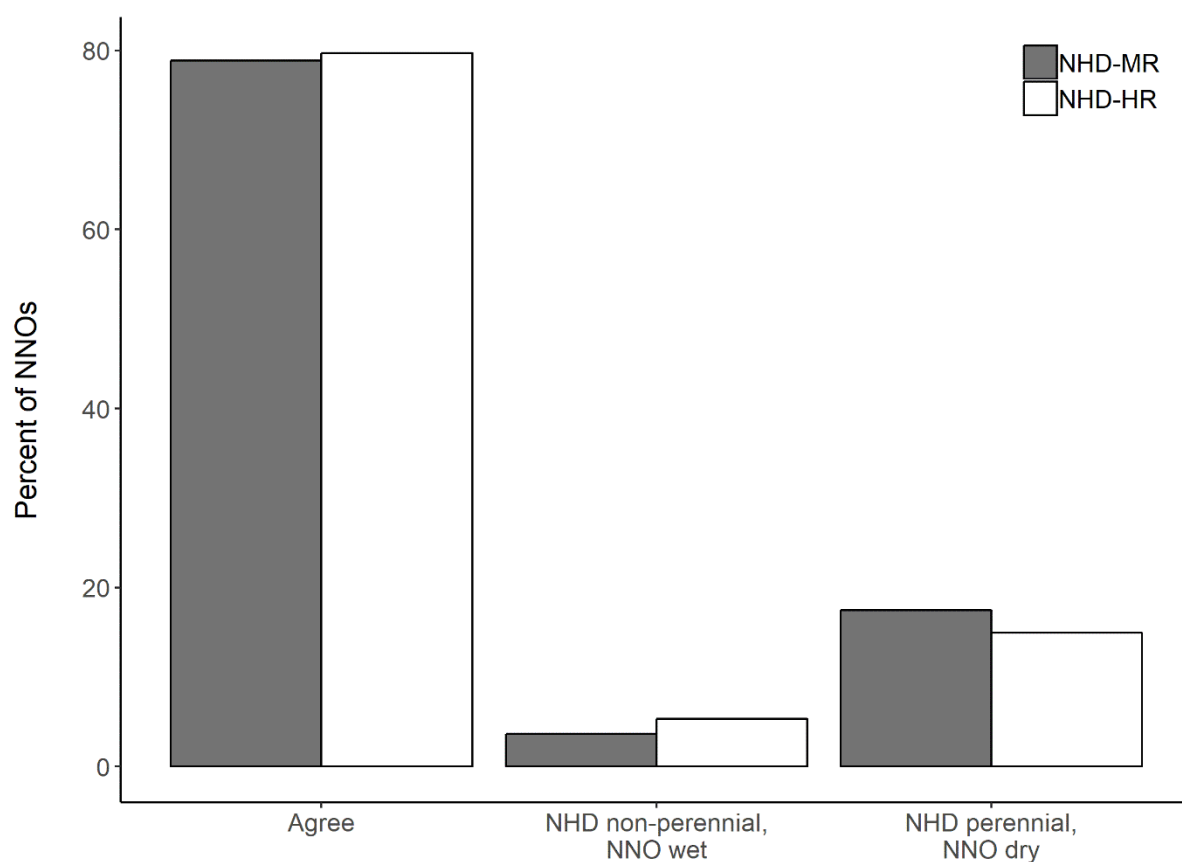


Figure 2.4. Agreement and disagreement between the high resolution and medium resolution versions of the National Hydrography Dataset (NHD-HR and NHD-MR, respectively) and non-NHD streamflow observations (NNO) in the Pacific Northwest during the years 1977-2016.

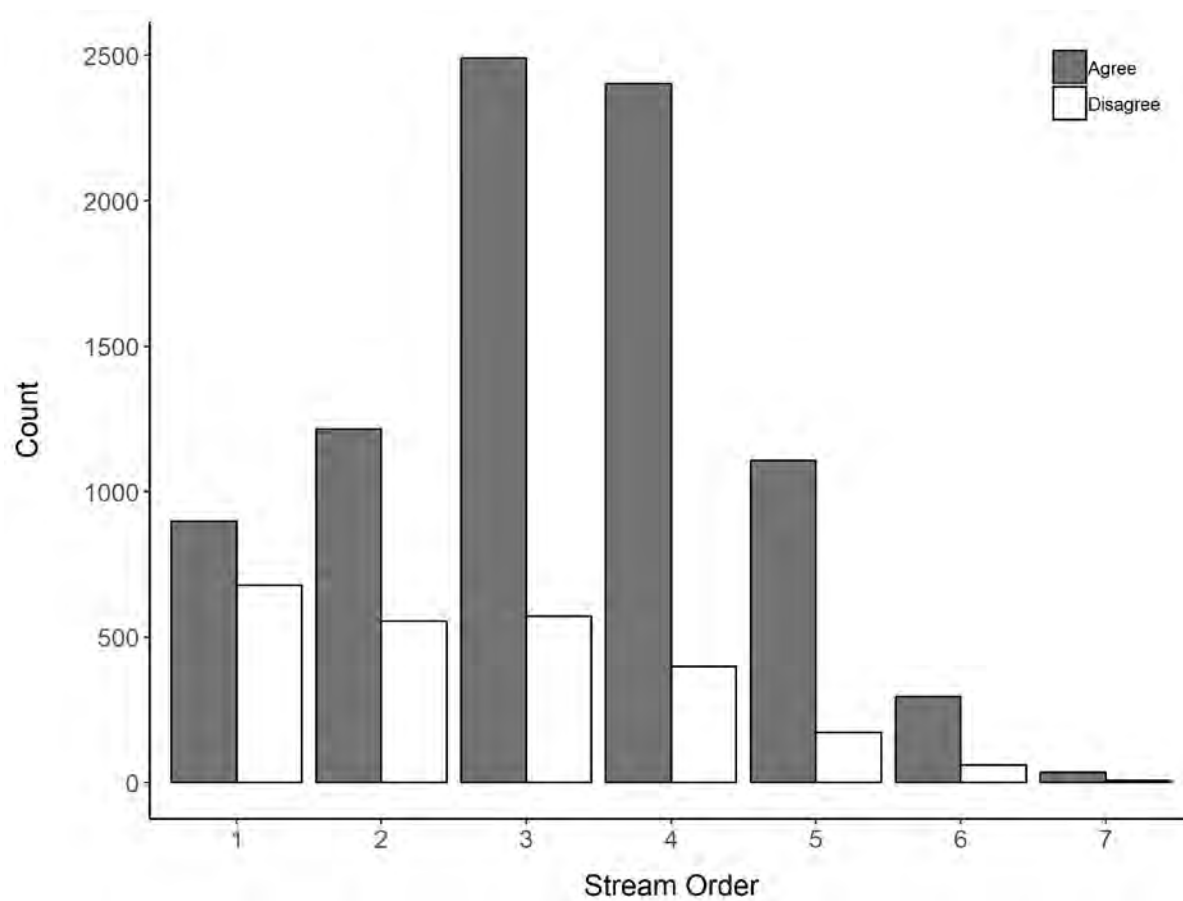


Figure 2.5. Disagreement between non-NHD streamflow observations (NNO) and High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classifications in the Pacific Northwest grouped by NHD-HR stream order.

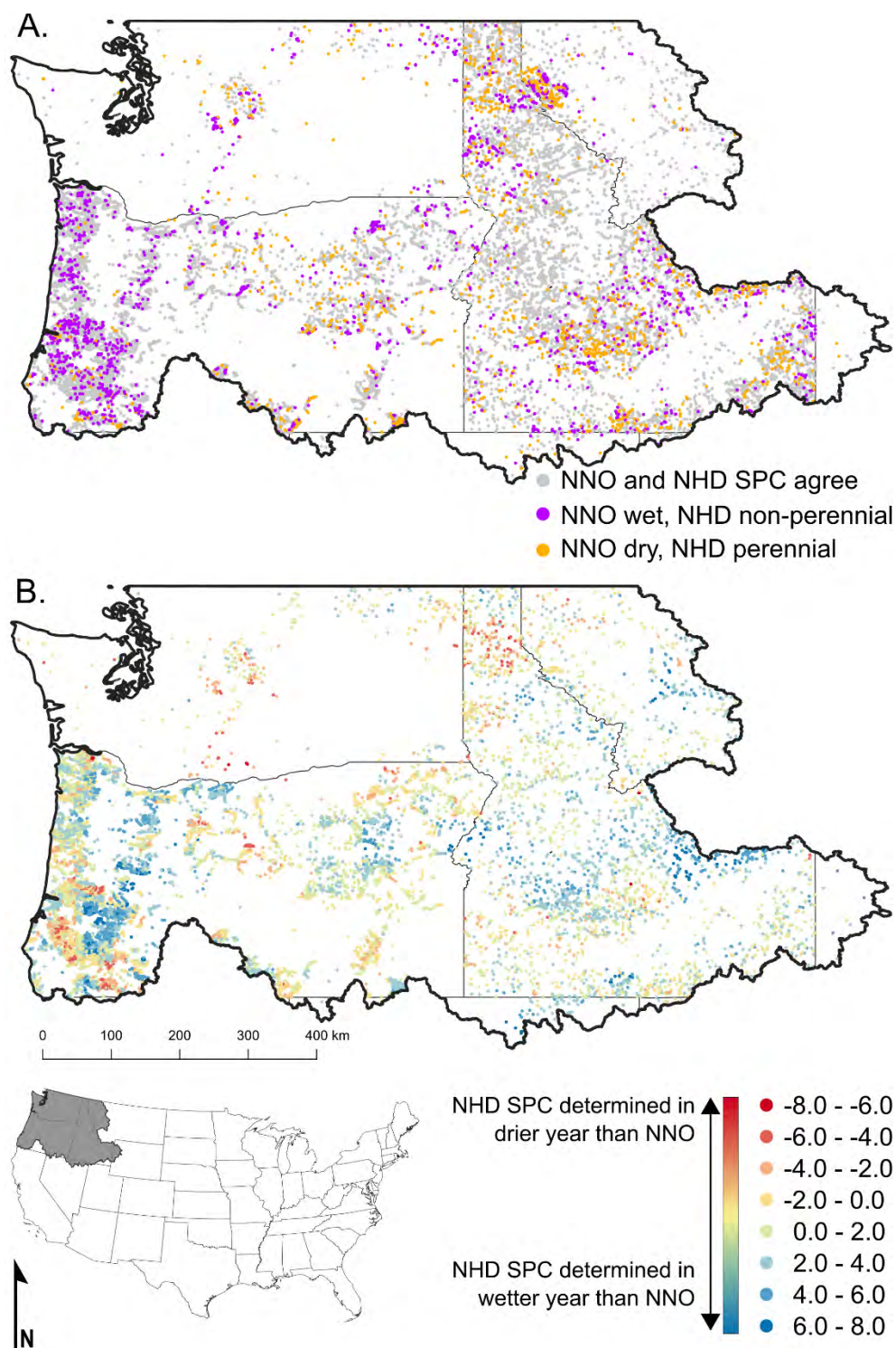


Figure 2.6. A) Disagreement between non-NHD streamflow observations (NNO) and High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classifications and B) the difference in the self-calibrated Palmer Drought Severity Index (PDSI) between the year the NHD stream permanence classification was made (*i.e.* the year of the topographic field survey for a 7.5-minute USGS quadrangle map) and the year the NNO was collected.

*Effects of Climate on NHD-NNO Disagreement in the Pacific Northwest*

Results from logistic regression analysis indicate that disagreements between NNO and NHD-HR SPC increase with increasing dPDSI and decrease with increasing stream order (Table 2; Figure 2.7). On first-order streams, wet NNO were seven-fold more likely to disagree with NHD-HR SPC than dry NNO ( $p < 0.0001$ ), and the probability of disagreement increased as  $dPDSI \geq 0$  increased ( $p < 0.0001$ ; Table 2). There was not a significant effect of  $dPDSI < 0$  ( $p = 0.2856$ ). The probability that a dry NNO disagreed with the NHD-HR SPC increased for stream orders two ( $p < 0.0001$ ) through six ( $p < 0.0001$ ), with a slight decrease from stream orders six to seven ( $p = 0.0301$ ). This slight decrease is likely explained by small numbers of dry observations on larger stream orders (Figure 2.7). These trends can be observed in Figure 2.7, which shows decreased probability of wet NNO disagreeing with NHD-HR SPC as SPC were surveyed in wetter conditions (i.e. as dPDSI becomes increasingly positive) than NNO were collected. Conversely, Figure 2.7 shows the probability of a dry NNO disagreeing with NHD-HR SPC increasing as SPC were surveyed in wetter conditions than NNO. Interactions between OT and both  $dPDSI \geq 0$  and SO are also supported (Table 2). As annual climate during topographic field surveys became increasingly wetter than the climate during the year of a NNO, the probability of a wet observation disagreeing with the NHD-HR SPC decreased. For wet NNO, the probability of disagreement with NHD-HR also decreased for SO greater than one (Table 2).

Table 2.2. Parameter estimates for a logistic regression model to estimate the probability that a streamflow observation disagreed with the High Resolution National Hydrography Dataset (NHD-HR) stream permanence classification (perennial or non-perennial) for a given reach. Model parameters are observation type (OT; wet or dry), difference in Palmer Drought Severity Index (dPDSI) between the year of an observation and the year of the NHD topographic field survey for a reach separated by positive (dPDSI<sub><0</sub>) and negative (dPDSI<sub>≥0</sub>) values, and Strahler stream order (SO; 1-7).

<b>Parameter</b>	<b>Estimate (log odds)</b>	<b><i>p</i>-value</b>
Intercept	-1.368	<0.0001
OT(wet)	1.988	<0.0001
dPDSI <sub>&lt;0</sub>	-0.038	0.2856
dPDSI <sub>≥0</sub>	0.171	<0.0001
SO(2)	0.921	<0.0001
SO(3)	1.530	<0.0001
SO(4)	1.644	<0.0001
SO(5)	1.867	<0.0001
SO(6)	2.066	<0.0001
SO(7)	1.849	0.0301
OT(wet)*dPDSI <sub>&lt;0</sub>	0.065	0.2803
OT(wet)*dPDSI <sub>≥0</sub>	-0.301	<0.0001
OT(wet)*SO(2)	-3.039	<0.0001
OT(wet)*SO(3)	-5.106	<0.0001
OT(wet)*SO(4)	-6.008	<0.0001
OT(wet)*SO(5)	-5.493	<0.0001
OT(wet)*SO(6)	-5.951	<0.0001
OT(wet)*SO(7)	-5.572	<0.0001

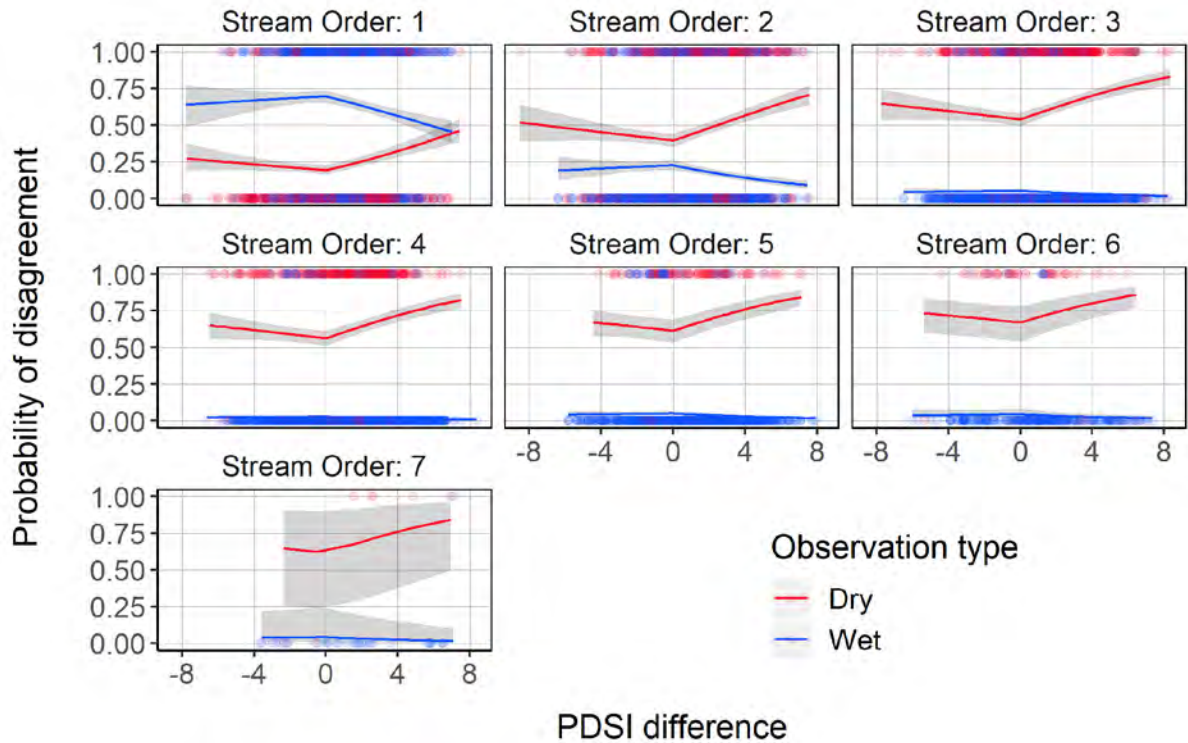


Figure 2.7. Probability that a non-NHD streamflow observation (NNO) disagreed with a High Resolution National Hydrography Dataset (NHD-HR) perennial/non-perennial classification in the Pacific Northwest. Results shown are from a logistic regression model that includes covariates of NNO type (wet or dry) and the difference in Palmer Drought Severity Index (PDSI) between the year the NHD-HR stream permanence classification was made (*i.e.* the year of the topographic field survey for a 7.5-minute USGS quadrangle map) and the year the NNO was made. Red and blue open circles show disagreement/agreement of individual NNO with NHD-HR. Red and blue lines show the modeled (logistic regression) probability of disagreement for dry and wet NNO, respectively.

### Discussion

Overall, we observed approximately 20% disagreement between NNO and NHD SPC. Disagreement was highest on first-order streams (approximately 50%), which was similar to disagreement rates of other studies in the PNW (Fritz et al., 2013; Nadeau et al., 2015). We also observed an effect of differences in annual climate on the probability of disagreement. When field surveys of 7.5 min maps were conducted in a wetter year than NNO were collected, greater annual climate differences resulted in a greater probability of disagreement. These results indicate that perennial streams may be overestimated by NHD, especially when 7.5-min maps were field surveyed in wetter than normal years. The probability of disagreement was highest on first order (headwater) streams (Figure 2.5; Figure 2.7). Our observation of 45% disagreement on first order streams is consistent with other studies in the Pacific Northwest that observed similar (~50%) disagreement rates (Fritz et al., 2013; Nadeau et al., 2015).

Herein we only consider the effect of PDSI on disagreements between NHD SPC and NNO, but there are other factors that may contribute to this disagreement. On average, NNO occurred 30 years after 7.5-min map field validation (range=11 years before to 68 years after; SD=14 years); thus, watershed land-use and water-use could have changed substantially between the time of topographic field surveys and NNO. NNO with known regulation (such as dams or diversions) upstream were excluded from the final dataset (McShane et al., 2017), but other flow regulations including, but not limited to, beaver dams, log jams, or groundwater withdrawals, all of which occur frequently in the PNW, may also alter streamflow patterns in some streams (Gibson and Olden, 2014; Hough-Snee et al., 2014; Konikow and Kendy, 2005). Considerable land-use/land-cover changes, related to logging, wildfire, urbanization and conversion to agriculture, have occurred throughout much of the PNW since 1950 and have important impacts on streamflow (Black et al., 1998). Additionally, cartographic methods for creating 7.5-min maps may have influenced accuracy of NHD-HR SPC. For example, some 7.5-min maps did not go through the entire quality control process and were released as ‘provisional’ maps (p-maps). These p-maps were still used in the creation of NHD. None of the NNO within the extents of p-maps disagreed with NHD SPC in the PNW study area. Nevertheless, p-maps could contribute to disagreement between NHD and observations in other parts of CONUS.

In addition, the definition used to define non-perennial and perennial streams on the 7.5-min maps has changed through time. The definition of perennial and intermittent streams in the original USGS mapping instructions (Beaman, 1928) were “one that flows throughout the year” and “one that is dry for a considerable time each year, say for three months or longer,” respectively. In a subsequent USGS report, Rowland (1955) distinguishes between perennial and intermittent water features as, “Intermittent streams, lakes, and ponds are those that are usually dry at least 6 months of the year. All others are perennial.” In another USGS report describing early efforts to digitize topographic maps, Gupta (1990) gives the definition of perennial and intermittent streams currently used by the NHD (U.S. Geological Survey, NHD User’s Guide. Accessed February 13, 2020, <https://nhd.usgs.gov/userguide.html>) as “containing water throughout the year (except for infrequent periods of severe drought)” and “containing water only part of the year, but more than just after rainstorms and at snowmelt”, respectively. Personal communication from Keven Roth (USGS National Mapping Division, to K. Hafen, November 14, 2018) establishes that the USGS topographic mapping guidelines, based on Beaman (1928), were updated through time without formal publication and that the current perennial/intermittent definitions were implemented in 1961. Thus, it is possible that different SPC definitions have created inconsistency in NHD SPC between 7.5-min maps. Nevertheless, regardless of the SPC definition used during topographic field surveys, NHD SPC are based on the observations of the survey crews. Only 3.1% (313 of 10,055) of NNO used in this study

occurred on 7.5-min maps that were field surveyed prior to 1961, so these inconsistencies would likely play a minor role in our analyses.

The NNO dataset provides over 10,000 observations throughout the PNW but also introduces uncertainty due to inconsistencies in data collection methods. Inconsistent NNO collection methods could have led to different ‘wet’ or ‘dry’ classification on similar reaches. For example, if a stream was not flowing but had isolated pools it could potentially be characterized as ‘wet’ or ‘dry’ depending on the methodology. We do not have explicit knowledge of the methods used; thus, there may be methodological differences we have not accounted for in this study. Additionally, some of the studies that contributed observations potentially used NHD for site selection. For example, an aquatic species study may have used NHD to target only perennial streams. Therefore, NNO from this study would only give observations on streams classified as perennial by NHD, and a consequence could be that no NNO were made on streams classified as non-perennial by NHD. Examples like these may have potentially introduced sample bias into the NNO data. Some evidence of this may be seen in the NNO dataset as there are about 2.5 times more ‘wet’ observations than ‘dry’ observations and about half of the ‘dry’ observations occurred on streams classified as perennial by NHD. Nevertheless, this study (and previous studies, e.g. Jaeger et al., 2019) illustrates the usefulness of simple, easy to collect wet/dry point observations for regional studies. Standardization of data collection procedures would result in greater consistency between observers and potentially create opportunities for data collection by citizen scientists.

Despite the inconsistencies inherent with data collected from various sources over a wide time span, we observed a significant effect of PDSI difference between NNO and NHD SPC on the disagreement between NNO and the NHD. Other studies have documented NHD accuracy across space and over short time periods in the PNW (Fritz et al., 2013; Nadeau et al., 2015), but have not considered the annual climate differences between topographic field survey years and NNO years as a factor affecting accuracy and have not made quantifications at the extent of the entire PNW. The influence of climate on disagreements between NNO and NHD provides support for the continued development of statistical and process-based models that dynamically estimate SPC (Jaeger et al., 2019; Sando and Blasch, 2015; Williamson et al., 2015) as opposed to static NHD classifications. Additionally, practitioners are wise to exercise when using NHD SPC for decision making. This is especially true for perennial streams, which we found to be overestimated by NHD, particularly when topographic field surveys were conducted during wetter climate periods. From our analysis, NHD SPC are most accurate for non-perennial streams. Because NHD is widely used for a variety of purposes (including management and policy decisions), understanding the potential error in the NHD



SPC, and the factors (e.g. climate) that influence errors dynamically, are essential to properly using NHD within its limits.

### Literature Cited

- Alley, W.M., 1984. "The Palmer Drought Severity Index: Limitations and Assumptions." *Journal of Climate and Applied Meteorology* 23:1100–1109.
- Beaman, W.M., 1928. "Topographic Instructions of the United States Geological Survey." In: *Topographic Mapping*, 161-378. U.S. Geological Survey Numbered Series 788. doi:10.3133/b788E.
- Benstead, J.P. and D.S. Leigh, 2012. "An Expanded Role for River Networks." *Nature Geoscience* 5:678–679.
- Black, A.E., E. Strand, R.G. Wright, J.M. Scott, P. Morgan, and C. Watson, 1998. "Land Use History at Multiple Scales: Implications for Conservation Planning." *Landscape and Urban Planning* 43:49–63.
- Caruso, B.S., 2014. "GIS-Based Stream Classification in a Mountain Watershed for Jurisdictional Evaluation." *Journal of the American Water Resources Association (JAWRA)* 50:1304–1324.
- Caruso, B.S. and J. Haynes, 2011. "Biophysical-Regulatory Classification and Profiling of Streams Across Management Units and Ecoregions." *Journal of the American Water Resources Association (JAWRA)* 47:386–407.
- Chorley, R.J. and P.F. Dale, 1972. "Cartographic Problems in Stream Channel Delineation." *Cartography* 7:150–162.
- Colson, T., J. Gregory, J. Dorney, and P. Russell, 2008. "Topographic and Soil Maps Do Not Accurately Depict Headwater Stream Networks." *National Wetlands Newsletter* 30:25–28.
- Cooper, M.G., J.R. Schaperow, S.W. Cooley, S. Alam, L.C. Smith, and D.P. Lettenmaier, 2018. "Climate Elasticity of Low Flows in the Maritime Western U.S. Mountains." *Water Resources Research* 54:5602–5619.
- Drummond, R.R., 1974. "When Is a Stream a Stream?" *The Professional Geographer* 26:34–37.
- Fritz, K.M., E. Hagenbuch, E. D'Amico, M. Reif, P.J. Wigington, S.G. Leibowitz, R.L. Comeleo, J.L. Ebersole, and T.-L. Nadeau, 2013. "Comparing the Extent and Permanence of Headwater Streams From Two Field Surveys to Values From Hydrographic Databases and Maps." *Journal of the American Water Resources Association (JAWRA)* 49:867–882.
- Gibson, P.P. and J.D. Olden, 2014. "Ecology, Management, and Conservation Implications of North American Beaver (*Castor Canadensis*) in Dryland Streams." *Aquatic Conservation: Marine and Freshwater Ecosystems* 24:391–409.
- Godsey, S.E. and J.W. Kirchner, 2014. "Dynamic, Discontinuous Stream Networks: Hydrologically Driven Variations in Active Drainage Density, Flowing Channels and Stream Order." *Hydrological Processes* 28:5791–5803.
- Guptill, S.C., 1990. "An Enhanced Digital Line Graph Design." U.S. Geological Survey Numbered Series Report 1048. doi: 10.3133/cir1048.
- Hill, R.A., M.H. Weber, S.G. Leibowitz, A.R. Olsen, and D.J. Thornbrugh, 2016. "The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States." *Journal of the American Water Resources Association (JAWRA)* 52:120–128.

- Hough-Snee, N., A. Kasprak, B.B. Roper, and C.S. Meredith, 2014. "Direct and Indirect Drivers of Instream Wood in the Interior Pacific Northwest, USA: Decoupling Climate, Vegetation, Disturbance, and Geomorphic Setting." *Riparian Ecology and Conservation* 2:14–34.
- Isaak, D.J., M.K. Young, C.H. Luce, S.W. Hostetler, S.J. Wenger, E.E. Peterson, J.M. Ver Hoef, M.C. Groce, D.L. Horan, and D.E. Nagel, 2016. "Slow Climate Velocities of Mountain Streams Portend Their Role as Refugia for Cold-Water Biodiversity." *Proceedings of the National Academy of Sciences of the United States of America* 113:4374–4379.
- Jaeger, K.L., R. Sando, R.R. McShane, J.B. Dunham, D.P. Hockman-Wert, K.E. Kaiser, K. Hafen, J.C. Risley, and K.W. Blasch, 2019. "Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest." *Journal of Hydrology X* 2:100005.
- Johnston, C.M., T.G. Dewald, T.R. Bondelid, B.B. Worstell, L.D. McKay, A. Rea, R.B. Moore, and J.L. Goodall, 2009. "Evaluation of Catchment Delineation Methods for the Medium-Resolution National Hydrography Dataset." *U.S. Geological Survey Scientific Investigations Report 2009-5233*. <http://pubs.er.usgs.gov/publication/sir20095233>. Accessed 31 Oct 2018.
- Konikow, L.F. and E. Kendy, 2005. "Groundwater Depletion: A Global Problem." *Hydrogeology Journal* 13:317–320.
- Leibowitz, S.G., R.L. Comeleo, P.J. Wigington, M.H. Weber, E.A. Sproles, and K.A. Sawicz, 2016. "Hydrologic Landscape Characterization for the Pacific Northwest, USA." *Journal of the American Water Resources Association (JAWRA)* 52:473–493.
- Leopold, L.B., 1994. *A View of the River*. Harvard University Press.
- Macfarlane, W.W., J.M. Wheaton, N. Bouwes, M.L. Jensen, J.T. Gilbert, N. Hough-Snee, and J.A. Shivik, 2017. "Modeling the Capacity of Riverscapes to Support Beaver Dams." *Geomorphology* 277:72–99.
- Mark, D.M., 1983. "Relations Between Field-Surveyed Channel Networks and Map-Based Geomorphometric Measures, Inez, Kentucky." *Annals of the Association of American Geographers* 73:358–372.
- McShane, R.R., R. Sando, and P.D. Hockman-Wert, 2017. Streamflow Observation Points in the Pacific Northwest, 1997-2016. U.S. Geological Survey data release. <https://doi.org/10.5066/F7BV7FSP>.
- Nadeau, T.-L., G.S. Leibowitz, J.P. Wigington, L.J. Ebersole, M.K. Fritz, A.R. Coulombe, L.R. Comeleo, and A.K. Blocksom, 2015. "Validation of Rapid Assessment Methods to Determine Streamflow Duration Classes in the Pacific Northwest, USA." *Environmental Management* 56:34–53.
- Nadeau, T.-L. and C.M. Rains, 2007. "Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy." *Journal of the American Water Resources Association (JAWRA)* 43. <http://doi.wiley.com/10.1111/j.1752-1688.2007.00010.x>.
- National Rivers and Streams Assessment 2008-2009: A Collaborative Survey. U.S. Environmental Protection Agency, Washington, D.C., 2016.
- O'Connor, B.L., Y. Hamada, E.E. Bowen, M.A. Grippio, H.M. Hartmann, T.L. Patton, R.A. Van Lonkhuyzen, and A.E. Carr, 2014. "Quantifying the Sensitivity of Ephemeral Streams to Land Disturbance Activities in Arid Ecosystems at the Watershed Scale." *Environmental Monitoring and Assessment* 186:7075–7095.

- Palmer, C.W., 1965. Meteorological Drought. U.S. Weather Bureau.
- PRISM Climate Group, Oregon State University. <http://prism.oregonstate.edu/>. Accessed 31 Oct 2018.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Rowland, J.B., 1955. Features Shown on Topographic Maps. US Geological Survey Circular 368.
- Roy, A.H., A.L. Dybas, K.M. Fritz, and H.R. Lubbers, 2009. "Urbanization Affects the Extent and Hydrologic Permanence of Headwater Streams in a Midwestern US Metropolitan Area." *Journal of the North American Benthological Society* 28:911–928.
- Ruprecht, J.S., K.R. Hersey, K. Hafen, K.L. Monteith, N.J. DeCesare, M.J. Kauffman, and D.R. MacNulty, 2016. "Reproduction in Moose at Their Southern Range Limit." *Journal of Mammalogy* 97:1355–1365.
- Sando, R. and W.K. Blasch, 2015. "Predicting Alpine Headwater Stream Intermittency: A Case Study in the Northern Rocky Mountains." *Ecohydrology & Hydrobiology* 15. <https://linkinghub.elsevier.com/retrieve/pii/S1642359315000154>.
- Sheng, J., J.P. Wilson, N. Chen, J.S. Devinny, and J.M. Sayre, 2007. "Evaluating the Quality of the National Hydrography Dataset for Watershed Assessments in Metropolitan Regions." *GIScience & Remote Sensing* 44:283–304.
- Stanislawski, L.V., 2009. "Feature Pruning by Upstream Drainage Area to Support Automated Generalization of the United States National Hydrography Dataset." *Computers, Environment and Urban Systems* 33:325–333.
- Stoddard, J., D. Peck, A. Olsen, D. Larsen, J. Van Sickle, C. Hawkins, R. Hughes, T. Whittier, G. Lomnický, and A. Herlihy, 2005. "Environmental Monitoring and Assessment Program (EMAP): Western Streams and Rivers Statistical Summary." Washington, DC: US Environmental Protection Agency, Office of Research and Development.
- Strahler, A.N., 1952. "Hypsometric (Area-Altitude) Analysis of Erosional Topography." *GSA Bulletin* 63:1117–1142.
- Strahler, A.N., 1957. "Quantitative Analysis of Watershed Geomorphology." *Eos, Transactions American Geophysical Union* 38:913–920.
- USGS US Topo and Historical Topographic Map Collection, . <https://geonames.usgs.gov/apex/f?p=262:1:0>. Accessed 5 Apr 2019.
- Wells, N., S. Goddard, and M.J. Hayes, 2004. "A Self-Calibrating Palmer Drought Severity Index." *Journal of Climate* 17:2335–2351.
- Williamson, T.N., C.T. Agouridis, C.D. Barton, J.A. Villines, and J.G. Lant, 2015. "Classification of Ephemeral, Intermittent, and Perennial Stream Reaches Using a TOPMODEL-Based Approach." *Journal of the American Water Resources Association (JAWRA)* 51:1739–1759.
- Wood, M.S., A. Rea, K.D. Skinner, and J.E. Hortness, 2009. "Estimating Locations of Perennial Streams in Idaho Using a Generalized Least-Squares Regression Model of 7-Day, 2-Year Low Flows." *U.S. Geological Survey Scientific Investigations Report 2009-5015*. <https://pubs.er.usgs.gov/publication/sir20095015>

## **Chapter 3: Precision of headwater stream permanence estimates from a monthly water balance model in the Pacific Northwest, USA**

### **Abstract**

The presence and duration of surface water in streams has important biological and ecological implications. It is also the metric that determines a stream's regulatory status in the United States. However, there are currently no models or products adequately describe surface water presence well enough for regulatory determinations. We modified the widely implemented Thornthwaite monthly water balance model (MWBm) with a flow threshold parameter to estimate flow permanence and evaluated the model's accuracy and precision on over 1.3 million headwater streams in the United States' Pacific Northwest (PNW). Stream reaches were assigned to one of eight calibration groups by unsupervised classification based on sensitivity to MWBM parameters. MWBM stream permanence estimates were benchmarked against the accuracy of National Hydrography Dataset (NHD) stream permanence classifications (SPC) on headwater streams, as determined by previous studies, and model precision was assessed for parameter sets that were at least as accurate as NHD SPC. Parameter sets that met benchmarks were identified for five of eight calibration regions (60% of streams). MWBM stream permanence estimates exhibited low precision for approximately 20% of headwater streams, resulting in 40% of the 1.3 million headwater stream reaches in the PNW that exhibited high precision and met accuracy benchmarks. Results point to the need for increased collection of surface water presence observations to improve calibration and assessment of stream permanence models. Additionally, implementation of MWBM indicates potential for process-based models to predict stream permanence condition with future development.

### **Introduction**

In the United States, stream permanence classifications (i.e. perennial, intermittent, ephemeral) are a primary consideration to determine stream regulatory status under the Clean Water Act and are also an important indicator of environmental conditions and biodiversity (Fritz et al., 2008; Nadeau et al., 2015; Vander Vorste et al., 2020). Currently, the National Hydrography Dataset (NHD) is the most comprehensive dataset describing stream permanence for the contiguous United States (CONUS; Nadeau and Rains, 2007). However, NHD stream permanence classifications (SPC) have been shown to exhibit up to 50% disagreement with in situ observations (Fritz et al., 2013; Nadeau et al., 2015) and the highest disagreement rates occur on headwater streams (Hafen et al., 2020). The US Environmental Protection Agency and Army Corps of Engineers have determined that

NHD SPC derived decades in the past are not always adequate for regulatory determinations, and that more reliable SPC mapping products are required that consider dynamic changes in climate and land use in headwater and low stream order environments

([https://www.epa.gov/sites/production/files/2020-01/documents/nwpr\\_fact\\_sheet\\_-\\_mapping.pdf](https://www.epa.gov/sites/production/files/2020-01/documents/nwpr_fact_sheet_-_mapping.pdf)).

A primary reason for the uncertainty of NHD SPC is that designations originated from observations made by topographic survey crews throughout the 1900s (Beaman, 1928; Fritz et al., 2013; Hafen et al., 2020). This classification method incorporates first-hand knowledge of stream reaches but is limited to climatic conditions during the survey year (or recent past). Thus, climatic conditions during the survey year may not have represented long-term average conditions, and observations may not have captured the full range of variability for surface water conditions at a stream reach (Godsey and Kirchner, 2014; Hafen et al., 2020). Another reason for uncertainty is the SPC designations themselves. The definitions for perennial, intermittent, and ephemeral streams have changed through time, and were not given definitions that could be accurately assessed by survey crews over a short period of time (Beaman, 1928; Guptill, 1990; Hafen et al., 2020). For example, perennial streams were defined as “streams that flow continuously except in periods of extreme drought”, but extreme drought is not defined (Guptill, 1990). Additionally, surveys for adjacent topographic maps were often made multiple years apart. Thus, SPC designations for part of watershed could have been made under very different climate conditions than other parts of the same watershed (Hafen et al., 2020).

Uncertainties in NHD SPC also arise from representing surface water presence, which is a dynamic occurrence (Godsey and Kirchner, 2014), with a static classification. Multiple efforts have sought to add additional context to NHD SPC by modeling streamflow for NHD stream reaches (McKay et al., 2012; Miller et al., 2018). Two efforts include the NHD enhanced runoff method (EROM), which is a unit hydrograph approach that is integrated into the NHDPlus itself (McKay et al., 2012), and a machine learning approach developed by the U.S. Geological Survey (Miller et al., 2018; hereafter referred to as the “Miller data” or “Miller streamflow model”). Key to both modeling efforts were runoff estimates that are generated from the USGS

Thornthwaite Monthly Water Balance Model (MWBM; McCabe and Markstrom, 2007). The MWBM has been implemented in multiple studies to evaluate various components of the hydrological cycle, water supply, and water demand for the CONUS (Bock et al., 2016, 2018; McCabe and Wolock, 2007, 2008, 2011). While these efforts address runoff and streamflow, they do not necessarily inform surface water presence.

Both EROM and the Miller streamflow models are calibrated to gauged streamflow time-series, which is the general standard for hydrological modeling. However, results can only be

evaluated where continuous gaging stations exist. In the United States stream gages are strategically placed to primarily monitor water supply and flood risk (Hester et al., 2006). As a result of these priorities the majority of stream gages are located on larger-order rivers and streams, with very few gages in headwater catchments (DeWeber et al., 2014). It is in headwater catchments where the greatest uncertainty of NHD SPC is documented and where few continuous records of streamflow exist to inform traditional modeling efforts. As a result, performance (e.g. accuracy and uncertainty) of regional- and national-extent models in headwater streams is largely unknown. Assessing model performance in headwater streams is especially salient because more than 50% of all stream reaches are classified as headwaters (Lowe and Likens, 2005).

To increase the spatial coverage of calibration and validation data from fixed stream gages, some statistical and process-based models have used simple surface water presence observations (SWPO) to develop and evaluate stream permanence estimates (Gendaszek et al., 2020; Jaeger et al., 2019; Jensen et al., 2017; Ward et al., 2018; Williamson et al., 2015). Surface water presence observations are easy to collect and represent more spatial locations than streamflow gages (Beaufort et al., 2019). While SWPO represent many different locations, they are often not repeated over time and, thus, have much lower temporal resolution than stream gage data (Beaufort et al., 2019). Nevertheless, SWPO have been used to develop statistical models at local (Gendaszek et al., 2020; Jensen et al., 2017) and regional extents (Jaeger et al., 2019; Yu et al., 2018) and process-based models (Ward et al., 2018, 2020; Williamson et al., 2015) at local extents where existing stream gage networks did not provide sufficient information to model stream permanence on headwaters and other small tributary streams.

Recently, stream intermittence was assessed in Australia with a daily water balance model calibrated and validated with stream gage data (Yu et al., 2018, 2020). These results indicate that the MWBM may be useful for modeling stream permanence, especially since the MWBM can easily be applied at regional, continental, and even global extents. On the other hand, the structure of the MWBM also suggests it may not be suitable for modeling stream permanence, particularly in headwaters and small tributaries. Runoff estimates from the MWBM are generated at the spatial resolution of input climate data, usually 1-4 km (1-16 km<sup>2</sup>, depending on the data source). Thus, the spatial resolution of runoff estimates is coarse compared to the drainage area of headwater catchments, which are mostly smaller than 2 km<sup>2</sup>. Additionally, flow processes in the MWBM are partitioned into direct runoff (i.e. infiltration excess flow) and all other flow (e.g. saturation excess flow, baseflow, lateral flow). The MWBM does not explicitly represent saturation excess flow,

baseflow, or lateral flow, which can be important components for estimating streamflow in headwater streams (Fischer et al., 2015; Kosugi et al., 2006).

Due to the MWBM's usefulness for large extent studies, and the ability to incorporate changing climate inputs (Ward et al., 2020), it is a reasonable starting place for modeling dynamic stream permanence at a regional extent. Our primary objective is to determine if the MWBM can generate dynamic SPC estimates with accuracy similar to NHD SPC (e.g. 60-65%; Fritz et al., 2013; Hafen et al., 2020). To accomplish this objective, we first assess the accuracy of annual stream permanence estimates generated from the MWBM on headwater streams to identify parameter combinations that perform at least as well as NHD SPC. Then, with the suitable parameter combinations, we calculate the precision of MWBM SPC on headwater streams throughout the United States' Pacific Northwest (PNW). Information from this study will help establish how the MWBM can be used to assess stream permanence at large spatial extents and identify where more data and different modeling approaches may be necessary to improve stream permanence estimates.

## Methods

### *Monthly Water Balance Model*

The MWBM has been previously implemented at global and CONUS extents. Monthly runoff values are generated by estimating the magnitude of hydrologic processes that supply and demand water. Equations governing hydrological processes are explained in detail by McCabe and Markstrom (2007). The basic model logic progresses as follows and is similarly explained by Bock et al. (2016).

Model inputs are mean monthly temperature ( $T$ ), total monthly precipitation ( $P$ ), and soil water holding capacity. Monthly estimates of  $P$  and  $T$  were obtained from PRISM (~4 km) data (PRISM Climate Group, Oregon State University). Soil water holding capacity estimates were obtained from STATSGO polygons (Schwarz and Alexander, 1995). STATSGO polygons were intersected with the extents of PRISM grid cells to obtain the water holding capacity corresponding to the footprint of each PRISM grid cell. The MWBM allows precipitation to occur as rain or snow, determined by  $T$ . Snowfall accumulates from month to month to form a snowpack, which melts as temperatures warm. Rainfall can be converted to direct runoff, evapotranspiration (ET), soil moisture storage, and surplus water. Monthly potential evapotranspiration (PET) is determined by  $T$  and latitude per the Hamon equation (Hamon, 1961). When the sum of rainfall and snowmelt for a month is less than PET, actual evapotranspiration (AET) is the sum of rainfall, snowmelt, and the portion of water that is evaporated from the soil. When the sum of rainfall and snowmelt is greater than or equal to PET, AET is equal to PET. Water remaining after AET recharges soil water storage. Water in excess of AET and soil water storage becomes surplus. A specified proportion of surplus is converted to runoff each month, and the



remaining surplus is temporarily held in storage. Thus, water is lost through AET and total runoff is the sum of direct runoff and surplus runoff (Figure 3.1). Runoff was summed for each NHDPlus HR catchment and multiplied by catchment area to arrive at a runoff volume for each month. Runoff volumes were converted to mean monthly flow (volume/s).

In addition to ET, water fluxes in the MWBM are modulated by five parameters (Table 1). Rain temperature (TR) is the temperature above which all precipitation falls as rain. Snow temperature (TS) is the temperature below which all precipitation falls as snow. When T is less than TR and greater than TS the proportion of rain to snow is determined by linear interpolation. Snow-melt coefficient (MC) is the maximum proportion of snow storage that can melt in a single month. Direct runoff (DR) is the proportion of precipitation and snowmelt that becomes overland runoff. Runoff factor (RF) determines the proportion of watershed storage that is converted to runoff each month. We also added precipitation factor (PF) and temperature addition (TA) parameters to adjust climate inputs to the MWBM. PF and TA were included in sensitivity and calibration analyses to identify the impact of climate inputs on MWBM performance. Descriptions, units, and value ranges of model parameters are presented in Table 3.1.

To convert MWBM runoff estimates to permanent or nonpermanent stream classifications, we added the flow threshold parameter (FT). Streams were classified as permanent for a month (model time-step) when mean monthly streamflow was greater than FT and non-permanent for a month when mean monthly streamflow was less than FT. Flow threshold has been used in previous stream permanence modeling studies to identify surface water presence and absence (Yu et al., 2020; Ward et al., 2018). We classified a stream to be permanent for a calendar year (annual permanence) when the MWBM predicted a stream to be permanent for each month of the year. Streams the MWBM predicted to be dry at least one month of a calendar year were classified as non-permanent for the calendar year. We determined stream permanence at an annual time-step to align with the observational data available for calibration. Additionally, the current best data describing stream permanence for headwater streams use perennial, and non-perennial classifications which describe permanence at a minimum time-step of one year.

## Monthly Water Balance Model

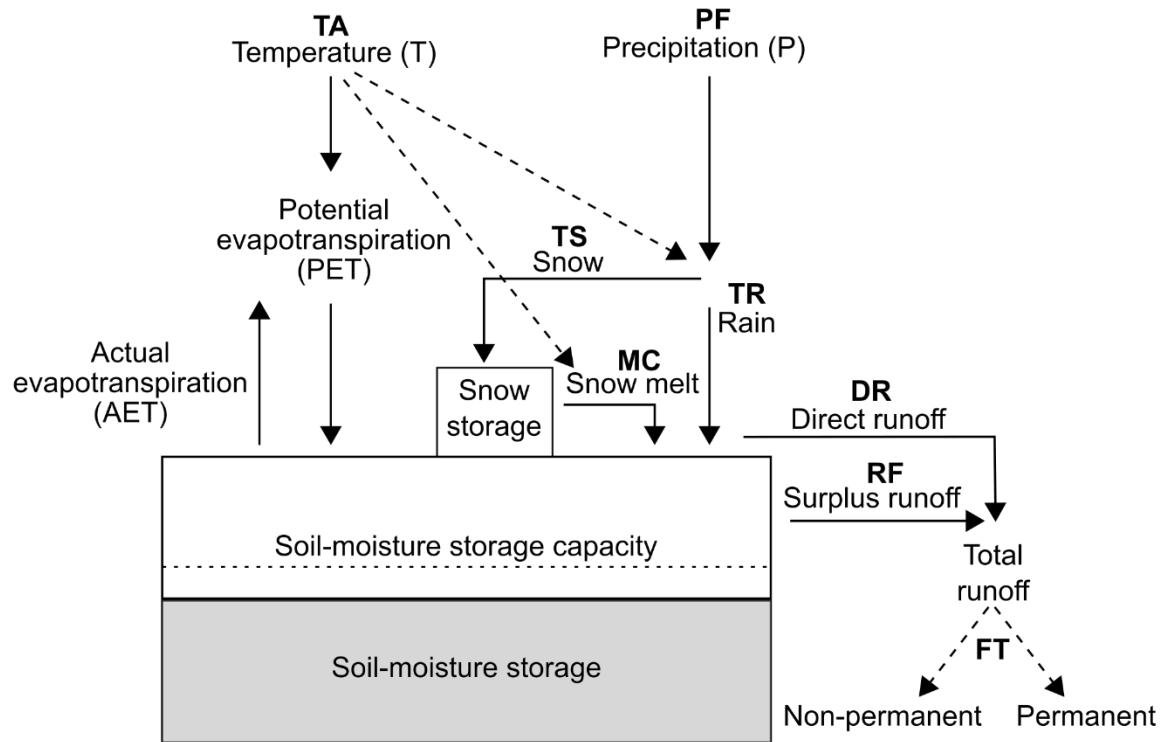


Figure 3.1. Diagram of the monthly water balance model. Bold parameter abbreviations correspond to the parameter descriptions and ranges in Table 1. Adapted from McCabe and Markstrom, 2007.

Table 3.1. Parameters assessed in sensitivity analysis and calibration of the monthly water balance model.

Parameter Name	ID	Description	Units	Range
Runoff Factor	RF	Proportion of catchment storage that is converted to streamflow each month	-	0.0 - 1.0
Direct Runoff Factor	DR	Proportion of precipitation that is converted to streamflow without infiltrating or evaporating	-	0.0 - 0.5
Snow Temperature	TS	Temperature below which all precipitation is snow	°C	-10.0 - 2.0
Rain Temperature	TR	Temperature above which all precipitation is rain	°C	0.0 - 10.0
Snow-melt Coefficient	MC	The maximum proportion of snow water equivalent that can melt in a single month	-	0.0 - 1.0
Flow Threshold	FT	Mean monthly flow above which a stream segment is considered permanent	L/s	0.0 - 14.2
Precipitation Factor	PF	Multiplier for input PRISM precipitation	-	0.1 - 2.0
Temperature Addition	TA	Value added to increase or decrease mean monthly temperature	°C	-2.0 - 2.0

### Study Area

The study area consisted of the Pacific Northwest region of the United States (PNW; Figure 3.2). We define the PNW as Hydrographic Region 17 of the United States as defined by Seaber et al, (1987). PNW is bounded on the north by the USA-Canada border, the West by the Pacific Ocean, and the interior by the boundaries of the Columbia River Basin, coastal watersheds, and several endorheic basins. Elevations in the region range from sea level in the coastal regions to over 4,000 m in the Cascade Mountain Range east of the Pacific Ocean. Annual precipitation ranges from 200 mm in the rain shadow of the Cascade Mountains to over 5,000 mm in the coastal regions. Previous studies have examined stream permanence in this region through field observation and modeling, resulting in a rich dataset of publicly available surface water presence observations that are imperative for this modeling study (Hafen et al., 2020; Jaeger et al., 2019, 2020; McShane et al., 2017).

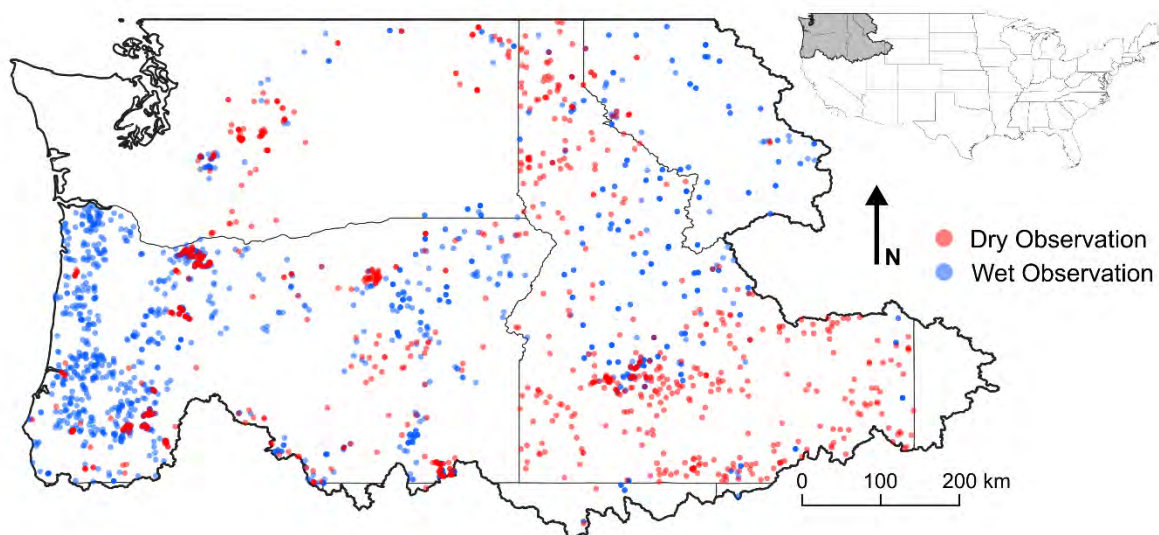


Figure 3.2. Location of *in situ* observations of surface water presence or absence on headwater streams in the Pacific Northwest from 1977-2019.

### Model Application

Stream catchments and flowlines in the Pacific Northwest were represented by the High-Resolution version of the National Hydrography Dataset with added attributes (NHDPlus HR; Moore et al., 2019). The MWBM was applied to headwater streams. We defined headwater streams and catchments as those with no upstream segments or tributaries. With the MWBM we calculated total monthly runoff within each headwater catchment in the PNW at the resolution of the input PRISM grid (~4 km; Figure 3.3). Total monthly runoff volume for headwater catchments was calculated by multiplying MWBM runoff depth by catchment area. When catchments intersected multiple PRISM

grid cells, the volume was calculated for each intersecting area and summed to obtain the total volume. We calculated the mean monthly flow rate (L/s) at the outlet of each headwater stream segment as the catchment's total monthly runoff volume divided by the number of seconds in a given month. Thus, MWBM results represent runoff estimates at the outlet (most downstream point) of each headwater stream in the PNW.

Monthly stream permanence was determined with the FT parameter. When mean monthly flow was less than FT the stream reach was classified as non-permanent for that month. When mean monthly flow was greater than or equal to FT the stream reach was classified as permanent for that month. Stream reaches that were classified as permanent every month of a calendar year were classified as permanent for that year. Any stream reach that was classified as non-permanent for least one month during a calendar year was classified as non-permanent for that year. This analysis grouped both intermittent and ephemeral streams into the non-permanent category.

We implemented the MWBM as presented by Markstrom and McCabe (2007) in the Python programming language in order to adapt the MWBM to headwater streams and scale millions of model runs on a super computer. All MWBM model runs were completed on the USGS Yeti supercomputer (USGS Advanced Research Computing).

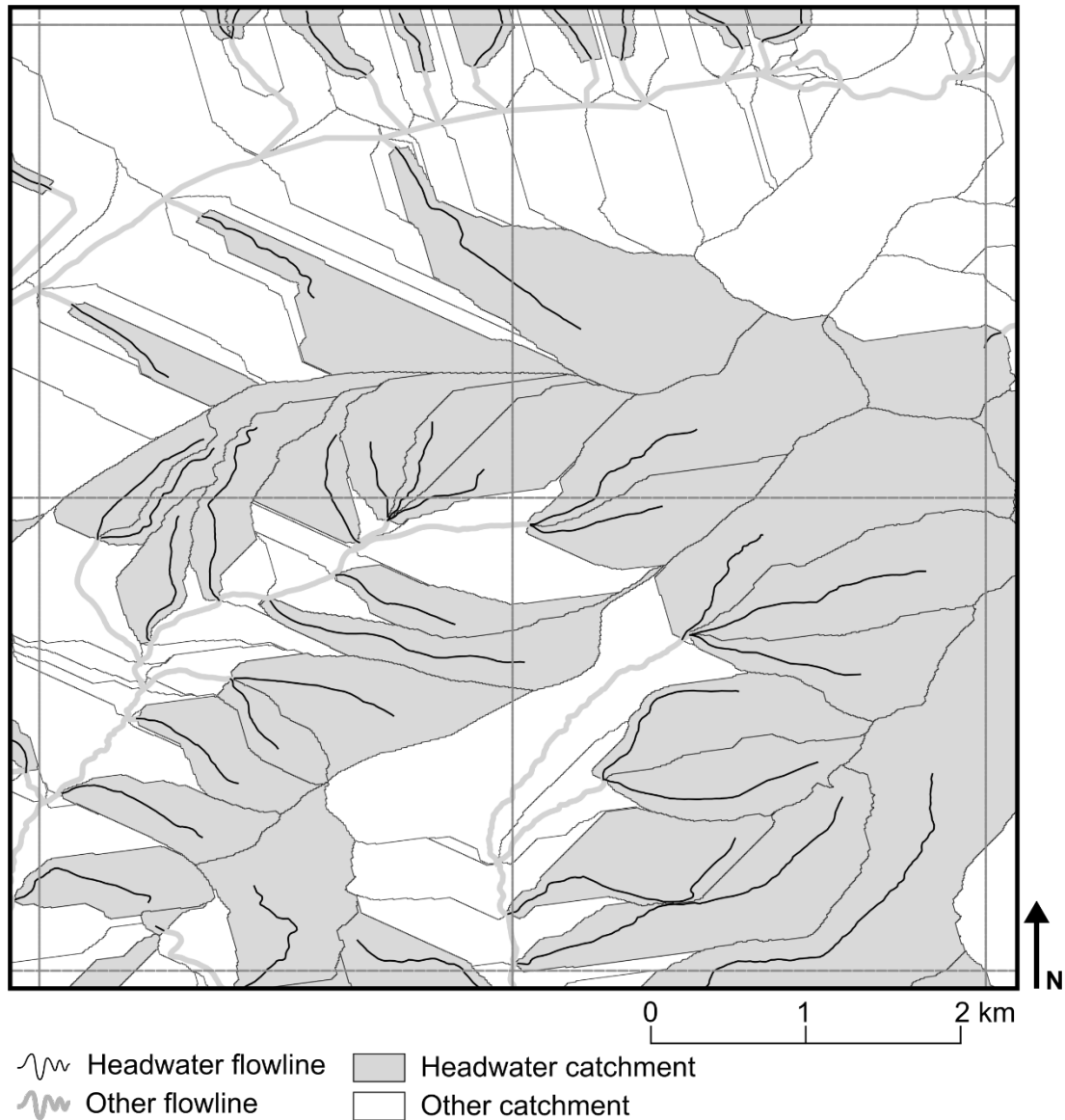


Figure 3.3. Catchments and streams that show the geospatial fabric to which the monthly water balance model was applied. The grid shows the extent of individual PRISM grid cells (~4 km)

#### *Observation Data*

Model calibration and precision analyses were conducted with observational data that described presence or absence of surface water in headwater streams. Because few stream gages exist on headwater streams these observations of surface water presence provide the best spatial coverage of headwater streams in the PNW. Observational data were primarily acquired from a dataset compiled as part of the Probability of Stream Permanence (PROSPER; Jaeger et al., 2019) modeling effort in the PNW (McShane et al., 2017). These data were supplemented with more recent observations collected with the FLOWPER application (Jaeger *et al.*, 2020). Surface water presence

observations were made by a variety of agencies, including the Sauk-Suiattle Indian Tribe, Idaho Department of Environmental Quality, Oregon Department of Fish and Wildlife, US Bureau of Land Management, US Forest Service, USGS, and the US Environmental Protection Agency. Observations spanned the years 1977-2019. Each observation recorded the date (some older observations only specified the year) the observation was made, the geographic coordinates of the observation location, and the presence (wet) or absence (dry) of surface water in the observed stream channel.

All dry observations were used to represent streams that were annually non-permanent. Only wet observations made in August or September were used to represent annually permanent streams. An observation point was assumed to represent the entire flowline (i.e. headwater stream reach). Each observation was joined to the nearest NHDPlus HR flowline. Observations farther than 50 m from a flowline were excluded from the analysis. All observations greater than 20 m from a flowline were manually inspected to ensure they were associated with the correct flowline. Similar methodologies were implemented by Hafen et al, (2020) and Jaeger et al, (2019) to identify observations for stream permanence assessment and modeling. In all, 2,804 observations (1,120 dry and 1,684 wet) were used for model calibration (Figure 3.2). Though the MWBM produces runoff and stream permanence estimates at a monthly time-step, we assessed the model on annual permanence because approximately half of the dry observations only specified the year of the observation and not the day and month.

### *Sensitivity Analysis*

We assessed the relative sensitivity of annual stream permanence classification from the MWBM to each of the eight model parameters with the Fourier Amplitude Sensitivity Test (FAST) over the period 1977-2019. Sensitivity analysis was conducted with the SALib Python module (Herman and Usher, 2017). In all we tested 1,200 parameter combinations for ~1.3 million headwater catchments in the PNW. Parameter values were selected from uniform distributions displayed in Table 1. For many stream reaches, the sensitivity analysis produced not-a-number values. Upon inspection of the model runs we observed that most of the not-a-number results occurred on stream reaches where greater than 95% of simulations predicted the same outcome (i.e. permanent, non-permanent) for all years. This indicated low sensitivity to model parameters for these stream reaches. To represent this sensitivity in further analysis, the sensitivity value for stream reaches where the initial FAST result produced not-a-number, but where greater than 99% of FAST simulations produced the same result, was set to 0.01.

## *Model Precision Analysis*

### Parameter Regionalization

Parameters for headwater stream reaches were regionalized by grouping reaches that responded similarly to changes in model parameters (e.g. Beck et al., 2016; Bock et al., 2016). We performed an unsupervised, K-means classification (Pedregosa et al., 2011) based on MWBM parameter sensitivities to assign stream reaches to calibration regions, or groups. Each headwater stream reach, as defined by NHDPlus HR, was classified based on its sensitivity to the eight MWBM parameters as calculated in the previous section. We selected the number of calibration groups by qualitatively balancing the K-means classification error (squared distance of each point to the class center) with the number of observations available for each group. The number of observations for each group decreased as the number of calibration groups increased. We tested the K-means classification for 2-15 groups and determined that eight calibration groups best minimized K-means classification error while simultaneously maximizing the number of observations in each group. Physical characteristics, parameter sensitivities and number of observations for each calibration group are presented in the Results section.

### Parameter Set Selection

We evaluated model accuracy for one million randomly selected (Monte Carlo) parameter sets from the same uniform distributions with the same ranges used for sensitivity analysis (Table 1). For parameter set selection, we only ran simulations for the stream reaches where observations were located. The accuracy of a parameter set was evaluated against the surface water presence observations as the sum of simulated stream permanence classifications that agreed with observations divided by the total number of observations. We used previous studies that quantified the accuracy of NHD SPC on headwater streams to range from 50-65% (e.g. Fritz et al., 2013; Hafen et al., 2020; Nadeau et al., 2015) as benchmarks to identify parameter sets that produced a good model ‘fit’ in each calibration region. Therefore, we required suitable parameter sets to have greater than 65% overall accuracy against all observations and greater than 60% accuracy against wet or dry observations individually. All parameter sets that produced suitable accuracy results in a calibration region were retained to assess model precision.

### Precision Analysis

Model results were calculated for each suitable parameter set in each calibration region. Model precision was assessed by agreement of simulated stream permanence classifications between adequate parameter sets. High model precision was exhibited when all suitable parameter sets simulated the same permanence classification. Model confidence decreased as different parameter

sets predicted different permanence classifications for the same stream. We represented model precision as the proportion of parameter sets resulting in a permanent classification. Model precision values ranged from 0.0-1.0 where values of 0.0 and 1.0 indicated all parameter sets simulated non-permanent (0.0) and permanent (1.0) classifications, respectively. A value of 0.5 indicated that half of the parameter sets simulated permanent and the other half simulated non-permanent, resulting in low model precision.

## **Results**

### *Sensitivity Analysis*

Generally, parameter sensitivities were low for interior mountain ranges, moderate for coastal mountain ranges, and high on the interior plains and plateaus (Figure 3.4). Most headwater streams exhibited some degree of sensitivity to the RF and FT parameters. This is expected as RF is the main parameter controlling monthly flow rates and FT is the primary parameter controlling the model result (permanent or non-permanent).



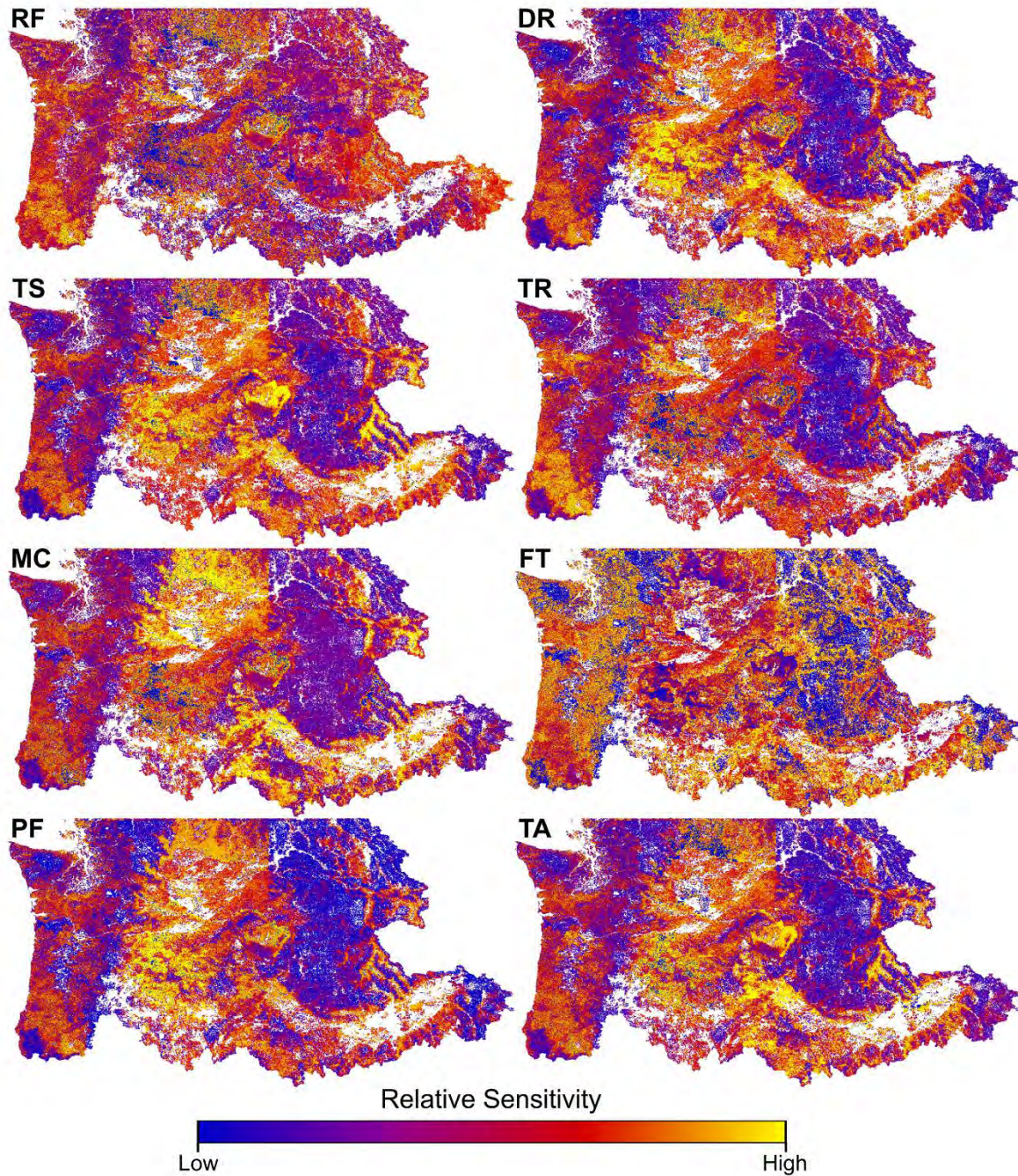


Figure 3.4. Spatial distribution of stream permanence classification (permanent or nonpermanent) to the monthly water balance model parameters presented in Table 1.

#### *Model Precision Analysis*

##### Parameter Regionalization

The K-means classification of parameter sensitivities and the number of observational data points available for each group best supported eight calibration groups (Table 2). Headwater streams

in groups 1-5 exhibited relatively high sensitivity to RF and FT and moderate sensitivity to all other parameters (Figure 3.5). Streams in group 6 exhibited highly variable sensitivities to FT and PF and low sensitivities to all other parameters. Headwater streams in groups 7-8 showed high sensitivity to all parameters, especially FT. Observations were not distributed equally between calibration groups (Table 2). The majority of observations occurred on stream reaches in calibration groups 1 and 8. Most headwater streams in Idaho fall in these calibration groups and the majority of observations were located in Idaho. Only 2.4% of headwaters stream reaches were assigned to calibration group 4, and only 18 observations fell in this group.

Table 3.2. Number of wet and dry observations used to calibrate the monthly water balance model for each calibration group.

ID	%HW	Observations	
		Dry	Wet
1	26.7	425	525
2	13.6	41	119
3	13.2	25	145
4	2.4	3	15
5	17.1	98	202
6	13.8	42	68
7	8.2	102	100
8	5.0	192	255

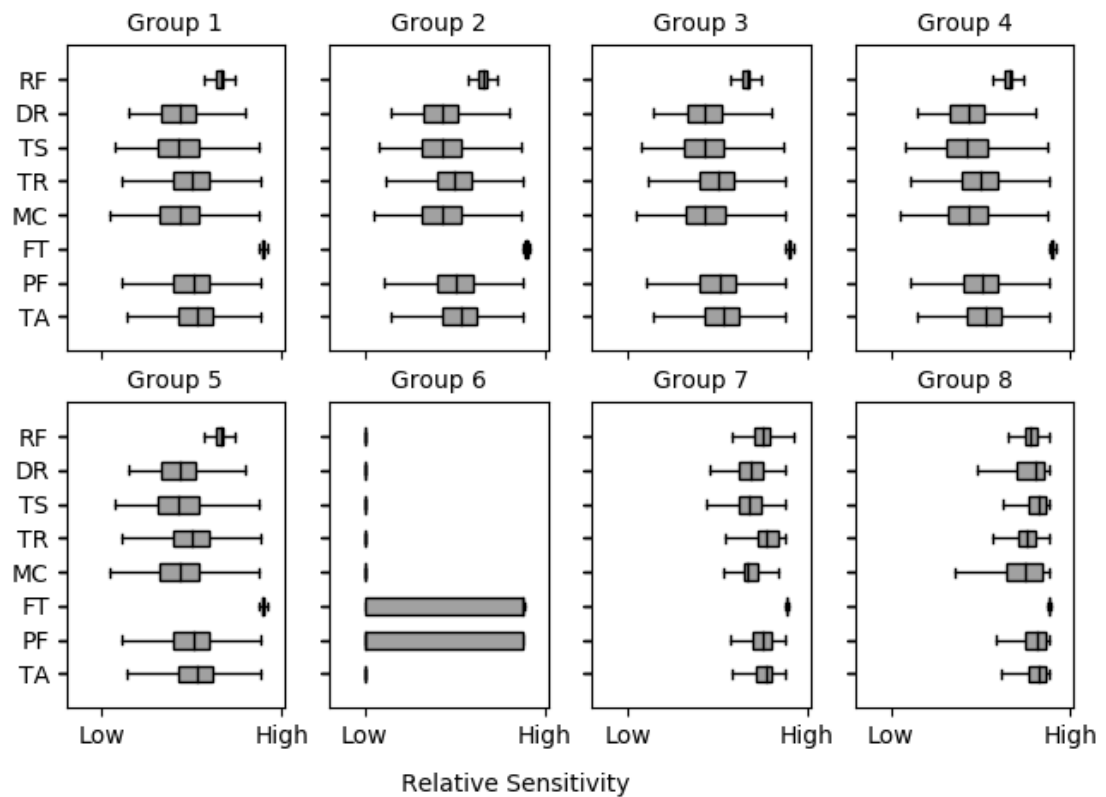


Figure 3.5. Sensitivity of streamflow estimates to each parameter in Table 1 for each calibration group.

Calibration groups also exhibited geographical similarities and spatial autocorrelation (Figure 3.6). Group 8 tended to represent mountainous regions of the interior and some mountainous, coastal regions. Groups 2, 3 and 5 tended to represent coastal mountains. Groups 1, 6, and 7 tended to represent headwater streams in the plains, foothills, and plateaus of the interior. While geographical similarities are apparent within calibration groups it is also apparent that headwater streams near to each other may exhibit different sensitivities to model parameters.



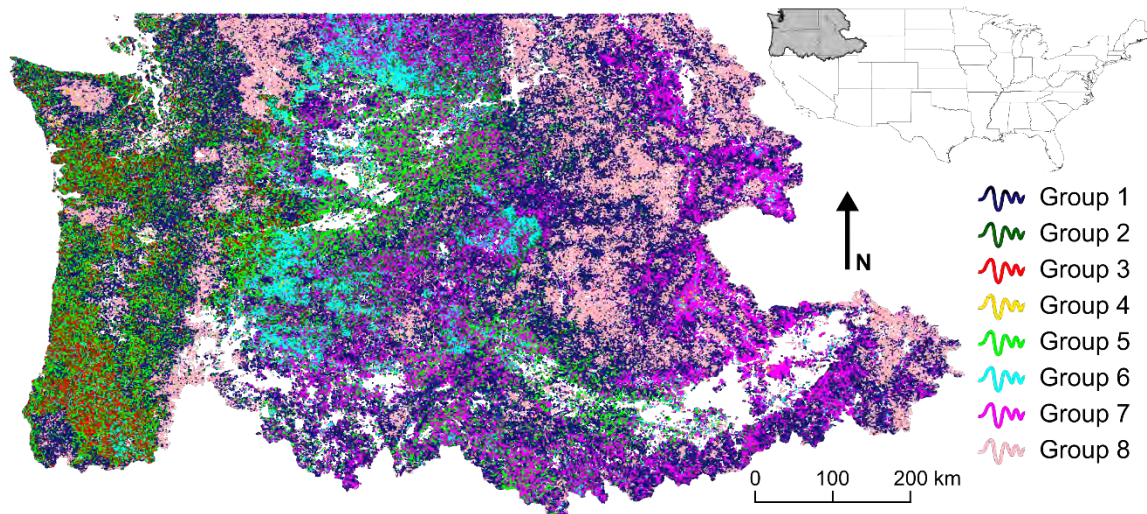


Figure 3.6. Spatial occurrence of stream reaches in each calibration group as determined by the parameter regionalization methods.

Calibration groups displayed some physiographic differences. Drainage areas for groups 2-6 were quite small ( $<0.25 \text{ km}^2$ ), but these groups were separated by differences in elevation and annual precipitation (Figure 3.7). While there is overlap in catchment area, maximum elevation, and annual precipitation between the eight calibration groups, it is also apparent that the groups are somewhat distinct from each other. For example, Group 8 contained the largest catchments and had the greatest average maximum elevation. Group 4 primarily represented small, arid catchments. Groups 3 and 5 were similar in catchment size and elevation and had only slightly different precipitation distributions (Figure 3.7).

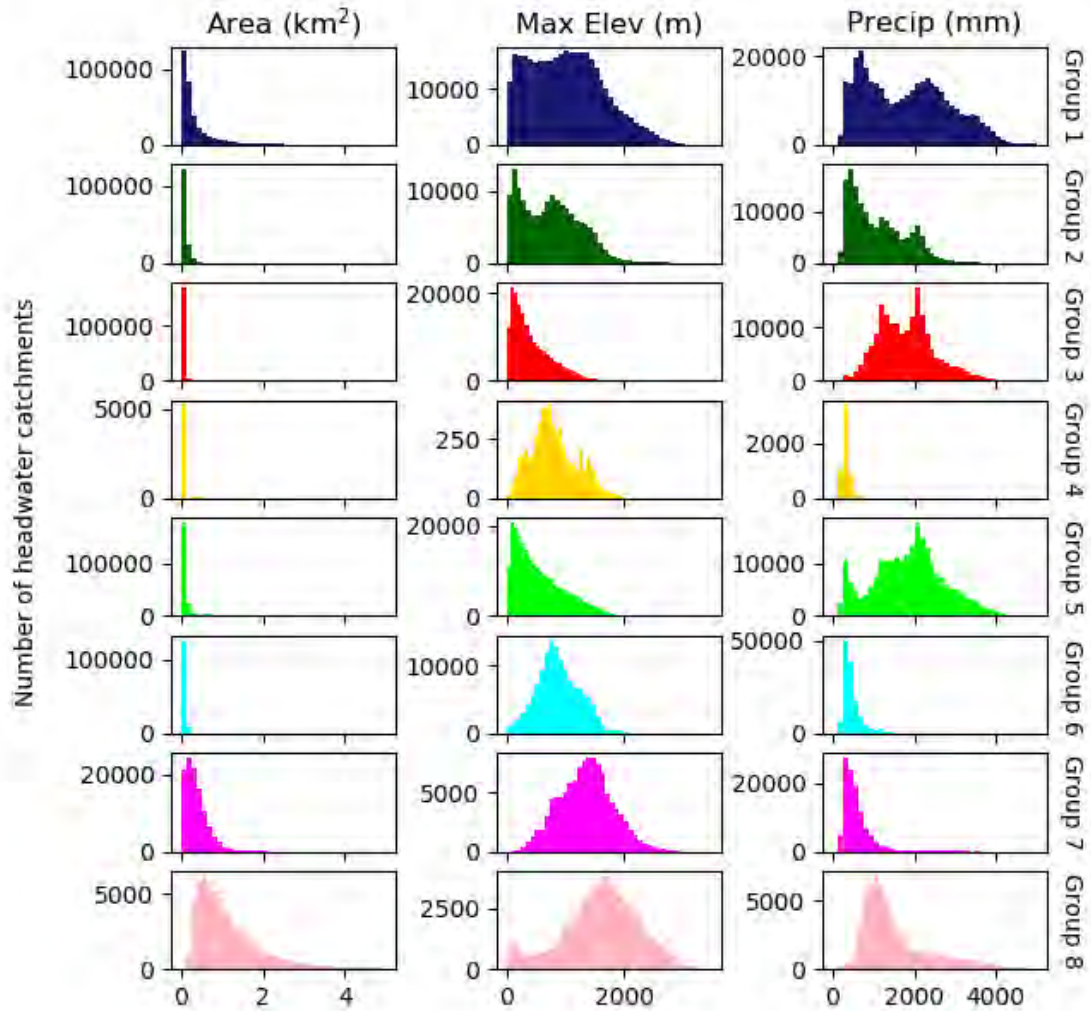


Figure 3.7. Distributions of drainage area, maximum elevation, and total annual precipitation for headwater catchments in each calibration group.

#### Parameter Set Selection

We identified multiple ( $n = 13-92$ ), suitable parameter sets for calibration groups 2-6 (Table 3). Overall accuracy ranged from 65-90%. The highest accuracy was simulated for group 4, which only contained 18 observations. No parameter sets met accuracy constraints for calibration groups 1, 7, and 8 (Table 3). The majority of surface water presence observations occur in 1, 7, and 8 and these groups represent 39.9% of headwater streams in the PNW.

Suitable parameters values for groups 2-6 occurred across nearly the entire distribution for each parameter, with some distinctions between calibration groups (Figure 3.8). The exception was FT, which tended to values near zero. RF values tended to be less than 0.5, except for group 5 where a more uniform distribution was observed between 0.3 and 1.0. TS for groups 2 and 6 tended towards zero, while the distribution was relatively uniform for the other calibration groups. Overall, the relatively uniform distributions of suitable parameters appear to indicate a high degree of equifinality in model parameterization.

Table 3.3. Number and accuracy range of parameter sets for each calibration region that met accuracy constraints for annual and monthly simulations.

ID	n	Annual Accuracy		
		Dry	Wet	Overall
1	0	-	-	-
2	32	0.61 - 0.78	0.61 - 0.72	0.65 - 0.70
3	13	0.60 - 0.76	0.64 - 0.74	0.65 - 0.73
4	92	0.67 - 1.0	0.67 - 0.93	0.67 - 0.89
5	41	0.60 - 0.74	0.61 - 0.75	0.65 - 0.71
6	19	0.67 - 0.79	0.60 - 0.68	0.65 - 0.69
7	0	-	-	-
8	0	-	-	-

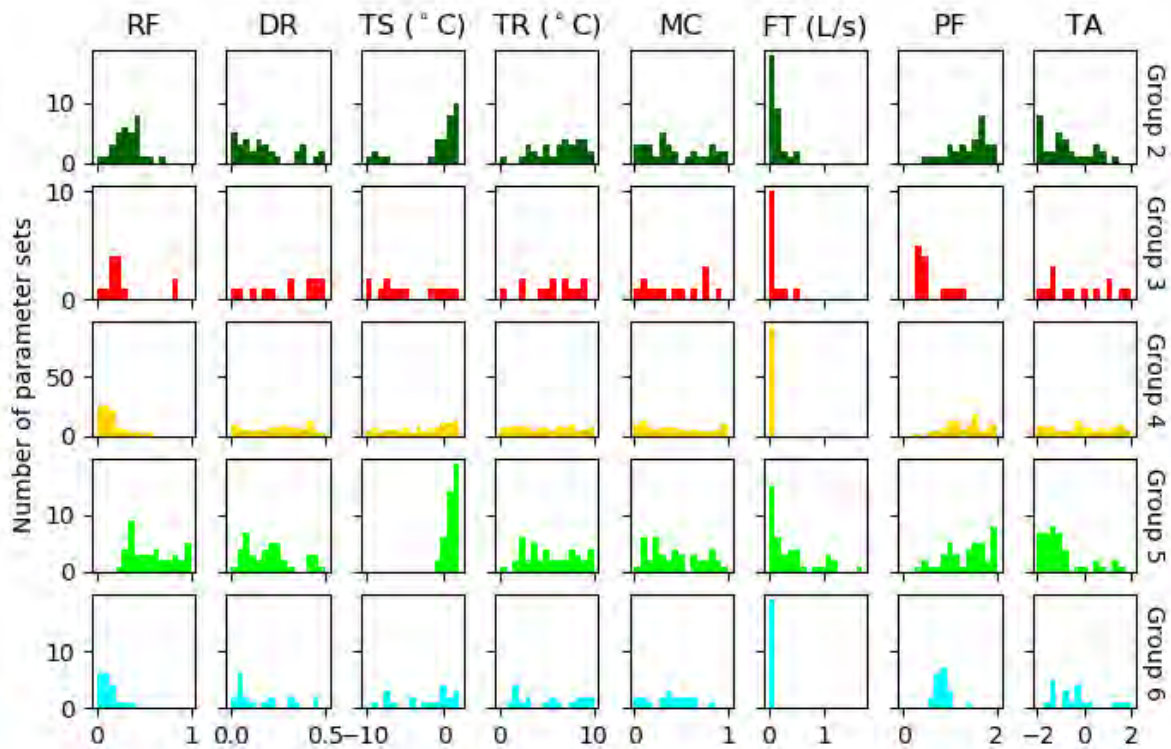


Figure 3.8. Distributions of parameter values for parameter sets that met benchmark constraints. The x-axis represents the sample range of parameter values (Table 1), except for FT, which is subset to 0.0-2.0 L/s to represent the full range of suitable values determined during calibration.

### Model Precision

Model precision for permanent and non-permanent streams varied with climate conditions. In general, model precision was higher for permanent streams in wet years and for non-permanent streams in dry years (Figure 3.9). Through time, model precision was highest and most stable for calibration group 6 (Figure 3.10). Group 3 exhibited a high amount of variability year-to-year. Precision was higher for permanent streams in groups 2-5 than for non-permanent streams. Group 6 showed slightly higher precision for non-permanent streams but was more balanced overall.



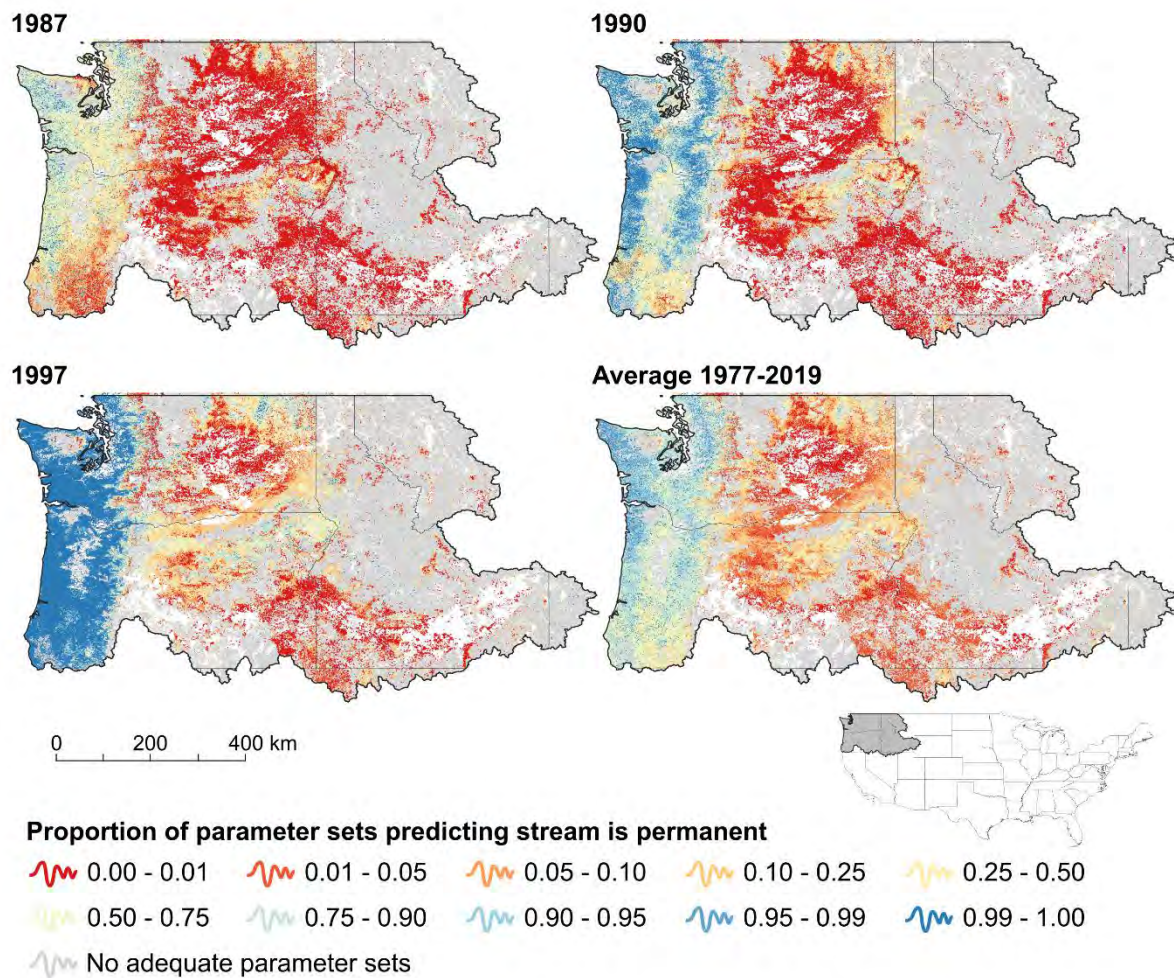


Figure 3.9. Model precision estimates for headwater streams in the Pacific Northwest during a drier than normal year (1987), approximately normal year (1990), wetter than normal year (1997), and the average for the modeled period (1977-2019).



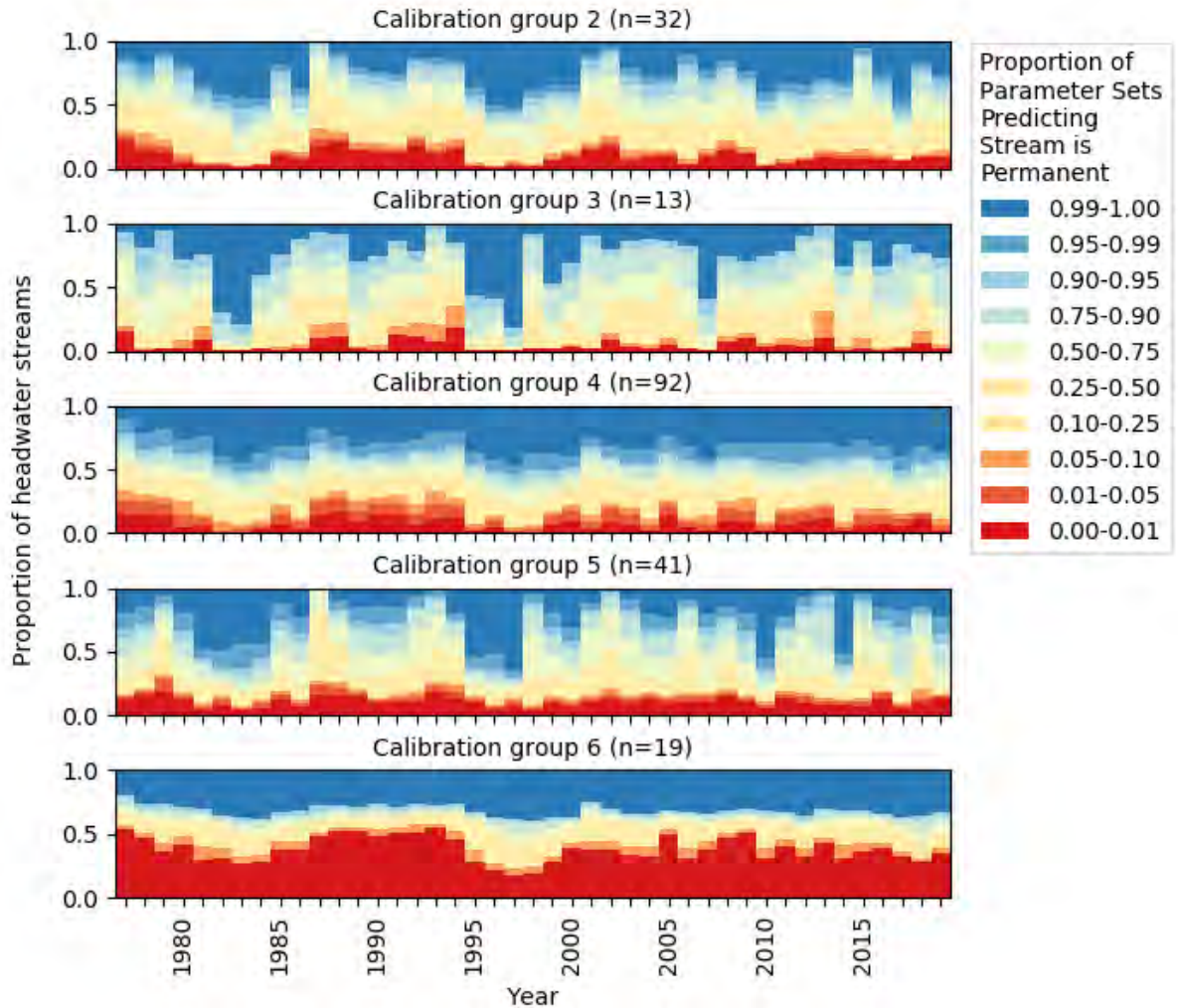


Figure 3.10. Estimated model precision for each calibration group during each year of the study period, where n is in the number of suitable parameter sets tested for each calibration group.

For the entire PNW region, model precision was generally higher for permanent classifications than non-permanent classifications. However, precision for non-permanent streams was less variable (Figure 3.10). Overall, the percentage of headwater streams in the PNW where model precision exceeded 90% ranged from 28-45%. Forty percent of headwater streams in the PNW did not have suitable parameterization so, each year, 15-32% of modeled stream reaches had questionable precision (Figure 3.11). These results only describe how well different model parameterizations agreed with each other and not the accuracy of each parameterization.

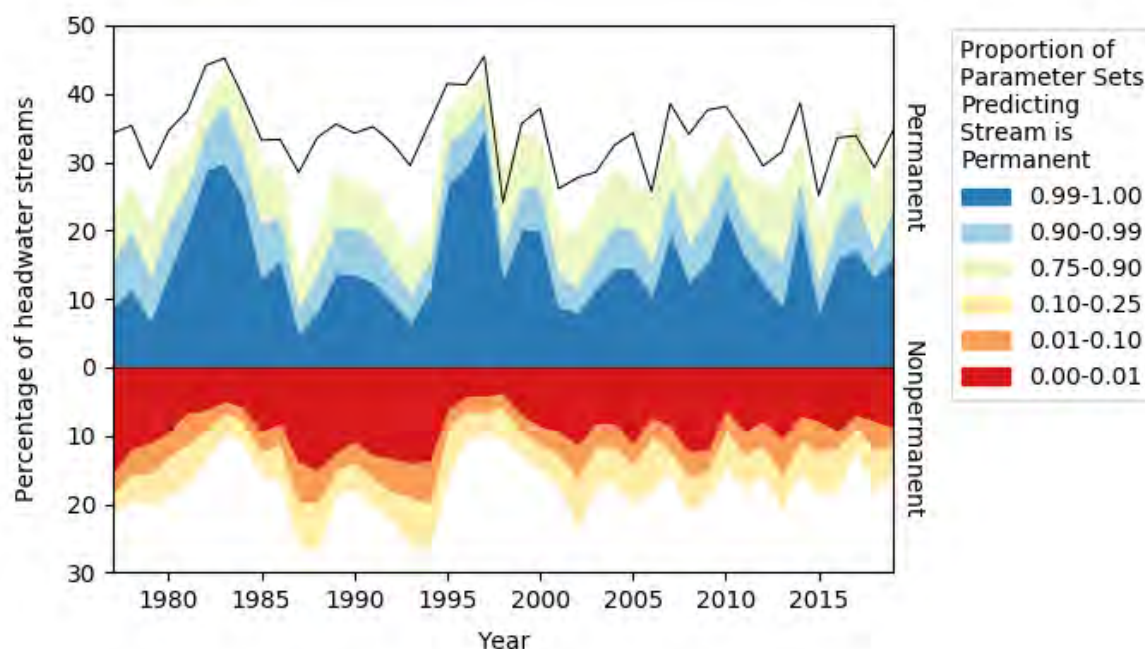


Figure 3.11. Estimated model precision summed for the entire study area (PNW) for each year of the study. The black line is the total percentage of headwater streams with greater than 90% precision. That is, the sum of the two darkest red and darkest blue areas on the plot.

### Discussion

The primary objective of this study was to evaluate if the MWBM could generate SPC estimates with similar accuracy to NHD SPC, but that were annually dynamic instead of static. Previous studies that calibrated the MWBM to discharge data from stream gages produced good results for large streams over regional and national extents (e.g. Bock et al., 2016; McCabe and Wolock, 2011). These previous results indicated the MWBM has potential for modeling stream permanence over similar spatial extents for headwater streams but needed observational data to validate this possibility. Additionally, by adding a parameter to the MWBM we were able to generate dynamic stream permanence estimates for all headwater streams in the PNW. Our methods and results highlight important considerations for future collection of SWPO and stream permanence modeling.

MWBM accuracy was best for calibration groups with smaller catchments. The largest headwater catchments in the PNW were represented by calibration groups 1, 7, and 8 and no suitable parameter sets were identified for any of those calibration groups (Table 3). Meanwhile, multiple parameter sets produced overall accuracies greater than 65% for calibration groups 2-7, when compared to SWPO. The maximum catchment area for groups 2-7 was about 0.5 km<sup>2</sup>. Groups 1, 7, and 8 all had catchments greater than 2 km<sup>2</sup>. Calibration groups 1, 7, and 8 also represented greater ranges of catchment area than groups 2-6.

Stream permanence estimates generated from the suitable parameter combination had high precision for approximately 40% of headwater stream reaches in the Pacific Northwest (Figure 3.11). Of the remaining headwater reaches, suitable parameter combinations were not found for 40% and approximately 20% had low, or variable, precision. Precision was more variable for calibration groups 2, 3, and 5. These groups exhibited higher annual precipitation, and a greater range of annual precipitation, than groups 4 and 6 (Figure 3.6).

Better estimates for calibration groups that represented smaller ranges of catchment area and annual precipitation may indicate that more calibration groups would result in better parameterization. Accuracy and precision results were best for calibration groups 4 and 6, which, generally, represented small, arid catchments. Calibration groups 2, 3, and 5, which represented small catchments across a wide range of annual precipitation, had good accuracy results, but more variable precision estimates. The spatial scale of PRISM data (~4 km) may not account for important heterogeneity in climate inputs at the scale of some headwater catchments. As shown in Figure 3.3, one PRISM grid cell often represented temperature and precipitation for multiple headwater catchments. Because topography is not an input to the MWBM, catchments are represented by climatic conditions and area. Other studies have also indicated that PRISM data do not always capture climatic heterogeneity in mountainous regions (Blasch *et al.*, 2006; Strachan and Daly, 2017). Thus, differences in catchment morphology could be an important factor contributing to stream permanence, and incorporation of downscaled climate data or other remotely sensed data may aid in stream permanence classifications in headwater catchments.

Calibration groups were determined by how catchments responded to changes in model parameters (i.e. parameter sensitivity). Given these results, it would be useful to test a regionalization method based on physiographic characteristics of catchments. For example, grouping catchments based on a suite of factors like geology, temperature, and precipitation regime. We did not take this approach because, while a thorough dataset of physiographic variables has been linked to NHDPlus (medium resolution) catchments (Hill *et al.*, 2016), a similar dataset does not yet exist for NHDPlus HR catchments, which better represent headwater streams. This could be an area of important future research that would aid in regional and continental hydrologic studies. Additionally, the number of calibration groups we considered was limited by the number of available SWPO available for accuracy assessment in each calibration group. Increasing the number of SWPO would make it possible to consider additional calibration groups that are more similar.

Sparse SWPO may have also prevented adequate MWBM calibration for some calibration groups. While SWPO represented many more headwater locations than stream gages, there were no

repeat observations at SWPO locations. This results in a space-for-time substitution, which assumes observations made in different locations account for the range of conditions stream reaches experience through time and assumes that catchment physiography and climate variability are well represented. As noted above, calibration groups representing larger ranges of catchment area and total annual precipitation had poorer performance than those with narrower ranges, indicating catchment physiography and climate variability were not well represented by SWPO in each calibration group.

Furthermore, SWPO were collected by multiple agencies for multiple uses using different methods. Observations were often obtained as ancillary information to other objectives and, thus, studies were not designed to create a robust dataset for modeling purposes. Because few SWPO were available we did not conduct a cross-validation accuracy assessment for each calibration group, but instead used multiple suitable parameter sets to assess model precision. A larger SWPO dataset would support and encourage more robust accuracy assessment. Indeed, other stream permanence modeling efforts, conducted over smaller spatial extents, were able to achieve high accuracies with fuller underlying datasets. This points to the need for increased collection of surface water presence data to support future NHD SWP modeling efforts.

Some hydrological processes and topographic features that influence surface water presence are not explicitly represented by the MWBM. Specifically, baseflow is often the major contribution to summer streamflow, especially in mountainous regions. However, in the MWBM, baseflow is lumped into the RF parameter, which also includes runoff contributions from lateral flow and saturation excess flow. Timing of water delivery to the stream channel is different for each of these mechanisms. With a monthly time step it may be reasonable to lump lateral flow and saturation excess flow, but the MWBM may benefit from explicit representation of baseflow for estimating stream permanence. Additionally, there is no representation of valley-bottom width, depth, or topography in the MWBM. Valley-bottom geomorphology and soil characteristics control the ability to transport water in the subsurface. Valley-bottom representation should be an important inclusion in future stream permanence models.

While the MWBM produced precise results for only 40% of headwater streams in the PNW it is important to consider these results in context. Ungauged headwater streams can be difficult to model even when abundant data are present to characterize the catchment (Wagner *et al.*, 2004). The fact that the MWBM generated precise results for 40% of headwater catchments indicates the potential for development of regional stream permanence models. Inputs to the MWBM are precipitation, temperature, and soil water holding capacity. While the simple MWBM has its shortcomings for this application, results indicate potential for future development of simple, stream permanence models.

The shortcomings identified for this particular use of the MWBM highlight important considerations and processes to include in future modeling efforts. Development of predictive stream permanence models is extremely important for assessment of how changes to land cover and climate may influence stream regulation and stream ecology in the United States and worldwide.

### **Conclusion**

The sheer number of headwater streams in the US precludes data collection on each individual stream. Thus, dynamic estimates of stream permanence in the headwaters will require modeling approaches to assess locations where data cannot be collected. We modified a MWBM by adding a simple flow threshold model to estimate stream permanence on over 1.3 million headwater streams in the PNW, then benchmarked MWBM stream permanence estimates against observed accuracies of NHD SPC to assess precision of the MWBM estimates.

For three of eight calibration groups – about 40% of headwater streams – no parameter combinations produced stream permanence results at least as accurate as NHD SPC. Precision of MWBM estimates for suitable parameter sets was low on an additional 20% of headwater streams. Thus, the MWBM may be suitable to estimate stream permanence for approximately 40% of headwater streams in the PNW, given the data and methods used in this study. MWBM precision was best for calibration groups with narrow ranges for catchment area and total annual precipitation.

More SWPO that are intentionally collected for modeling studies could greatly improve model development and calibration. The number of SWPO available for this study was much less than the number of modeled stream reaches. Thus, the characteristics of some stream reaches may not have been represented by the observational data used to calibrate the MWBM and assess its accuracy. Additionally, the MWBM does not explicitly represent baseflow and valley-bottom processes which can be important for stream permanence determinations. Future modeling efforts should consider the effect of these processes.

Use of a simple MWBM with primary inputs of monthly precipitation and temperature produced precise results with comparable accuracy to the NHD standard for 40% of headwater streams in the PNW. While these results indicate the MWBM is not suitable to consistently estimate stream permanence for headwater streams, they do indicate potential for predictive models to produce reliable stream permanence estimates. Lessons learned from this modeling application will aid in development of future stream permanence models over large spatial extents.

### Literature Cited

- Beaman, W.M., 1928. Topographic Instructions of the United States Geological Survey. Topographic Mapping. U.S. Geological Survey. doi:10.3133/b788E.
- Beaufort, A., J. Carreau, and E. Sauquet, 2019. A Classification Approach to Reconstruct Local Daily Drying Dynamics at Headwater Streams. *Hydrological Processes* 33:1896–1912.
- Beck, H.E., A.I. van Dijk, A. De Roo, D.G. Miralles, T.R. McVicar, J. Schellekens, and L.A. Bruijnzeel, 2016. Global-scale Regionalization of Hydrologic Model Parameters. *Water Resources Research* 52:3599–3622.
- Blasch, K.W., J.P. Hoffmann, L.F. Graser, J.R. Bryson, and A.L. Flint, 2006. Hydrogeology of the Upper and Middle Verde River Watersheds, Central Arizona. Scientific Investigations Report, U. S. Geological Survey.
- Bock, A.R., W.H. Farmer, and L.E. Hay, 2018. Quantifying Uncertainty in Simulated Streamflow and Runoff from a Continental-Scale Monthly Water Balance Model. *Advances in Water Resources* 122:166–175.
- Bock, A.R., L.E. Hay, G.J. McCabe, S.L. Markstrom, and R.D. Atkinson, 2016. Parameter Regionalization of a Monthly Water Balance Model for the Conterminous United States. *Hydrology and Earth System Sciences* 20:2861–2876.
- DeWeber, J.T., Y.-P. Tsang, D.M. Krueger, J.B. Whittier, T. Wagner, D.M. Infante, and G. Whelan, 2014. Importance of Understanding Landscape Biases in USGS Gage Locations: Implications and Solutions for Managers. *Fisheries* 39:155–163.
- Fischer, B.M., M. Rinderer, P. Schneider, T. Ewen, and J. Seibert, 2015. Contributing Sources to Baseflow in Pre-alpine Headwaters Using Spatial Snapshot Sampling. *Hydrological Processes* 29:5321–5336.
- Fritz, K.M., E. Hagenbuch, E. D’Amico, M. Reif, P.J. Wigington, S.G. Leibowitz, R.L. Comeleo, J.L. Ebersole, and T.-L. Nadeau, 2013. Comparing the Extent and Permanence of Headwater Streams From Two Field Surveys to Values From Hydrographic Databases and Maps. *JAWRA Journal of the American Water Resources Association* 49:867–882.
- Fritz, K.M., B.R. Johnson, and D.M. Walters, 2008. Physical Indicators of Hydrologic Permanence in Forested Headwater Streams. *Journal of the North American Benthological Society* 27:690–704.
- Gendaszek, A.S., J.B. Dunham, C.E. Torgersen, D.P. Hockman-Wert, M.P. Heck, J. Thorson, J. Mintz, and T. Allai, 2020. Land-Cover and Climatic Controls on Water Temperature, Flow Permanence, and Fragmentation of Great Basin Stream Networks. *Water* 12:1962.
- Godsey, S.E. and J.W. Kirchner, 2014. Dynamic, Discontinuous Stream Networks: Hydrologically Driven Variations in Active Drainage Density, Flowing Channels and Stream Order. *Hydrological Processes* 28:5791–5803.
- Guptill, S.C., 1990. An Enhanced Digital Line Graph Design. USGPO; Books and Open-File Reports Section, US Geological Survey,.
- Hafen, K.C., K.W. Blasch, A. Rea, R. Sando, and P.E. Gessler, 2020. The Influence of Climate Variability on the Accuracy of NHD Perennial and Nonperennial Stream Classifications. *JAWRA Journal of the American Water Resources Association* 56:903–916.

- Hamon, W., 1961. Estimating Potential Evaporation. JoH Division., pp. 107–120.
- Herman, J. and W. Usher, 2017. SALib: An Open-Source Python Library for Sensitivity Analysis. *Journal of Open Source Software* 2:97.
- Hester, G., K. Carsell, and D. Ford, 2006. Benefits of USGS Streamgaging Program—Users and Uses of USGS Streamflow Data. *National Hydrologic Warning Council*:1–18.
- Hill, R.A., M.H. Weber, S.G. Leibowitz, A.R. Olsen, and D.J. Thornbrugh, 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *JAWRA Journal of the American Water Resources Association* 52:120–128.
- Jaeger, K.L., J. Burnett, E.D. Heaston, S.M. Wondzell, N. Chelgren, J.B. Dunham, S. Johnson, and M. Brown, 2020. FLOWPER User Guide—For Collection of FLOW PERmanence Field Observations. Open-File Report, US Geological Survey. <https://pubs.er.usgs.gov/publication/ofr20201075>.
- Jaeger, K.L., R. Sando, R.R. McShane, J.B. Dunham, D.P. Hockman-Wert, K.E. Kaiser, K. Hafen, J.C. Risley, and K.W. Blasch, 2019. Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest. *Journal of Hydrology X* 2:100005.
- Jensen, C.K., K.J. McGuire, and P.S. Prince, 2017. Headwater Stream Length Dynamics across Four Physiographic Provinces of the Appalachian Highlands. *Hydrological Processes* 31:3350–3363.
- Kosugi, K., S. Katsura, M. Katsuyama, and T. Mizuyama, 2006. Water Flow Processes in Weathered Granitic Bedrock and Their Effects on Runoff Generation in a Small Headwater Catchment. *Water Resources Research* 42.
- Lowe, W.H. and G.E. Likens, 2005. Moving Headwater Streams to the Head of the Class. *BioScience* 55:196–197.
- McCabe, G.J. and S.L. Markstrom, 2007. A Monthly Water-Balance Model Driven by a Graphical User Interface. Geological Survey (US).
- McCabe, G.J. and D.M. Wolock, 2007. Warming May Create Substantial Water Supply Shortages in the Colorado River Basin. *Geophysical Research Letters* 34. doi:10.1029/2007GL031764.
- McCabe, G.J. and D.M. Wolock, 2008. Joint Variability of Global Runoff and Global Sea Surface Temperatures. *Journal of Hydrometeorology* 9:816–824.
- McCabe, G.J. and D.M. Wolock, 2011. Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States. *Water Resources Research* 47. doi:10.1029/2011WR010630.
- McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012. NHDPlus Version 2: User Guide. US Environmental Protection Agency.
- McShane, R.R., R. Sando, and P.D. Hockman-Wert, 2017. Streamflow Observation Points in the Pacific Northwest, 1977-2016. U.S. Geological Survey. <https://doi.org/10.5066/F7BV7FSP>.
- Miller, M.P., D.M. Carlisle, D.M. Wolock, and M. Wiczorek, 2018. A Database of Natural Monthly Streamflow Estimates from 1950 to 2015 for the Conterminous United States. *JAWRA Journal of the American Water Resources Association* 54:1258–1269.

- Moore, R.B., L.D. McKay, A.H. Rea, T.R. Bondelid, C.V. Price, T.G. Dewald, and C.M. Johnston, 2019. User's Guide for the National Hydrography Dataset plus (NHDPlus) High Resolution. US Geological Survey.
- Nadeau, T.-L., G.S. Leibowitz, J.P. Wigington, L.J. Ebersole, M.K. Fritz, A.R. Coulombe, L.R. Comeleo, and A.K. Blocksom, 2015. Validation of Rapid Assessment Methods to Determine Streamflow Duration Classes in the Pacific Northwest, USA. *Environmental Management* 56:34–53.
- Nadeau, T.-L. and C.M. Rains, 2007. Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy. *Journal of the American Water Resources Association* 43. <http://doi.wiley.com/10.1111/j.1752-1688.2007.00010.x>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, 2011. Scikit-Learn: Machine Learning in Python. *The Journal of Machine Learning Research* 12:2825–2830.
- PRISM Climate Group, Oregon State University. <http://prism.oregonstate.edu/>. Accessed 31 Oct 2018.
- Schwarz, G.E. and R. Alexander, 1995. State Soil Geographic (STATSGO) Data Base for the Conterminous United States. USGS Numbered Series. <https://pubs.er.usgs.gov/publication/ofr95449>.
- Seaber, P.R., F.P. Kapinos, and G.L. Knapp, 1987. Hydrologic Unit Maps.
- Strachan, S. and C. Daly, 2017. Testing the Daily PRISM Air Temperature Model on Semiarid Mountain Slopes. *Journal of Geophysical Research: Atmospheres* 122:5697–5715.
- USGS Advanced Research Computing, USGS Yeti Supercomputer. U.S. Geological Survey. <https://doi.org/10.5066/F7D798MJ>.
- Vander Vorste, R., R. Sarremejane, and T. Datry, 2020. Intermittent Rivers and Ephemeral Streams: A Unique Biome With Important Contributions to Biodiversity and Ecosystem Services.
- Wagener, T., H. Wheeler, and H.V. Gupta, 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. ImperiL College Press.
- Ward, A.S., N.M. Schmadel, and S.M. Wondzell, 2018. Simulation of Dynamic Expansion, Contraction, and Connectivity in a Mountain Stream Network. *Advances in Water Resources* 114:64–82.
- Ward, A.S., S.M. Wondzell, N.M. Schmadel, and S.P. Herzog, 2020. Climate Change Causes River Network Contraction and Disconnection in the HJ Andrews Experimental Forest, Oregon, USA. *Front. Water* 2: 7. Doi: 10.3389/Frwa.
- Williamson, T.N., C.T. Agouridis, C.D. Barton, J.A. Villines, and J.G. Lant, 2015. Classification of Ephemeral, Intermittent, and Perennial Stream Reaches Using a TOPMODEL-Based Approach. *JAWRA Journal of the American Water Resources Association* 51:1739–1759.
- Yu, S., N.R. Bond, S.E. Bunn, Z. Xu, and M.J. Kennard, 2018. Quantifying Spatial and Temporal Patterns of Flow Intermittency Using Spatially Contiguous Runoff Data. *Journal of Hydrology* 559:861–872.
- Yu, S., H.X. Do, A.I. van Dijk, N.R. Bond, P. Lin, and M.J. Kennard, 2020. Evaluating a Landscape-Scale Daily Water Balance Model to Support Spatially Continuous Representation of Flow



Intermittency throughout Stream Networks. *Hydrology and Earth System Sciences* 24:5279–5295.

## **Chapter 4: Estimating Stream Permanence with the Watershed Erosion Prediction Project Model: Implications for Surface Water Presence Modeling and Data Collection**

### **Abstract**

Process-based models have been widely useful to simulate how water quantity may be influenced by various drivers at local, regional, national, and global extents. Results from these models are valuable for water resources management but are less helpful for determining the regulatory status of streams, which depends on surface water presence, not quantity. Application of process-based models to simulate stream permanence has only occurred at local scales and models have been validated on data from just a handful of stream reaches. Herein, the Watershed Erosion Prediction Project (WEPP) hydrological model is applied to watersheds in the humid H. J. Andrews Experimental Watershed (HJA) and the more arid Willow and Whitehorse watersheds (WW), both in Oregon, to simulate daily and annual stream permanence. One thousand parameter combinations were tested to calibrate WEPP to observed streamflow in HJA watersheds and one-hundred parameter combinations were tested to calibrate WEPP to observed surface water presence time-series in WW watersheds. When calibrated to observed streamflow, WEPP correctly classified annual stream permanence for 83% of HJA stream reaches. In the WW, WEPP simulations correctly classified 63–82% of daily stream permanence observations and 59–87% of annual stream permanence classifications. Inclusion of a dry-day threshold (the maximum number of days a stream reach could be modeled ‘dry’ but still classified as permanent for the year) improved annual accuracy in three of four WW watersheds. Parameter sets producing the best daily accuracies in WW did not produce suitable annual accuracies. Results indicate the importance of evaluating process-based models on both permanent and nonpermanent streams at daily and annual time scales. Additionally, results suggest that strategic collection of surface water presence observations will be essential for robust calibration of process-based models to simulate stream permanence moving forward.

### **Introduction**

The majority of land surface area in any watershed drains to headwater streams (Downing et al., 2012). Because of this, water quality regulation in headwater streams has important implications for land use practices and downstream water quality and quantity (Alexander et al., 2007). Throughout the last century, research related to water resources has focused on modeling and characterizing various components of water quantity in larger basins. More recently, the ecological importance of

water resources has received attention with respect to maintaining habitats for sensitive, threatened, and endangered species. Once again, the focus has been on maintaining an appropriate quantity and quality of water for organisms to survive. However, the policies that govern whether a stream receives regulatory protection (i.e. is subject to water quality standards) under the US Clean Water Act is determined by the presence or absence of surface water in a stream reach, not by the quantity of water.

In the US, the majority of data quantifying surface water in streams comes from the USGS stream gage network. The primary priorities of this network are to assess water availability, flood risk, and climate-related trends (DeWeber et al., 2014; Hester et al., 2006). As a result, larger rivers are disproportionately represented, while fewer data exist on smaller, headwater streams (DeWeber et al., 2014). Experimental watersheds and long-term ecological research sites complement the USGS gage network by collecting long-term data over smaller spatial extents that are more representative of headwater systems. Like the USGS gages network, these experimental stations are generally designed to quantify surface water by monitoring streamflow magnitude at gauged locations. The type of data collected at stream gages, and their geographic locations, have largely influenced model development by determining data and spatial scales available for model validation. Of course, data collection and modeling are driven by the important need to forecast flooding, drought, and water availability. As a likely result of the more pressing need to understanding water availability, fewer data collection efforts and modeling studies have focused on identifying when and where surface water is present and the duration of surface water presence in stream channels (Hammond et al., 2020; Jaeger et al., 2019; Jensen et al., 2018; Williamson et al., 2015).

The National Hydrography Dataset (NHD) is the most comprehensive dataset describing stream permanence in the US (Nadeau and Rains, 2007). However, NHD stream permanence classifications (i.e. perennial, intermittent, and ephemeral) are not well structured to represent the dynamic nature of streamflow presence and may exhibit error rates up to 50% on headwater streams (Fritz et al., 2013; Hafen et al., 2020; Nadeau et al., 2015). In the last decade, many data collection and modeling efforts have begun to focus on quantifying, estimating, and predicting stream permanence.

To date (2021), most stream permanence models can be grouped into one of two general categories: 1) regional studies that use sparse data (in relation to the number of streams modeled) to make stream permanence estimates over large spatial extents (e.g. Jaeger et al., 2019; Sando and Blasch, 2015); and 2) local studies that intensively monitor and model streams in small watersheds (e.g. Jensen et al., 2017, 2018; Ward et al., 2018; Williamson et al., 2015). Models in both categories suffer from data-based limitations. At regional extents predictions or estimates are generated for

millions of stream reaches based on only a few observations (Jaeger et al., 2019; Sando and Blasch, 2015). At local scales models are tested with data that quantitatively detail only a small number of stream reaches (Ward et al., 2018; Williamson et al., 2015). Gendaszek et al. (2020) may be the first to produce results at the mesoscale (e.g. 1-10 km<sup>2</sup>) with a spatial statistical network model assessing stream permanence at over 200 monitored stream reaches.

Regional and mesoscale stream permanence modeling efforts have largely used statistical methods to identify relationships between climatic and physiographic variables that influence the presence of surface water in a particular location during a specific time period (Gendaszek et al., 2020; Jaeger et al., 2019; Sando and Blasch, 2015). These methods identify variables that influence stream permanence but the models are not readily adaptable to new locations and time periods. By contrast, implementation of physically-based models to estimate stream permanence has primarily occurred in a few small (<100 ha) watersheds. Theoretically, by representing the physical processes that govern streamflow, physically-based models could be readily applied to new locations, though parameterization is often also necessary. Furthermore, many hydrological models are becoming operationalized on cloud-based platforms making them accessible for use by land and water managers. While physically-based models have shown promise for representing stream permanence over small spatial extents, they have not been widely tested over larger extents. This is partially due to lack of time-series data available to adequately validate performance of simulated stream permanence from physically-based models over these larger extents (Jensen et al., 2018; Ward et al., 2018; Williamson et al., 2015).

The purpose of this work is to assess the performance of the cloud-based version (WEPPCloud) of the physically-based Watershed Erosion Prediction Project (WEPP) hydrological model (Flanagan et al., 2001; Flanagan and Nearing, 1995) to estimate stream permanence in both humid and arid environments of the United States. WEPP was selected because it has shown potential to generate accurate streamflow estimates in ungauged basins (Brooks et al., 2016) and has been implemented to model hydrology in a variety of environments including the hydrological effects of land surface disturbances (Dun et al., 2009; Srivastava et al., 2020, 2013; Zheng et al., 2020). Stream permanence estimates from WEPP are validated with a combination of time-series data from sensors and human observations in the arid Willow and Whitehorse watersheds in the Great Basin of southeast Oregon and in the humid HJ Andrews Experimental Watershed in the Cascade Mountains of western Oregon (Gendaszek et al., 2020; Jones and Grant, 1996; Schultz et al., 2017; Ward et al., 2020). This study presents evaluation of a physically-based hydrological model to estimate stream permanence at the mesoscale.

## Methods

### *Study Areas*

Two sets watersheds in Oregon, which represent different climates, were selected for this watershed. The humid climate of the western Cascades was represented by eight (HJA01, HJA02, HJA03, HJA06, HJA07, HJA08, HJA09, HJA10) of the ten gauged watersheds in the H. J. Andrews Experimental Forest of northwestern Oregon (Figure 1). The HJA has a rich history of hydrological experimentation, research, and debate and equally rich datasets to facilitate future research (Jones and Grant, 1996; Thomas and Megahan, 1998; Ward et al., 2020). Streamflow for all eight watersheds has been continuously gauged since 1995. Drainage areas of the eight gauged watersheds range from 9 ha (HJA09) to 101 ha (HJA03). Despite their small size, all gauged watersheds support perennial flow at the gage sites, though portions of upstream channels and tributaries regularly go dry in summer months. Elevations of the HJA watersheds range from 400-1200 m above sea level. Annual precipitation in the area averages 2300 mm and mean annual temperature is 9.2 deg. C (Ward et al., 2020). The HJA is heavily forested with Douglas fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*) and western red cedar (*Thuja plicata*) with red alder (*Alnus rubra*) occupying open riparian areas (Frady et al., 2007; Jones and Grant, 1996). Comprehensive descriptions of HJA climate, morphology, logging experiments, and geology are well described by others (Dyrness, 1969; Jones et al., 2000; Swanson and Dyrness, 1975; Swanson and James, 1975).

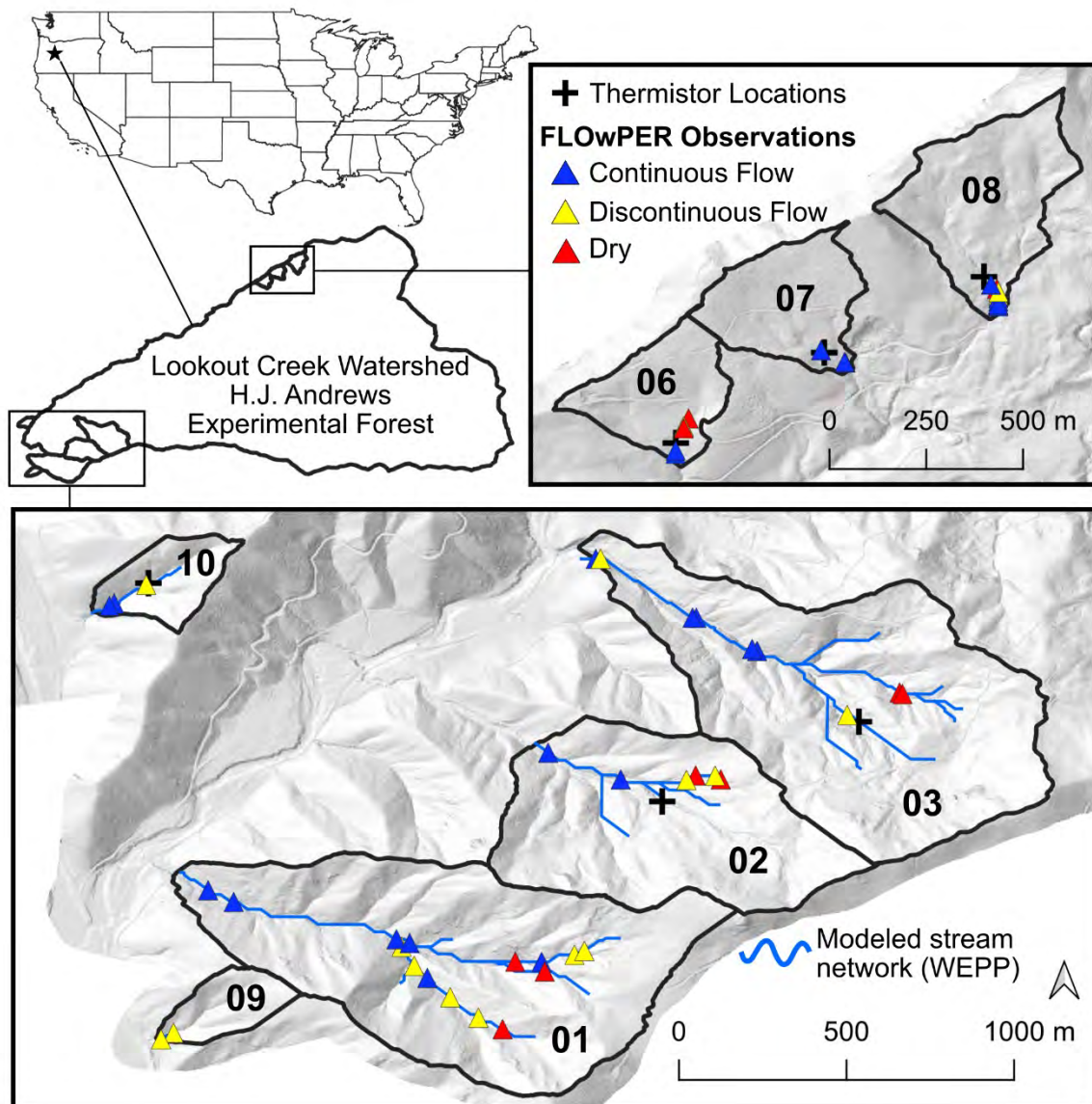


Figure 4.1. Watersheds in the H. J. Andrews Experimental forest where data collection and modeling occurred.

The Willow and Whitehorse watersheds (WW) in southeastern Oregon provide a stark climatological difference to the HJA watersheds. Four watersheds representing portions of Willow and Whitehorse creeks and their tributaries (WW01, WW02, WW03, WW04) were considered for this study (Figure 2). Watersheds were selected to be small enough for easy modeling with WEPP but large enough to represent as many data locations as possible (available data are described below). Areas of the modeled watersheds ranged from 1,594 ha (WW01) to 4,433 ha (WW03). Many WW streams are nonpermanent while a few maintain permanent flow each year (Schultz et al., 2017). Surface water presence was monitored at multiple sites in the WW watersheds from 2011-2017 as

part of studies assessing habitat for Lahontan cutthroat trout (Gendaszek et al., 2020; Schultz et al., 2017). At around 400 mm, average annual precipitation in the WW is nearly 2,000 mm less than in the HJA. Elevation in the WW watersheds ranged from 1,600-2,400 m and mean annual air temperature is 8.1 deg. C. Vegetation in the area is primarily composed of sagebrush (*Artemisia spp.*) with patches of aspen (*Populus tremuloides*) at higher elevations and willow (*Salix spp.*) occurring in riparian areas. Comprehensive details of climate, geology, and landcover of WW are provided elsewhere (Dunham et al., 2003; Gendaszek et al., 2020; Schultz et al., 2017).

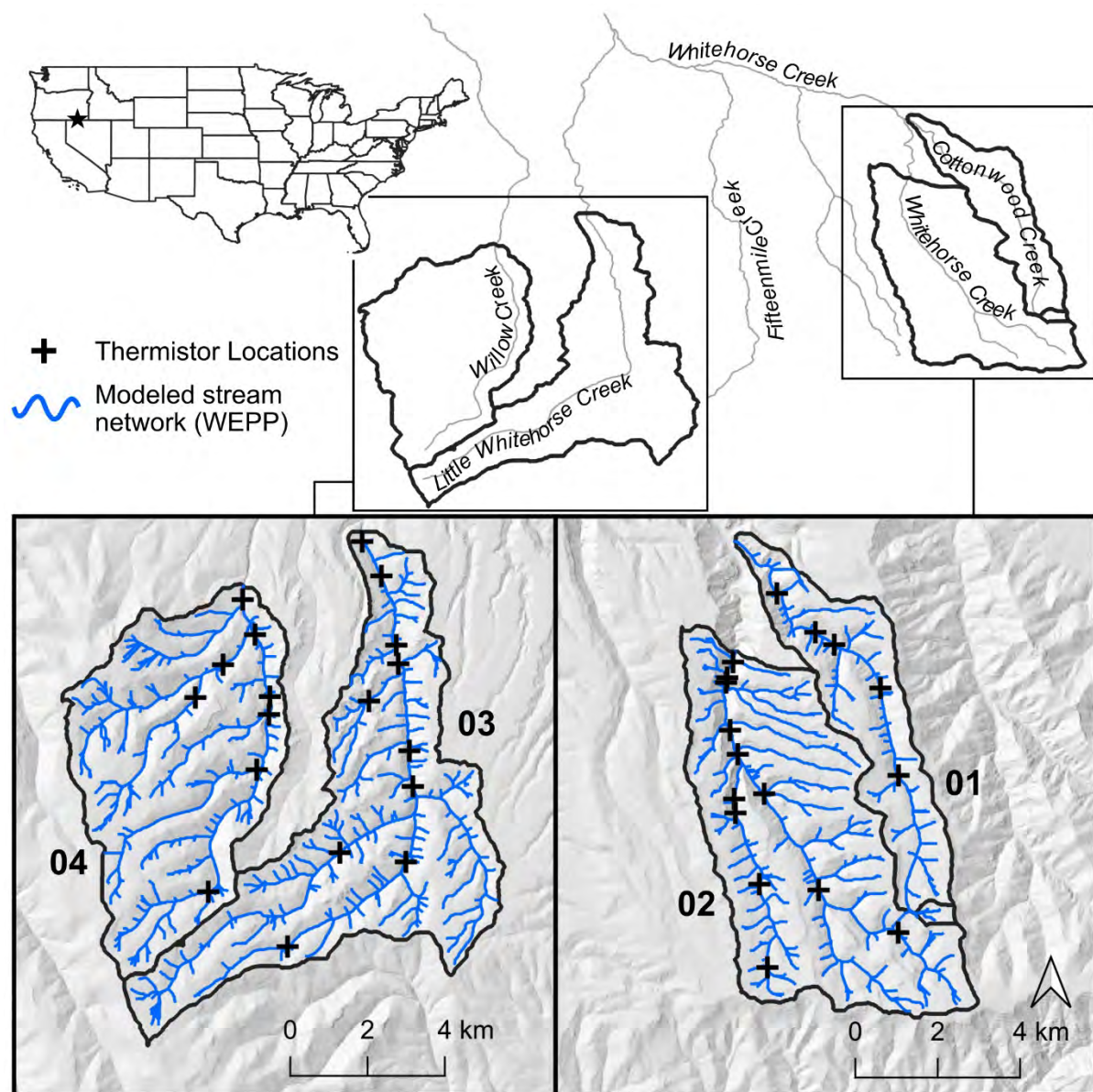


Figure 4.2. Watersheds in the Willow-Whitehorse basin where thermistors were deployed and modeling occurred.

### *Streamflow Data*

Streamflow for all eight HJA watersheds considered in this study was recorded at 15-minute intervals for multiple decades. Minimum, maximum, and mean daily streamflow values were calculated using the 15-minute time series. Based on recommendation of HJA scientists, streamflow data prior to the year 2000 (when the stage-discharge relationships were most recently updated) were not used. Streamflow data were not available for the WW watersheds.

### *Surface Water Presence Data*

During the summer and autumn of 2020 observations of surface water presence were made on stream reaches in gauged HJA watersheds using the USGS Flow Permanence data collection application (FLOwPER). Locations for FLOwPER observations were determined randomly and most locations were observed on two different dates. However, September wildfires prevented access to HJA watersheds during autumn 2020. With FLOwPER, observers designate a point on a stream reach to have ‘continuous flow’ if surface water is present in the stream channel 10 m upstream and downstream of the observer’s location, ‘discontinuous flow’ if water is present but the stream channel also contains channel-spanning segments without surface water, or ‘no flow’ if there is no surface water present in the stream channel. Locations of FLOwPER observations are presented in Figure 1.

In addition to FLOwPER observations, surface water presence was monitored using temperature sensors (thermistors) in HJA gauged watersheds during the summer and autumn of 2020 (Figure 1). The stream reaches where thermistors were placed were randomly selected and thermistors were placed in locations on the stream reach that were expected to go dry at some point during the summer or autumn. Two thermistors were deployed at each site; one in the stream channel, to record water temperature, and one adjacent to the stream channel to record air temperature. Surface water presence is derived from hourly temperature time series by comparing the magnitude and fluctuation of the in-channel thermistor and out-of-channel thermistor (where one exists), or the in-channel thermographs where an out-of-channel sensor was not deployed (Arismendi et al., 2017; Blasch et al., 2002; Gendaszek et al., 2020).

Surface water presence in the WW watersheds was also recorded with thermistors. Thermistors were deployed between 2011 and 2017. Thermistors were initially deployed to evaluate and model temperature and surface water presence for trout habitat in the watershed (Gendaszek et al., 2020; Schultz et al., 2017). Surface water presence in WW watersheds was derived from temperature time series following the same methods used for HJA watersheds.



To avoid potential misclassifications of surface water presence from thermistor data due to frozen streams, only thermistor observations recorded between April 1 and October 31 (inclusive) were used in this study. Hourly thermistor time-series were converted to daily values. Any day a thermistor location was determined to be dry for any hour that day was classified as ‘dry’, or absent surface water. Data were aggregated at the spatial scale of WEPP stream reaches (described below). All thermistor and FLOWPER data on each stream reach for each date were combined so that if any observation indicated absence of surface water at any location on any day the reach was classified as non-permanent, or ‘dry’ on that day. Any reach that had at least one ‘dry’ observation between April 1 and October 31 was classified as non-permanent for that year.

No thermistor observations in HJA recorded surface water absence. However, FLOWPER observations on the same stream reaches as thermistors (but at different locations) recorded ‘discontinuous flow’ or ‘no flow’ conditions on these reaches on the same dates. For the most accurate annual classification of stream permanence in HJA, we considered all FLOWPER and thermistor observations. Any stream reach where ‘discontinuous flow’ or ‘no flow’ was observed was classified as non-permanent for 2020. Because surface water presence time series from thermistors did not record any days without surface water, only annual stream permanence was considered in the HJA. Thermistors in the WW recorded sites which had both permanent and non-permanent surface water presence throughout a year. Both daily and annual stream permanence were considered for the WW. No FLOWPER data were collected in the WW.

### *WEPP Modeling*

WEPP models were generated using the University of Idaho’s online implementation of WEPP, named WEPPCloud (<https://wepp.cloud/weppcloud/>). WEPPCloud automates much of the acquisition and formatting of the topographic, soil, land cover, and climate data required by WEPP to create the WEPP input and run files.

To setup and run a WEPP model using WEPPCloud a user specifies the input digital elevation model (DEM) resolution, both 10 m and 30 m nationally available DEM products are available. This study used a 10 m DEM to better represent topography in small, headwater catchments. After the DEM resolution is selected, and a study area is located, WEPP implements the TOPAZ (Garbrecht and Martz, 2004) software to delineate channels. Only channels with a drainage of area of at least 3 ha and greater than 70 m in length were modeled. The user then selects a basin pour-point, after which TOPAZ delineates sub-catchments and hillslopes. Land cover (from the National Land Cover Database, NLCD) and soil data (from the Soil Survey Geographic Database (SSURGO; Soil Survey

Staff) or State Soil Geographic Database (STATSGO; Schwarz and Alexander, 1995) if SSURGO is not available) are summarized for each hillslope.

Climate inputs for WEPPCloud are also summarized by hillslope and can be gathered from a variety of sources including GRIDMET (Abatzoglou, 2013), PRISM (PRISM Climate Group, Oregon State University, <http://prism.oregonstate.edu>), and Daymet (Thornton et al., 2017). GRIDMET was used for this study because it was the only climate dataset available in WEPPCloud that had coverage for 2020 (at the time of model development), the year stream permanence data were collected in the HJA for this study.

WEPPCloud allows users to alter other WEPP-specific parameters. These parameters were left at the default values for the initial WEPP run. After the initial run we downloaded the WEPPCloud project to a local machine for further parameter calibration and analysis. The watershed size of a WEPPCloud run is limited so servers are not overwhelmed with model requests for very large areas. Thus, we constrained the size and location of the four WW watersheds to be small enough to run with WEPPCloud while also modeling areas in the Willow and Whitehorse watersheds with the greatest sensor densities (Figure 2). Limiting the size of WW watersheds was also necessary to efficiently test multiple parameter combinations for WEPP. Individual WEPP models were established for each of the eight HJA watersheds and each of the four WW watersheds. Mean daily streamflow was generated from WEPP for the period spanning the 2000-2020 water years in HJA watersheds and the 2010-2017 watersheds in the WW watersheds.

### WEPP Calibration

In the HJA watersheds, each WEPP model was calibrated to observed streamflow over the 2000-2019 water years. HJA streamflow data for the 2020 water year had not been finalized at the time of this study. The purpose of this study was to evaluate the accuracy of stream permanence classifications generated by WEPP streamflow estimates. To this end, we used the entire period of record for calibration to determine the accuracy of stream permanence estimates when the best streamflow calibration was used. Based on previous studies, and the results of initial WEPP runs, we identified four WEPP parameters to alter for streamflow calibration. The deep seepage coefficient (KS), which controls how much water recharges a deeper aquifer or leaves the watershed via lateral groundwater flow, and baseflow coefficient (KB), which determines the rate of the linear baseflow recession curve, were found to be important in a previous WEPP implementations in the northwest US (Brooks et al., 2016; Srivastava et al., 2020). Additionally, based on observations from preliminary model runs that WEPP was underestimating annual water yield and flood peak magnitude we also adjusted the vertical conductivity of the restrictive layer (KR), and the crop coefficient (KC),

a multiplier of Penman-Monteith evapotranspiration, which have also been adjust by calibration in WEPP applications (Srivastava et al., 2020). Parameter units and ranges are presented in Table 1.

WEPP was run 1,000 times for each HJA watershed. For each run, the set of four parameters was randomly selected from a uniform distribution, bounded as indicated in Table 1. Percent bias (PBIAS), Nash-Sutcliffe Efficiency (NSE), and NSE for the natural log of daily streamflow estimates ( $\log Q$ ), which gives a better metric for model fit during periods of low flows, were recorded for the results of each parameter set when compared with observed streamflow data.

Evaluating the fit of each WEPP run based on PBIAS, NSE, and NSE ( $\log Q$ ), a single parameter set was identified to represent streamflow for each HJA watershed. To identify the best parameter set, all model runs with  $PBIAS < 25\%$  and  $NSE > 0.3$  were selected. From these runs the parameter set that produced the greatest value of NSE ( $\log Q$ ) was selected to represent streamflow for a watershed. PBIAS values  $< 25\%$  and NSE values  $> 0.3$  represent satisfactory results in many streamflow modeling studies (Brooks et al., 2016; Foglia et al., 2009; Moriasi et al., 2007). NSE ( $\log Q$ ) was maximized because identifying non-permanent channels is dependent upon accurate simulated streamflow during low flow periods. In the event no parameter combinations fit values where  $PBIAS < 25\%$  and  $NSE > 0.3$  the parameter set with the greatest value of NSE ( $\log Q$ ) was selected for streamflow modeling.

WEPP calibration was conducted differently in the WW because there were no streamflow data available for calibration. Instead, WEPP was calibrated to best fit the daily and annual stream permanence observations. In essence the calibration and evaluation steps were combined for the WW watersheds to evaluate the performance of WEPP. Without streamflow data for calibration only the KB and KC parameters were adjusted for WW. Without streamflow data it is difficult to identify, which parameters need to be adjusted to create a good model fit. Logically, the shape of the baseflow recession and the amount of ET are important parameters that could influence stream permanence predictions. Additionally, the larger WW watersheds required much more time for WEPP to run. By limiting calibration to two parameter the parameter space could be represented effectively with 100 model runs.

Table 4.1. WEPP parameters and their samples ranges for HJA and WW watersheds in comparison to the default WEPPCloud parameters. Parameter abbreviations are as follows: baseflow coefficient=KB, deep seepage coefficient=KS, vertical conductivity of the restrictive layer=KR, and crop coefficient=KC.

<b>Study Area</b>	<b>KB</b> days <sup>-1</sup>	<b>KS</b> days <sup>-1</sup>	<b>KR</b> mm hr <sup>-1</sup>	<b>KC</b> -
HJA	0.0 – 0.1	0.0 – 0.1	0.0 – 0.5	0.8 – 1.2
WW	0.0 – 0.1	-	-	0.8 – 1.2
Default	0.04	0.0	144.0	0.95

### WEPP Evaluation

WEPP model simulations produced estimates of average daily streamflow for each modeled stream reach. To match the time period of thermistor data, only WEPP streamflow estimates between April 1 and October 31 were used. Daily streamflow estimates were converted to wet or dry classifications for each day. Any stream reach where the modeled streamflow value was zero was classified as dry. Any stream reach where the modeled daily streamflow was greater than zero was classified as wet. Any stream reach that was modeled as dry for at least one day was classified as non-permanent for the April-October period of a given year. Streams that were not modeled dry for any days were classified as permanent for the April-October period of a given year. Classifications were made for both WEPP streamflow estimates using the best parameter set (as described above) and the default WEPPCloud parameters (Table 1).

As described above, no dry observations were recorded at thermistor locations in HJA watersheds. However, at different locations on the same reaches, FLOWPER observations recorded dry conditions. Any reach in the HJA where a dry condition was observed in 2020 was classified as non-permanent. Any reach where a dry observation was not made was classified as permanent. WEPP stream permanence classifications were evaluated based on their agreement with stream permanence classifications made from FLOWPER and thermistor observations. Agreement was determined by dividing the number of stream reaches that agree in classification by the total number of stream reaches. In total, stream permanence classifications were made for 18 stream reaches in the HJA.

Daily and annual stream permanence classifications from WEPP streamflow estimates were determined for WW stream reaches in the same manner as HJA stream reaches. Because surface water presence and absence were both recorded at many WW thermistor sites, daily accuracy of WEPP wet/dry classifications was analyzed for the WW in addition to annual accuracy. Since observed streamflow data were not available for WW watersheds, the best WEPP parameter set was

identified by comparison to thermistor data. For results from each WEPP run, the accuracy of WEPP results for dry observations, wet observations, and all observations was assessed. This assessment was done for both daily surface water presence observations and annual permanent and non-permanent classifications. In addition to the accuracy values an adjusted accuracy value, which ranged from -1.0 to 1.0, was calculated as

$$\text{Adjusted Accuracy} = \text{Overall Accuracy} - |\text{Wet Accuracy} - \text{Dry Accuracy}|$$

Where wet accuracy is the accuracy on days thermistors recorded surface water presence (or the accuracy on observed permanent streams when considered annually), dry accuracy is the accuracy on days thermistors recorded surface water absence (or the accuracy on observed non-permanent streams when considered annually), and overall accuracy is the total number of days, or years, the WEPP classification matched the observed classification. Adjusted accuracy was used to give equal weight to surface water presence and absence observations (from thermistors) because overall accuracy is biased towards the category with more observations. For example, a site may be dry for 20 days out of a 200-day period. If the model predicts all 200 days to be wet the overall accuracy would be 90% (or 0.9) but the model would not have correctly classified any dry observations. Adjusted accuracy penalizes model results for over-predicting one category in relation to the other. Stream permanence simulations from the WEPP parameter sets that produced the highest adjusted accuracy values were used to assess the performance of WEPP stream permanence estimates.

A threshold for the number of dry days (dry-day threshold) allowed before a stream reach was classified as non-permanent was also tested for WW watersheds. For example, with a dry-day threshold of zero, WEPP would need to simulate surface water to be present in a stream channel every day from April 1 to October 31 for a stream reach to be classified as permanent. With a dry-day threshold of three, a stream reach with three or fewer simulated dry days would still be classified as permanent. Other studies (e.g. Ward et al., 2018; Williamson et al., 2015) have implemented flow thresholds, where modeled flows below the threshold are classified as dry even when the model estimates water in a stream channel, which adjust models that overpredict the number of wet days. The dry-day threshold serves the opposite purpose, to adjust WEPP in the event it overpredicts too many streams to be annually non-permanent. Dry-day thresholds of 0-20 days were tested against all parameter sets and the combination of the dry-day threshold value and parameter set that produced the highest adjusted accuracy was selected for further analysis.

## Results

### *WEPP Calibration*

With calibrated parameters (Table 2) WEPP satisfactorily modeled streamflow in most HJA watersheds (Figure 3). No parameter set for HJA08 and HJA09 met both the PBIAS and NSE constraints of 25% and 0.3, respectively. PBIAS was quite large for both watersheds and indicated WEPP greatly underestimated annual water yield. This may result from the 4 km resolution of the GRIDMET climate data, which may not have accurately represented precipitation for the small watersheds. The best PBIAS values for HJA08 was -49% and for HJA09 was -42%. It was also apparent that WEPP underpredicted or missed flood peaks in several of the watersheds. However, the NSE (log Q) values indicate that WEPP satisfactorily modeled low-flow periods, which are most important for identifying when (or if) surface flow may cease. The receding limbs of flood peaks also match relatively well between observed and modeled streamflow time series. Additionally, NSE values were greater than 0.3 for all watersheds except HJA09 over the 2000-2019 water years, indicating satisfactory streamflow simulation.

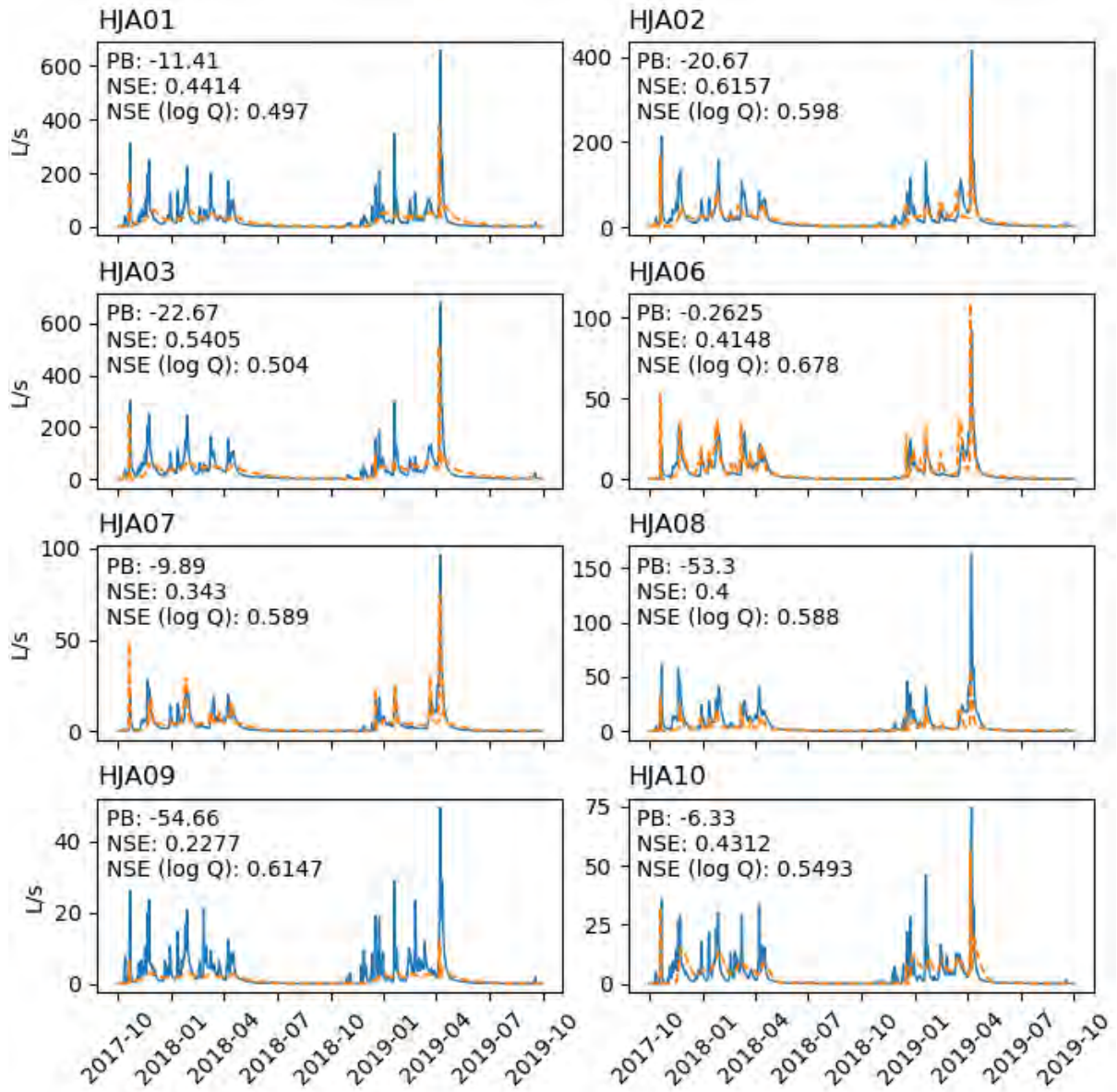


Figure 4.3. Comparison of observed (blue) and simulated (dashed orange) streamflow in HJA watersheds. Percent bias (PB), Nash Sutcliffe Efficiency (NSE), and NSE calculated on the natural log of discharge values (log Q) represent model fits over the 2000-2019 water years.

Table 4.2. WEPP parameters determined by calibration to observed streamflow for each HJA watershed. The default WEPP parameters are also shown for reference. KC=crop coefficient, KR=vertical conductivity of the restrictive layer, KS=deep seepage coefficient, and KB=baseflow recession coefficient.

WS	KC	KR	KS	KB
HJA01	0.912	0.206	0.0123	0.0072
HJA02	0.934	0.083	0.0067	0.0077
HJA03	1.029	0.180	0.0072	0.0113
HJA06	1.048	0.027	0.0090	0.0046
HJA07	1.091	0.111	0.0129	0.0037
HJA08	0.810	0.049	0.0013	0.0022
HJA09	0.946	0.074	0.0025	0.0011
HJA10	1.006	0.074	0.0089	0.0034
Default	0.950	144.0	0.0000	0.0400

#### *WEPP Stream Permanence Classification Accuracy*

##### H. J. Andrews Classification Accuracy

With the default WEPPCloud parameters, stream permanence classifications from WEPP streamflow estimates were only 39% accurate (Figure 4). In the upper reaches of the larger HJA watersheds (HJA01, HJA02, HJA03) and in three of the smaller watersheds (HJA06, HJA08, HJA10), WEPP predicted permanent conditions on non-permanent streams. Additionally, WEPP predicted the mainstem reaches of HJA01 and HJA02 (two segments) to be non-permanent when they were observed to be permanent. However, based on observations from other studies (Ward et al., 2018) the main stem reaches of HJA01 and HJA02 go dry in some places nearly every year (Sherri Johnson and Steve Wondzell, personal communication). Assuming those stream segments also had dry patches in 2020 the accuracy of the WEPPCloud default parameters would be 56%.



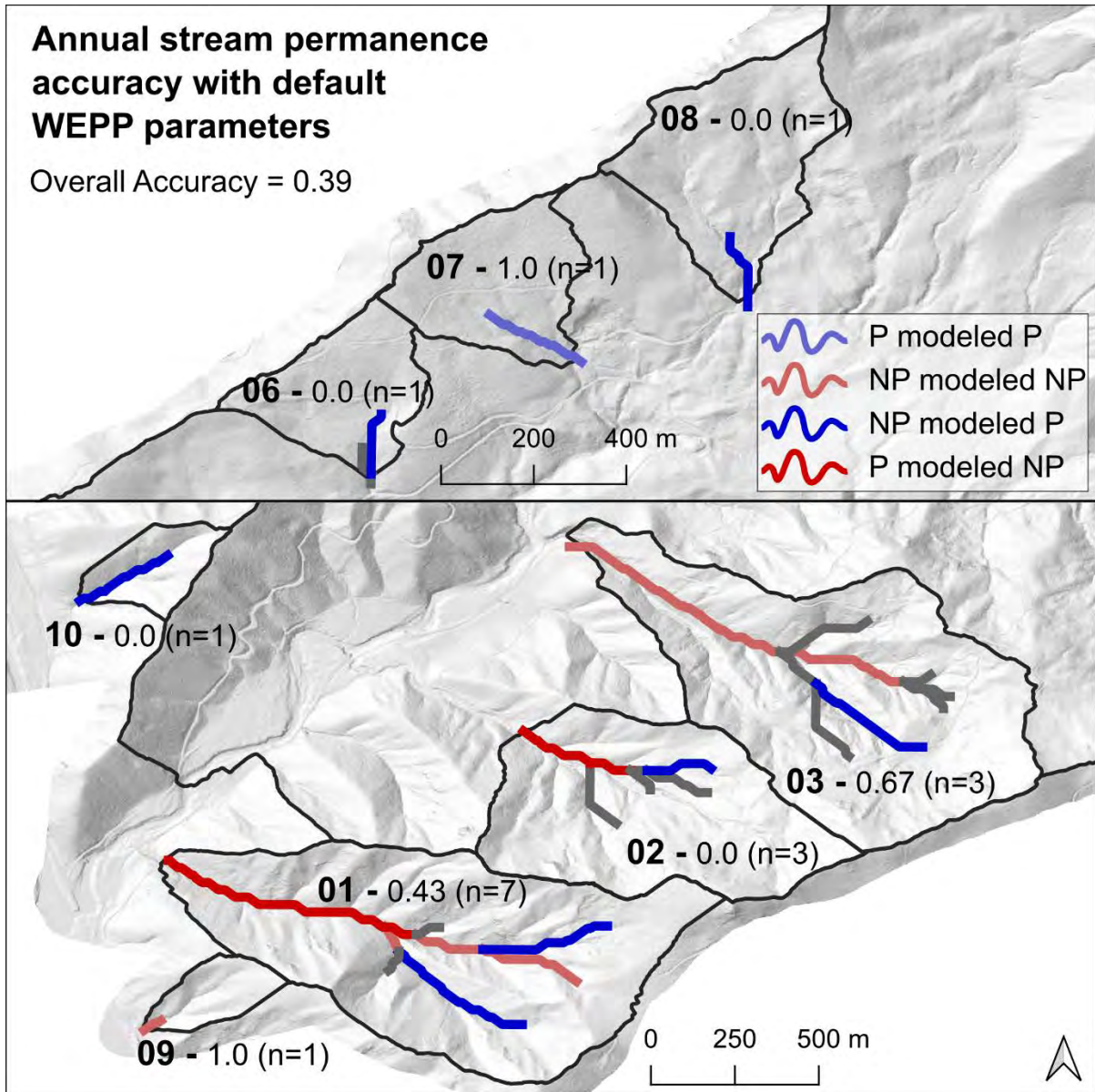


Figure 4.4. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each HJA watershed with default WEPP parameters where  $n$  is the number of modeled stream reaches in each watershed.

The WEPP model calibrated to observed streamflow performed considerably better for stream permanence classification than with the default WEPPCloud parameters, resulting in 61% accuracy with observed stream permanence classifications (Figure 5). Most errors occurred in the small HJA06, HJA07, and HJA08 watersheds. Once again, the WEPP estimates classified the main stem reaches of HJA01 and HJA02 as non-perennial when no dry observations were made on those reaches in 2020. As noted above HJA scientists observe dry portions of these stream segments nearly every year. It is important to note that the FLOWPER observations of these stream reaches did not survey the entire reach, but just the conditions in the stream channel within 10 m of a point. Two

FLOWPER observations were made on the mainstem reach of HJA01 and one observation on each of the two mainstem segments of HJA02. It is possible that the FLOWPER observations were made prior to a portion of the stream reach drying, or on a portion of the stream reach that didn't dry, while portions upstream of the observation or location point were dry. Assuming these three stream segments were non-permanent during 2020, as indicated by previous observations of HJA scientists, the accuracy of the calibrated stream permanence estimates would be 83%.

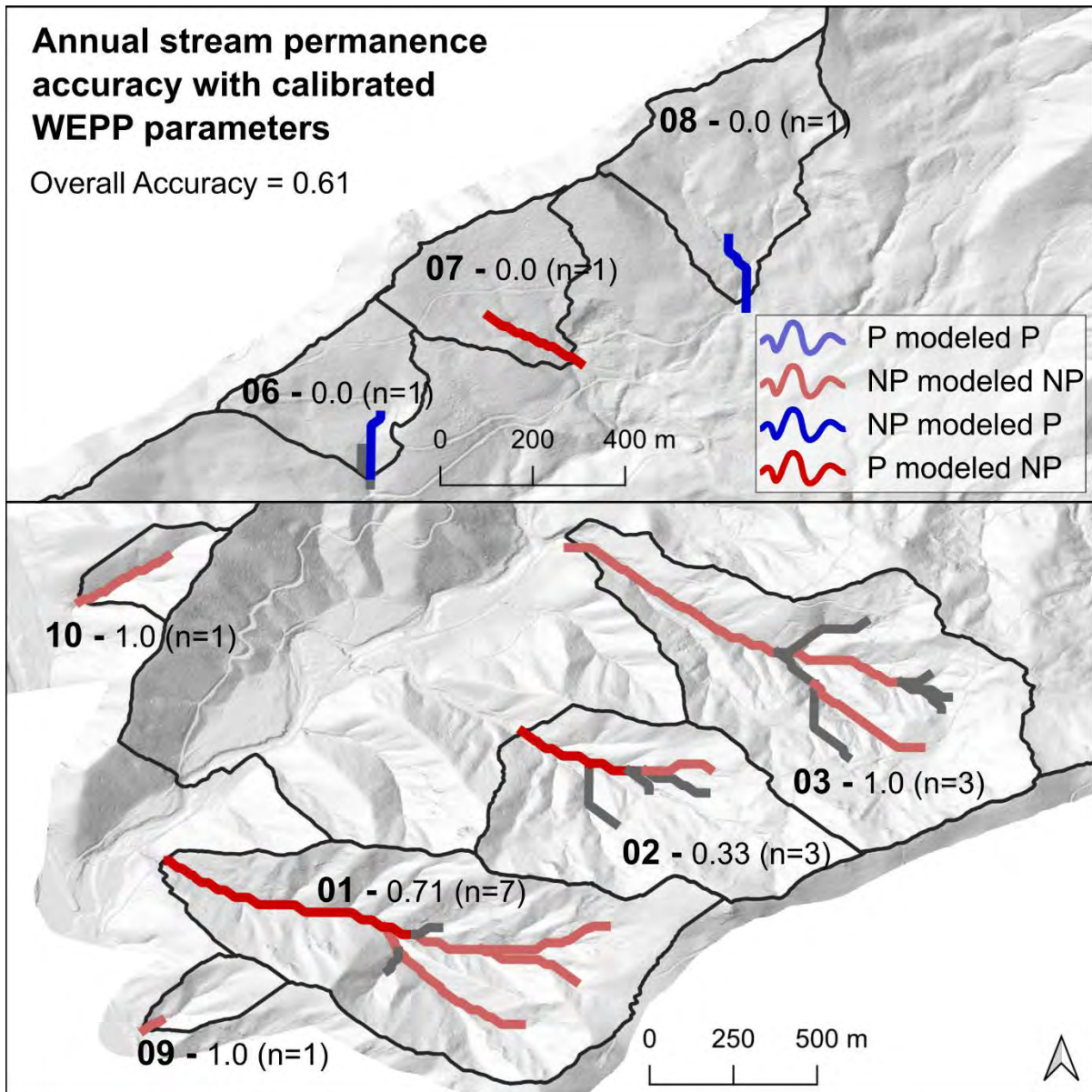


Figure 4.5. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each HJA watershed with WEPP parameters calibrated to streamflow where  $n$  is the number of stream reaches modeled in each watershed.

### Willow-Whitehorse Classification Accuracy

Maximum accuracy of daily WEPP estimates differed by watershed and ranged from 70% in WW02 to 100% in WW04 (Figure 6). However, it was apparent that maximum accuracies were inflated by a greater number of wet observations. Daily overall accuracies corresponding to the maximum adjusted accuracy were lower, ranging from 63% in WW01 to 82% in WW04. As indicated by Table 3, maximizing adjusted accuracy identified parameter sets that maximized the accuracy for both wet and dry classifications. However, accuracy of annual (permanent and nonpermanent) classifications was poor for parameter sets with the best daily adjusted accuracy with less than 50% of permanent stream reaches being classified correctly (Table 3). The discrepancies in daily and annual accuracy occur primarily on permanent streams, which were modeled as non-permanent, indicating the daily calibrations simulate too-many dry-days on permanent streams and indicate support for inclusion of a dry-day threshold.

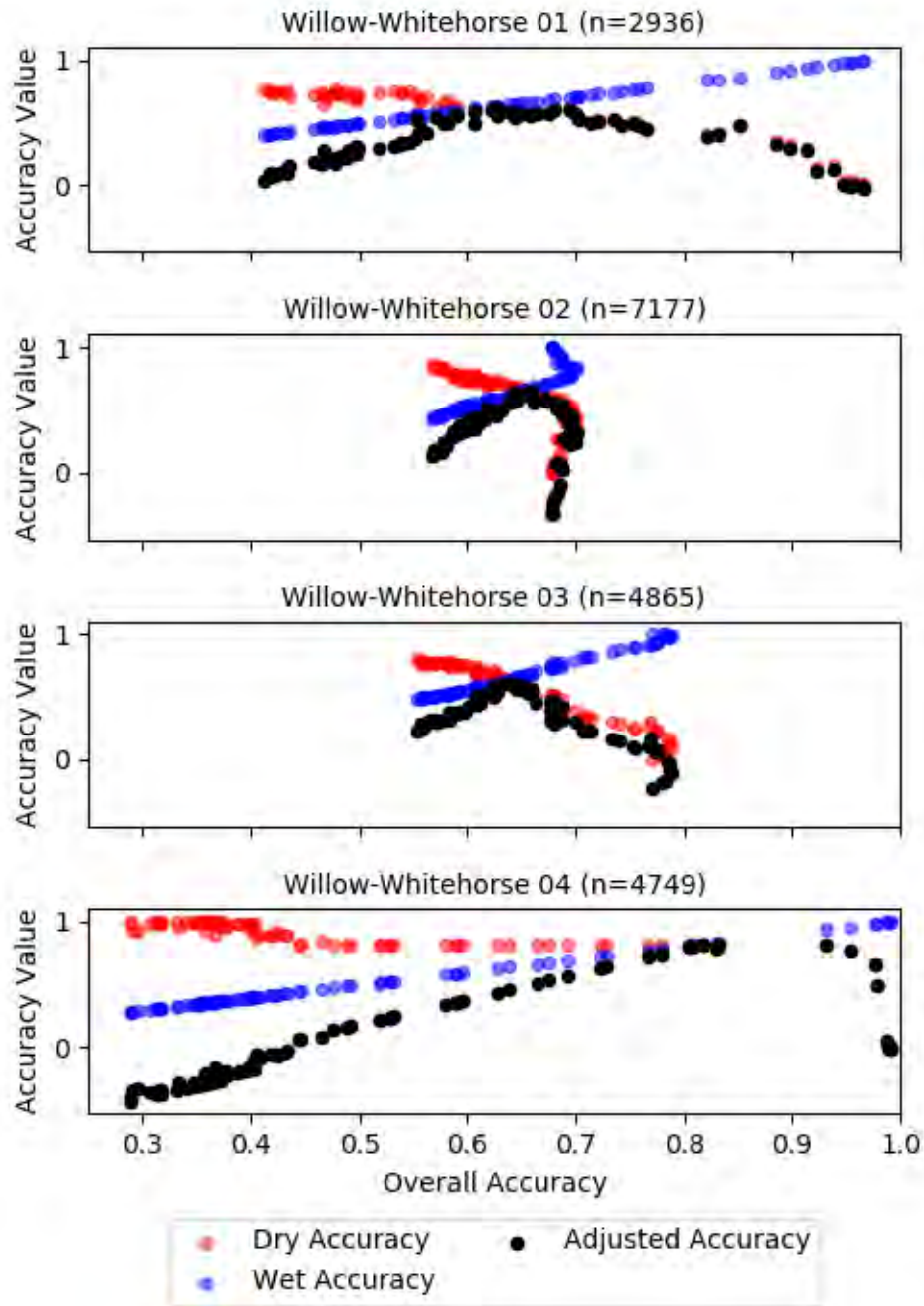


Figure 4.6. Daily accuracy of modeled (WEPP) surface water presence with different parameter sets when compared with observed surface water presence at each WW watershed where  $n$  is the sum of observed days at all thermistor sites in a watershed. Each vertical set red, blue, and black points indicates values for a different parameter set.

Table 4.3. Daily and annual accuracy of the WEPP parameter sets, calibrated to daily surface water presence observations, with the highest adjusted accuracy value for each WW watershed. WAcc and DAcc describe the modeled accuracy when compared to wet and dry surface water presence observations, respectively. PAcc and NPAcc describe annual modeled accuracy when compared to permanent and nonpermanent locations, respectively. KC is the crop coefficient parameter and KB the baseflow recession coefficient parameter.

<b>WS</b>	<b>Daily Accuracy</b>			<b>Annual Accuracy</b>			<b>WEPP params</b>	
	<b>Acc</b>	<b>WAcc</b>	<b>DAcc</b>	<b>Acc</b>	<b>PAcc</b>	<b>NPAcc</b>	<b>KC</b>	<b>KB</b>
WW01	0.63	0.63	0.61	0.20	0.00	1.00	1.139	0.030
WW02	0.66	0.65	0.66	0.67	0.07	1.00	1.076	0.052
WW03	0.64	0.65	0.62	0.47	0.00	1.00	1.153	0.064
WW04	0.82	0.82	0.82	0.60	0.57	1.00	1.105	0.014

Maximum overall accuracy of annual (non-permanent and permanent) WEPP estimates ranged from 60% in WW03 to 95% in WW04. As with daily accuracies, the maximum annual accuracies were inflated due to a greater number of observed permanent years than non-permanent years (Figure 7). Overall annual accuracies corresponding to the maximum adjusted accuracy were lower, ranging from 56% in WW03 to 85% in WW01 (Table 4). WEPP parameters corresponding to the best daily adjusted accuracy (Table 3) and best annual adjusted accuracy (Table 4) were different for all WW watersheds. This is an important finding because it indicates that a parameterization designed to maximize daily accuracy may not maximize annual accuracy and the annual permanence classification may be incorrect.



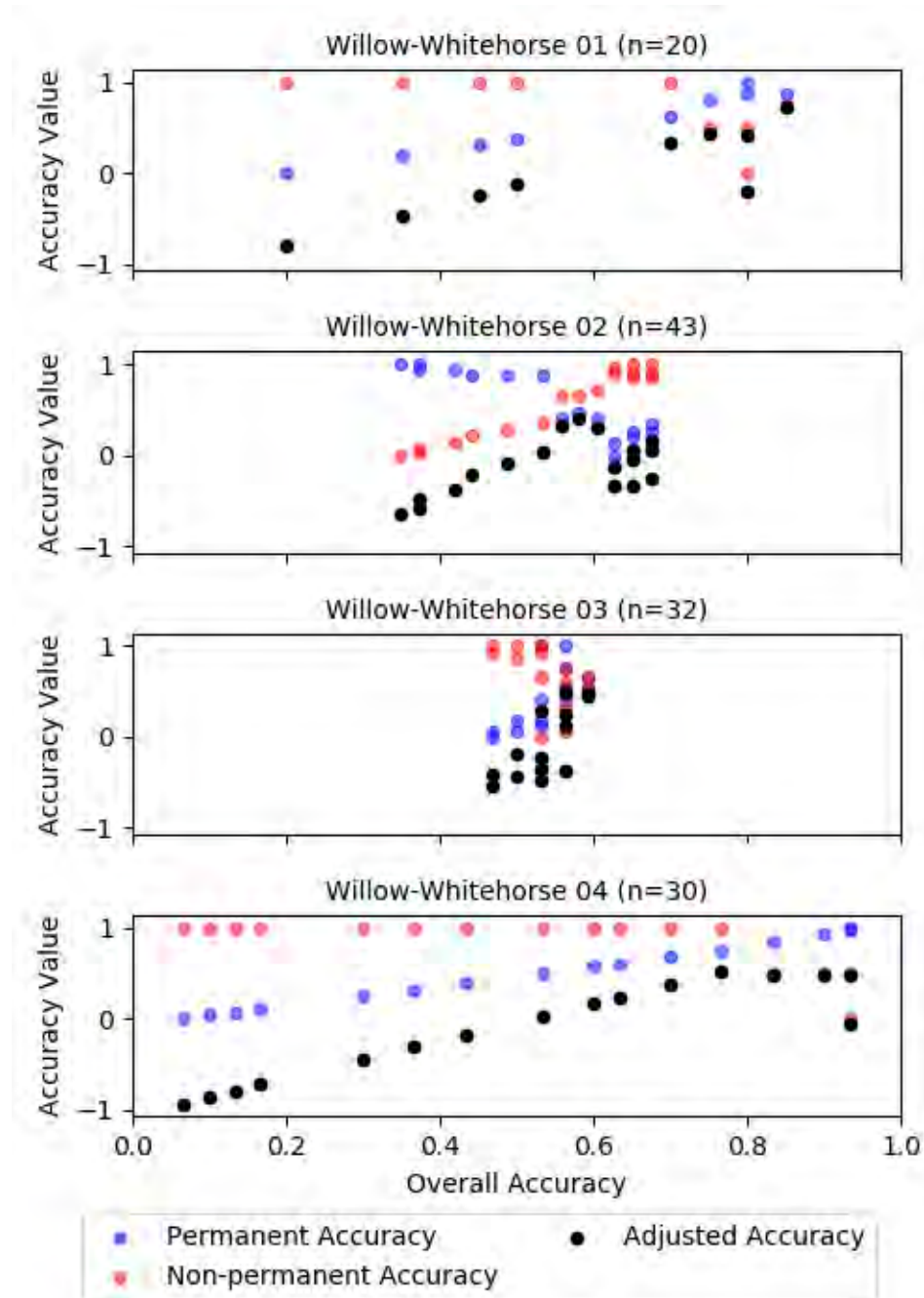


Figure 4.7. Annual analysis of WEPP accuracy for all tested parameter sets in WW watersheds where n is the number of years with observed data summed across all thermistor locations in each watershed.

Inclusion of the dry-day threshold was supported for three of the four WW watersheds. WW02 achieved its greatest annual adjusted accuracy when permanent streams were represented as streams with eight or fewer dry days, WW03 with three or fewer dry days, and WW04 with two or fewer dry days (Figure 8). WW01 achieved greatest annual adjusted accuracy with no dry-day threshold. Inclusion of the dry-day threshold improved accuracy only marginally for WW02 and

WW03 (2% and 3%, respectively). However, the dry-day threshold improved accuracy for WW04 by 10% (Table 4). This finding is similar to other studies (Ward et al., 2018; Williamson et al., 2015) that determined a daily streamflow threshold was necessary to eliminate incorrectly classified dry observations and indicates that WEPP (and potentially other models) may require an opposing threshold to adjust for permanence classifications at an annual time step.

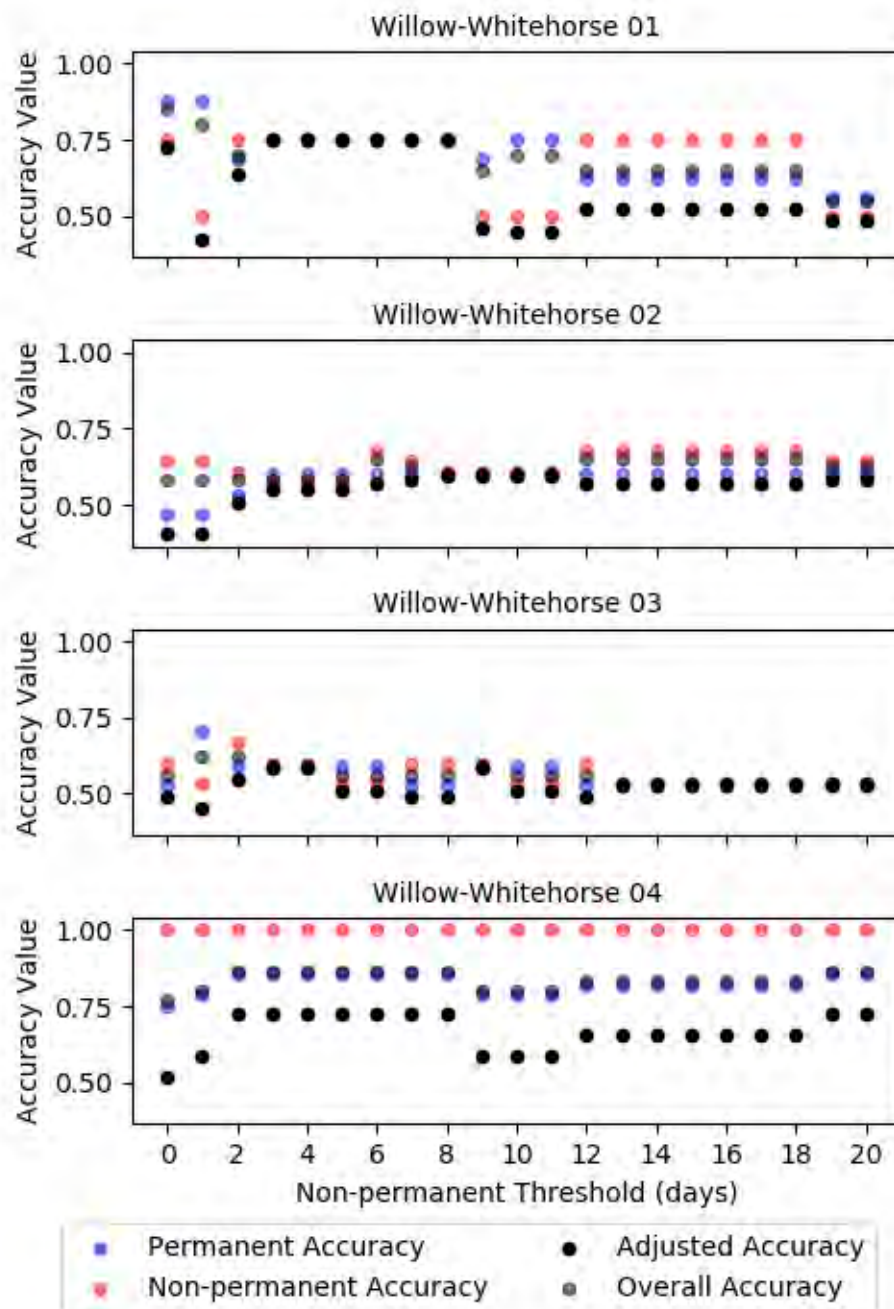


Figure 4.8. Annual accuracy values for the best WEPP parameter set for each dry-day threshold value in each WW watershed.

The WEPP parameterizations that produced the best annual accuracies for WW04 stayed consistent with addition of the dry-day threshold. However, WW02 and WW03 parameters changed with inclusion of the dry day threshold. KC and KB both increased slightly for WW02 while KC had a slight decrease and KB a slight increase for WW03 (Table 4).

Table 4.4. Annual WEPP accuracy (Acc) in each WW watershed (WS) with and without a dry-day threshold, and the corresponding values for crop coefficient (KC) and baseflow recession coefficient (KB).

<b>WS</b>	Without Dry-day Threshold			With Dry-day Threshold			
	<b>Acc</b>	<b>KC</b>	<b>KB</b>	<b>Acc</b>	<b>Days</b>	<b>KC</b>	<b>KB</b>
WW01	0.85	1.018	0.026	0.85	0	1.018	0.026
WW02	0.58	0.869	0.029	0.60	8	0.913	0.030
WW03	0.56	1.087	0.025	0.59	3	0.906	0.029
WW04	0.77	0.828	0.018	0.87	2	0.828	0.018

Analysis of daily and annual accuracies presented above indicates that high daily accuracies do not always result in correct permanent and non-permanent classifications for a year. Figure 9 through Figure 12 show the annual and daily overall accuracy (unadjusted) at each WW observed reach for each observed year as modeled with the best annual parameter set. These results show that time periods and locations with high daily accuracy may still be classified incorrectly annually (e.g. Figure 9) while time periods and locations with low accuracy may still be classified correctly (e.g. Figure 10). It is important to note that the level of daily accuracy for a correct non-permanent classification varies. For example, a non-permanent stream reach could be dry for 10 or 100 days during a year. The model only needs to simulate enough dry days to be greater than the dry-day threshold (if one is used) to correctly simulate the annual condition. On the other hand, permanent stream reaches require the model to correctly simulate all wet days (if no dry-day threshold is used) to produce the correct annual classification.



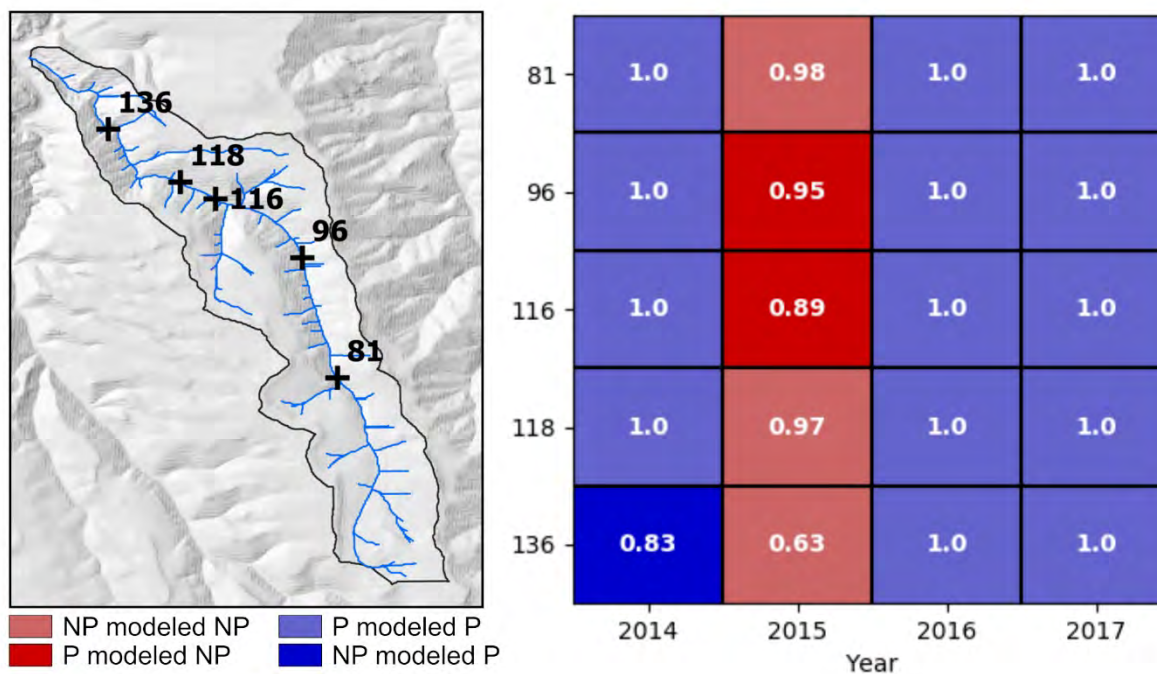


Figure 4.9. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW01. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

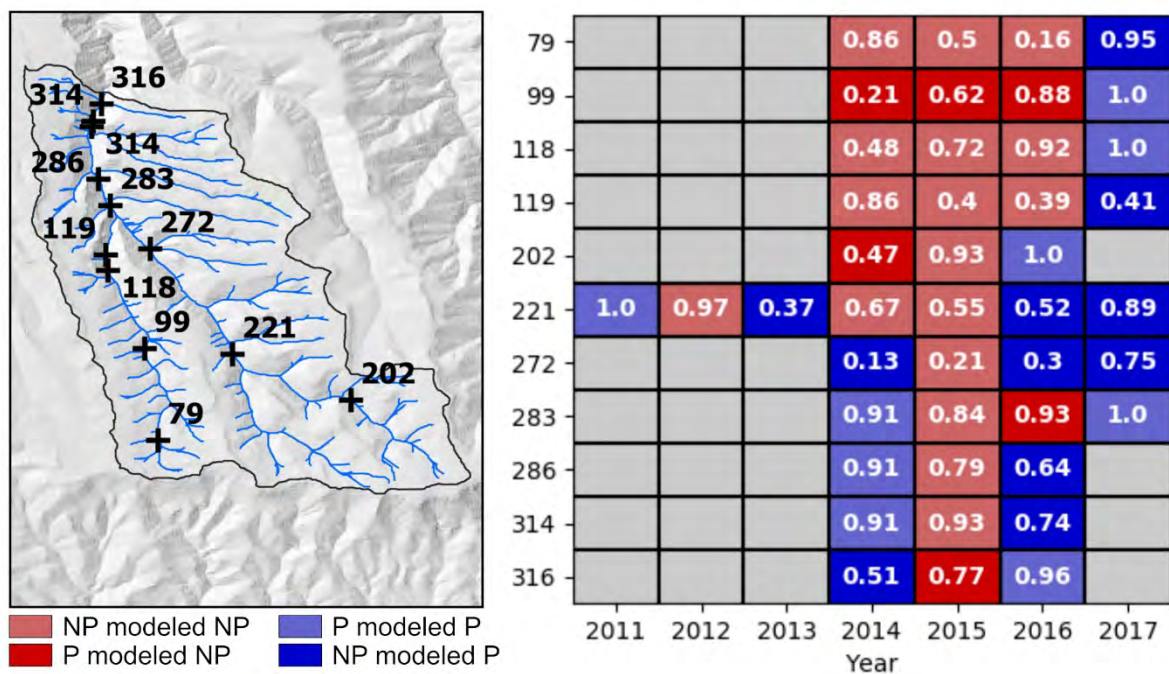


Figure 4.10. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW02. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

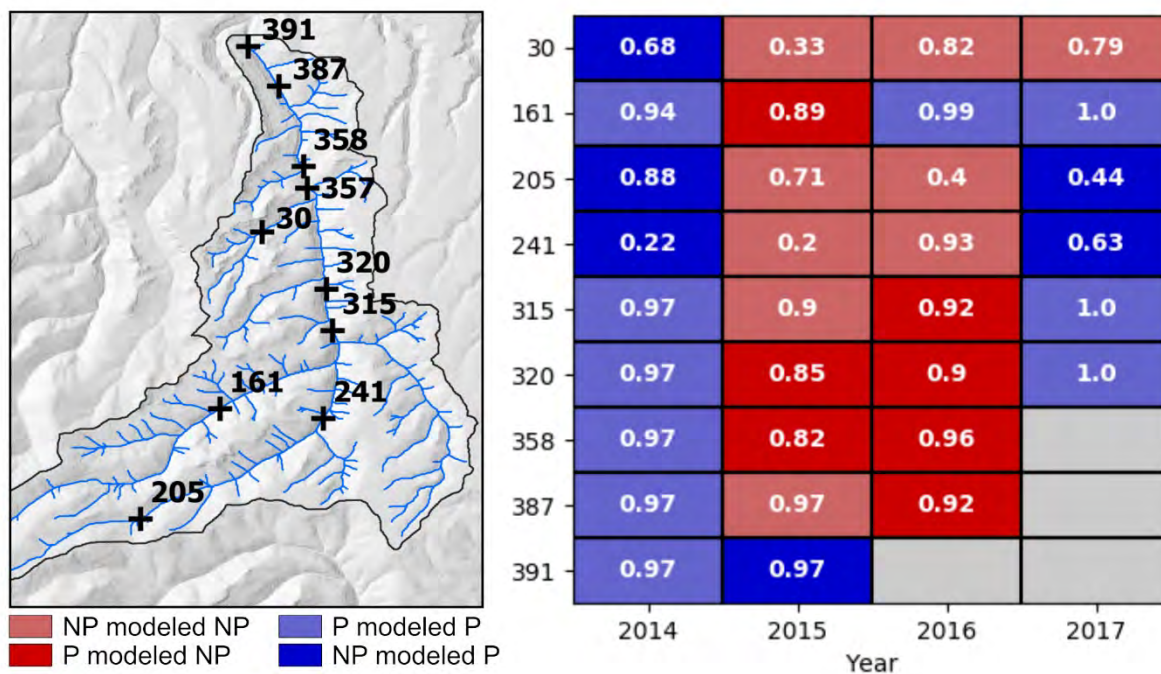


Figure 4.11. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW03. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

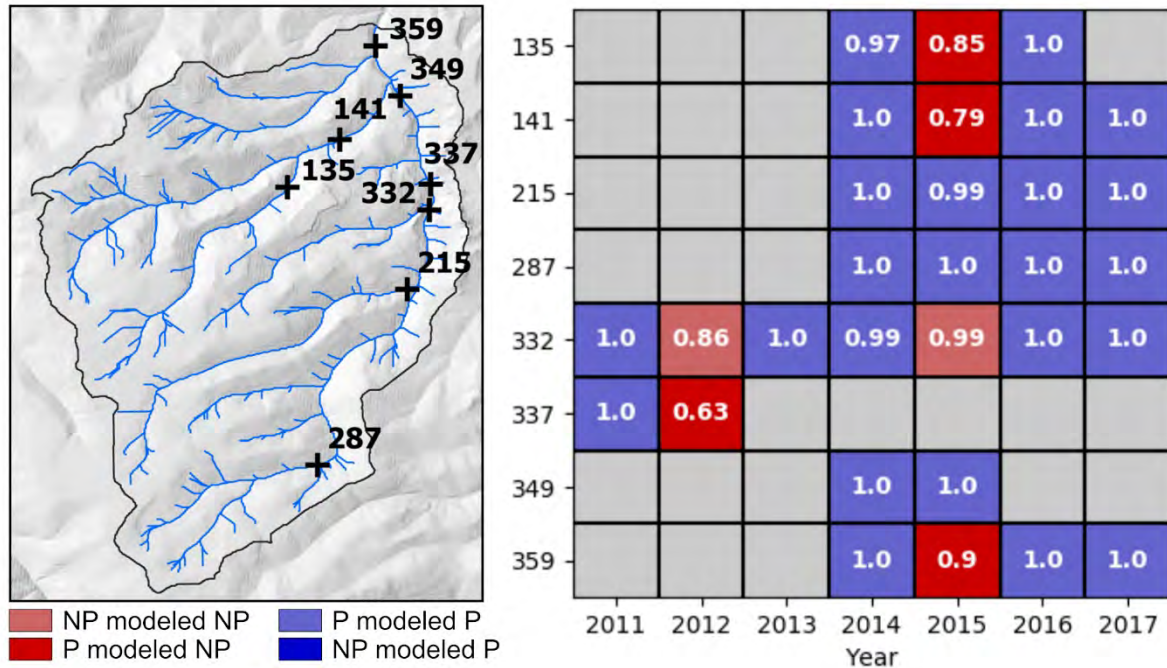


Figure 4.12. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW04. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

### Discussion

This work presents a case study of stream permanence modeling using the WEPP model in both a humid and an arid study area. Assessment of stream permanence classifications was performed at daily and annual time steps using continuously recorded values from thermistors and observations from human observers. Both data types have been used in previous modeling studies. Results indicate the importance of evaluating model simulations based on both daily and annual accuracy, assessing model performance on both permanent and non-permanent streams, and the need for targeted data collection to accurately describe the permanence condition of stream reaches.

Previous studies that examined the utility of physically-based models to simulate stream permanence have focused intensive data collection efforts on a small number of nonpermanent stream reaches (Ward et al., 2018; Williamson et al., 2015). Data for this study represented 40 unique stream reaches (18 in HJA and 32 in WW) but did not describe reaches in the same detail as other studies have. One advantage of this approach is that both permanent and nonpermanent streams are represented. Without data for both permanent and nonpermanent streams model utility is uncertain because a single daily miscalculation on a permanent stream reach can result in an annual classification error. Assessment of annual permanence classification accuracy is also important to

temper daily accuracy assessment which can be biased from unequal occurrence of wet and dry days. Results also show that correct annual classifications can be modeled even with lower daily accuracies.

In the HJA watersheds WEPP performed much better after streamflow estimates were calibrated to observed streamflow than with the default WEPPCloud parameterization. However, very few areas have the gage density of HJA, potentially making this a limited approach. Calibrated parameters were relatively similar between the gauged HJA watersheds but did display some variation. This indicates that a single parameterization based on streamflow on a larger stream reach may also produce suitable results. Future research could assess WEPP stream permanence estimates when WEPP is calibrated to streamflow on a larger stream, typical of most gage sites.

Misclassification of two non-permanent streams, one in HJA01 and one in HJA02, point to the uncertainty associated with using data from point observations to make reach-scale classifications. The HJA misclassifications may also help explain high uncertainty of another model (the PROSPER regional model) in the western portions of the Pacific Northwest Region (Jaeger et al., 2019). Despite relatively high-density data in this area the PROSPER model predictions were largely inaccurate. A similar phenomenon in the data driving PROSPER, where wet observations were made on non-permanent stream reaches without also recording dry observations, could describe these inconsistencies. These examples indicate that data collection for classifying stream permanence will be most effective when focused in areas where stream reaches are most likely to go dry. Features of these locations include wide valley-bottoms with deep alluvial soils where subsurface water flow capacity is likely high (Ward et al., 2018) or deep coarse bed sediments in sub alpine stream channels (Sando and Blasch, 2015). Dry observations indicate, with certainty, a stream reach was not permanent (for a given time period) while wet observations only serve to support the hypothesis that the stream is permanent but cannot confirm this hypothesis unless spatially and temporally continuous observations are made.

This study implemented a unique approach by using surface water presence observations to calibrate WEPP in the WW watersheds where streamflow data were not available. Overall, the calibration produced good results in WW01 and WW04 and moderate results in WW02 and WW03 for both daily and annual accuracy. Calibration to surface water presence observations could be useful for future development of physically-based streamflow models because collection of surface water presence/absence data are easier and less costly to acquire than continuous records of streamflow data. The drawback to this method, as previously mentioned, is that surface water sensors record the condition at a point and may not necessarily be indicative of a stream reach, the spatial scale that is important for regulatory determinations. Focusing data collection at locations that are more

representative of a stream reach (as mentioned above), along with reach scale classification from on-the-ground surveys and remotely sensed products will help narrow this knowledge gap. The application of WEPP with stream permanence observations would also be advantageous to describe a watershed to select the optimal installation of more permanent but costly continuous hydrological or ecological monitoring equipment.



### Literature Cited

- Abatzoglou, J.T., 2013. Development of Gridded Surface Meteorological Data for Ecological Applications and Modelling. *International Journal of Climatology* 33:121–131.
- Alexander, R.B., E.W. Boyer, R.A. Smith, G.E. Schwarz, and R.B. Moore, 2007. The Role of Headwater Streams in Downstream Water Quality. *JAWRA Journal of the American Water Resources Association* 43:41–59.
- Arismendi, I., J.B. Dunham, M.P. Heck, L.D. Schultz, and D. Hockman-Wert, 2017. A Statistical Method to Predict Flow Permanence in Dryland Streams from Time Series of Stream Temperature. *Water* 9:946.
- Blasch, K.W., T.P. Ferré, A.H. Christensen, and J.P. Hoffmann, 2002. New Field Method to Determine Streamflow Timing Using Electrical Resistance Sensors. *Vadose Zone Journal* 1:289–299.
- Brooks, E.S., M. Dobre, W.J. Elliot, J.Q. Wu, and J. Boll, 2016. Watershed-Scale Evaluation of the Water Erosion Prediction Project (WEPP) Model in the Lake Tahoe Basin. *Journal of Hydrology* 533:389–402.
- DeWeber, J.T., Y.-P. Tsang, D.M. Krueger, J.B. Whittier, T. Wagner, D.M. Infante, and G. Whelan, 2014. Importance of Understanding Landscape Biases in USGS Gage Locations: Implications and Solutions for Managers. *Fisheries* 39:155–163.
- Downing, J.A., J.J. Cole, C. Duarte, J.J. Middelburg, J.M. Melack, Y.T. Prairie, P. Kortelainen, R.G. Striegl, W.H. McDowell, and L.J. Tranvik, 2012. Global Abundance and Size Distribution of Streams and Rivers. *Inland Waters* 2:229–236.
- Dun, S., J.Q. Wu, W.J. Elliot, P.R. Robichaud, D.C. Flanagan, J.R. Frankenberger, R.E. Brown, and A.C. Xu, 2009. Adapting the Water Erosion Prediction Project (WEPP) Model for Forest Applications. *Journal of Hydrology* 366:46–54.
- Dunham, J., R. Schroeter, and B. Rieman, 2003. Influence of Maximum Water Temperature on Occurrence of Lahontan Cutthroat Trout within Streams. *North American Journal of Fisheries Management* 23:1042–1049.
- Dyrness, C., 1969. Hydrologic Properties of Soils on Three Small Watersheds in the Western Cascades of Oregon. *USDA For. Ser. Res. Note PNW* 111:17.
- Flanagan, D.C., J.C. Ascough, M.A. Nearing, and J.M. Laflen, 2001. The Water Erosion Prediction Project (WEPP) Model. *Landscape Erosion and Evolution Modeling*. Springer, pp. 145–199.
- Flanagan, D.C. and M.A. Nearing, 1995. *USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*. NSERL report.
- Foglia, L., M.C. Hill, S.W. Mehl, and P. Burlando, 2009. Sensitivity Analysis, Calibration, and Testing of a Distributed Hydrological Model Using Error-Based Weighting and One Objective Function. *Water Resources Research* 45. doi:10.1029/2008WR007255.
- Fraday, C., S. Johnson, and J. Li, 2007. Stream Macroinvertebrate Community Responses as Legacies of Forest Harvest at the HJ Andrews Experimental Forest, Oregon. *Forest Science* 53:281–293.
- Fritz, K.M., E. Hagenbuch, E. D’Amico, M. Reif, P.J. Wigington, S.G. Leibowitz, R.L. Comeleo, J.L. Ebersole, and T.-L. Nadeau, 2013. Comparing the Extent and Permanence of Headwater

- Streams From Two Field Surveys to Values From Hydrographic Databases and Maps. *JAWRA Journal of the American Water Resources Association* 49:867–882.
- Garbrecht, J. and L.W. Martz, 2004. TOPAZ User Manual.
- Gendaszek, A.S., J.B. Dunham, C.E. Torgersen, D.P. Hockman-Wert, M.P. Heck, J. Thorson, J. Mintz, and T. Allai, 2020. Land-Cover and Climatic Controls on Water Temperature, Flow Permanence, and Fragmentation of Great Basin Stream Networks. *Water* 12:1962.
- Hafen, K.C., K.W. Blasch, A. Rea, R. Sando, and P.E. Gessler, 2020. The Influence of Climate Variability on the Accuracy of NHD Perennial and Nonperennial Stream Classifications. *JAWRA Journal of the American Water Resources Association* 56:903–916.
- Hammond, J.C., M. Zimmer, M. Shanafield, K. Kaiser, S.E. Godsey, M.C. Mims, S.C. Zipper, R.M. Burrows, S.K. Kampf, and W. Dodds, 2020. Spatial Patterns and Drivers of Non-perennial Flow Regimes in the Contiguous US. *Geophysical Research Letters*:2020GL090794.
- Hester, G., K. Carsell, and D. Ford, 2006. Benefits of USGS Streamgaging Program—Users and Uses of USGS Streamflow Data. *National Hydrologic Warning Council*:1–18.
- Jaeger, K.L., R. Sando, R.R. McShane, J.B. Dunham, D.P. Hockman-Wert, K.E. Kaiser, K. Hafen, J.C. Risley, and K.W. Blasch, 2019. Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest. *Journal of Hydrology X* 2:100005.
- Jensen, C.K., K.J. McGuire, and P.S. Prince, 2017. Headwater Stream Length Dynamics across Four Physiographic Provinces of the Appalachian Highlands. *Hydrological Processes* 31:3350–3363.
- Jensen, C.K., K.J. McGuire, Y. Shao, and C. Andrew Dolloff, 2018. Modeling Wet Headwater Stream Networks across Multiple Flow Conditions in the Appalachian Highlands. *Earth Surface Processes and Landforms* 43:2762–2778.
- Jones, J. and G.E. Grant, 1996. Peak Flow Responses to Clear-cutting and Roads in Small and Large Basins, Western Cascades, Oregon. *Water Resources Research* 32:959–974.
- Jones, J.A., F.J. Swanson, B.C. Wemple, and K.U. Snyder, 2000. Effects of Roads on Hydrology, Geomorphology, and Disturbance Patches in Stream Networks. *Conservation Biology* 14:76–85.
- Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE* 50:885–900.
- Nadeau, T.-L., G.S. Leibowitz, J.P. Wigington, L.J. Ebersole, M.K. Fritz, A.R. Coulombe, L.R. Comeleo, and A.K. Blocksom, 2015. Validation of Rapid Assessment Methods to Determine Streamflow Duration Classes in the Pacific Northwest, USA. *Environmental Management* 56:34–53.
- Nadeau, T.-L. and C.M. Rains, 2007. Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy. *Journal of the American Water Resources Association* 43. <http://doi.wiley.com/10.1111/j.1752-1688.2007.00010.x>.
- Sando, R. and W.K. Blasch, 2015. Predicting Alpine Headwater Stream Intermittency: A Case Study in the Northern Rocky Mountains. *Ecohydrology & Hydrobiology* 15. <https://linkinghub.elsevier.com/retrieve/pii/S1642359315000154>.



- Schultz, L., M. Heck, D. Hockman-Wert, T. Allai, S. Wenger, N. Cook, and J.B. Dunham, 2017. Spatial and Temporal Variability in the Effects of Wildfire and Drought on Thermal Habitat for a Desert Trout. *Journal of Arid Environments* 145:60–68.
- Schwarz, G.E. and R. Alexander, 1995. State Soil Geographic (STATSGO) Data Base for the Conterminous United States. USGS Numbered Series. <https://pubs.er.usgs.gov/publication/ofr95449>.
- Soil Survey Staff, Soil Survey Geographic (SSURGO) Database. Natural Resources Conservation Service, United States Department of Agriculture. <https://sdmdataaccess.sc.egov.usda.gov/>. Accessed 1 Mar 2016.
- Srivastava, A., E. Brooks, M. Dobre, W. Elliot, J. Wu, D. Flanagan, J. Gravelle, and T. Link, 2020. Modeling Forest Management Effects on Water and Sediment Yield from Nested, Paired Watersheds in the Interior Pacific Northwest, USA Using WEPP. *Science of The Total Environment* 701:134877.
- Srivastava, A., M. Dobre, J.Q. Wu, W.J. Elliot, E.A. Bruner, S. Dun, E.S. Brooks, and I.S. Miller, 2013. Modifying WEPP to Improve Streamflow Simulation in a Pacific Northwest Watershed. *Transactions of the ASABE* 56:603–611.
- Swanson, F.J. and C. Dyrness, 1975. Impact of Clear-Cutting and Road Construction on Soil Erosion by Landslides in the Western Cascade Range, Oregon. *Geology* 3:393–396.
- Swanson, F.J. and M.E. James, 1975. Geology and Geomorphology of the HJ Andrews Experimental Forest, Western Cascades, Oregon.
- Thomas, R.B. and W.F. Megahan, 1998. Peak Flow Responses to Clear-cutting and Roads in Small and Large Basins, Western Cascades, Oregon: A Second Opinion. *Water Resources Research* 34:3393–3403.
- Thornton, P.E., M.M. Thornton, B.W. Mayer, Y. Wei, R. Devarakonda, R.S. Vose, and R.B. Cook, 2017. Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 3. ORNL DAAC. doi:<https://doi.org/10.3334/ORNLDAAAC/1328>.
- Ward, A.S., N.M. Schmadel, and S.M. Wondzell, 2018. Simulation of Dynamic Expansion, Contraction, and Connectivity in a Mountain Stream Network. *Advances in Water Resources* 114:64–82.
- Ward, A.S., S.M. Wondzell, N.M. Schmadel, and S.P. Herzog, 2020. Climate Change Causes River Network Contraction and Disconnection in the HJ Andrews Experimental Forest, Oregon, USA. *Front. Water* 2: 7. Doi: 10.3389/frwa.
- Williamson, T.N., C.T. Agouridis, C.D. Barton, J.A. Villines, and J.G. Lant, 2015. Classification of Ephemeral, Intermittent, and Perennial Stream Reaches Using a TOPMODEL-Based Approach. *JAWRA Journal of the American Water Resources Association* 51:1739–1759.
- Zheng, F., X.J. Zhang, J. Wang, and D.C. Flanagan, 2020. Assessing Applicability of the WEPP Hillslope Model to Steep Landscapes in the Northern Loess Plateau of China. *Soil and Tillage Research* 197:104492.

## Chapter 5: Conclusion

Surface water presence in streams is a spatially and temporally dynamic occurrence related to climate conditions (Godsey and Kirchner, 2014) that is not well-represented by static classifications (Hafen et al., 2020). Thus, incorporating climate variables into stream permanence classifications will be essential to generating dynamic stream permanence products. This work further establishes the importance of climatic variables for stream permanence classification by identifying relationships between NHD disagreements and climate conditions. Additionally, results from the modeling studies, herein, highlight the need for continued data collection and further development of data collection methods to quantify stream permanence at broader spatial extents with increased temporal frequency. Modeling efforts also speak to important practices for evaluating model errors and flow processes that may be important for stream permanence modeling.

Stream permanence modeling on headwater streams using the USGS Thornthwaite Monthly Water Balance Model, which has been extensively implemented to simulate hydrology over large spatial extents with satisfactory results (Bock et al., 2016, 2018; McCabe and Wolock, 2008, 2011), produced mixed results. Implementation of MWBM runoff estimates to drive additional streamflow models linked to the National Hydrography Dataset (NHD) stream network (McKay et al., 2012; Miller et al., 2018; Moore et al., 2019) indicated that it may also be useful to simulate stream permanence at regional (and potentially larger) extents. However, application of the MWBM to simulate stream permanence on headwater streams in the Pacific Northwest (PNW) failed to produce satisfactory results on 40% of headwater streams. In total, 28-45% of headwater streams exhibited satisfactory model precision depending on the year. These results highlight three important considerations for future regional modeling.

First, current surface water presence datasets may not be extensive enough to support regional modeling. The number of observations is far less than the number of modeled stream reaches. Even though surface water presence observations (SWPO) are often more evenly distributed across stream networks than streamflow gage stations, each SWPO represents only a single point in time and space and repeat SWPO at the same location are uncommon. Therefore, only low confidence can be given to a SWPO representing a stream channel as ‘wet’ because the conditions upstream, downstream, and in the future are unknown. This study utilized the most extensive data set of SWPO available in the US (at the time) with 2,804 observations on headwater streams. This dataset was not large enough to adequately perform error analysis (thus, model precision was considered instead) for the 1.3 million headwater streams in the PNW study area. Future data collection to support regional

models should focus on characterizing surface water presence at the stream-reach scale with multiple observations each year.

Second, the MWBM's ability to represent processes governing baseflow conditions in small catchments is largely unknown. Timing of runoff estimates from the MWBM are governed by a parameter controlling direct runoff and another that specifies the proportion that of stored water that is converted to lateral flow, baseflow, and saturation excess flow, which are lumped together in a single runoff value. As indicated by WEPP parameter values (Chapter 3) baseflow characterization can be important for correctly identifying no-flow periods. Altering the MWBM to explicitly represent additional flow processes, specifically baseflow, may improve stream permanence simulation accuracy.

Third, the MWBM's monthly time step may not provide adequate temporal resolution to identify nonpermanent streams. The number of days a nonpermanent stream does not contain surface water can vary from just a few days to the majority of a year (Beaufort et al., 2019; Jaeger et al., 2019; Sando and Blasch, 2015; Ward et al., 2018; Williamson et al., 2015). It is possible the monthly MWBM simulations may not capture streamflow variation within a month well enough to identify months with a small number of dry days. However, this remains to be tested and will require additional data collection that includes repeat observations to identify the limits of the MWBM for stream permanence simulation.

Implementing WEPP to simulate stream permanence for stream reaches presented results that add to the current body of knowledge regarding stream permanence modeling with physically-based models. The best parameter sets for daily simulations resulted in poor annual classifications of permanent and nonpermanent streams indicating that accuracy metrics from time-series data should be tempered by the annual outcome. Such inconsistencies arise from greater numbers of SWPO representing wet periods than dry periods. Using a metric like adjusted accuracy can help identify parameter sets or model runs to balance accuracy between the different observation categories but, as demonstrated in Chapter 3, can still be misleading of the accuracy for annual permanence classifications.

Misclassification of permanence status based on observations in the H. J. Andrews watersheds highlights the need for continued development of data collection procedures and methods to quantify stream permanence at reach scales. Future research could focus on identifying the portions of stream reaches where surface water is most likely to disappear during dry months and focus

sampling at those locations. This would increase confidence that observations of surface water presence are indicators of stream permanence at reach scales. Additionally, new analytical methods could be developed to evaluate SWPO as probabilities instead of categories to account for uncertainty of how surface water presence may change at locations spatially and temporally. Probability of detection methods have been applied to evaluate stream permanence classifications (Beaufort *et al.*, 2019) and applying these methods to the input data could, at a minimum, improve uncertainty estimates for model outputs.

Stream permanence modeling is currently in the exploratory phases as the utility of the comprehensive NHD stream permanence classifications have been questioned for regulatory purposes and new methods for identifying stream permanence are developed and evaluated. Models have been, and are being developed, but these efforts are somewhat limited by data availability and limited understanding of the data burden required to calibrate and validate models that can simulate stream permanence with the desired accuracy over desired extents. While gauged streamflow data that have been used for decades to support hydrological studies are useful for modeling stream permanence, data that are fundamentally different than streamflow time series at points are required to effectively assess stream permanence estimates through time and space. Current stream permanence models (including efforts herein) do provide important information to guide future data collection and model development. Results from these stream permanence studies may be useful for identifying where more data are needed and the spatial and temporal scales at which data collection is necessary to provide a calibration and validation base for improved models. Additionally, model development and application aids in determining processes and techniques that are required for effective stream permanence modeling. To date, hydrological sciences have invested substantial time and resources to quantifying when, where, and how much water is available for different purposes. Now it is time to devote more resources to focus on when and where water is present.

### Literature Cited

- Beaufort, A., J. Carreau, and E. Sauquet, 2019. A Classification Approach to Reconstruct Local Daily Drying Dynamics at Headwater Streams. *Hydrological Processes* 33:1896–1912.
- Bock, A.R., W.H. Farmer, and L.E. Hay, 2018. Quantifying Uncertainty in Simulated Streamflow and Runoff from a Continental-Scale Monthly Water Balance Model. *Advances in Water Resources* 122:166–175.
- Bock, A.R., L.E. Hay, G.J. McCabe, S.L. Markstrom, and R.D. Atkinson, 2016. Parameter Regionalization of a Monthly Water Balance Model for the Conterminous United States. *Hydrology and Earth System Sciences* 20:2861–2876.
- Godsey, S.E. and J.W. Kirchner, 2014. Dynamic, Discontinuous Stream Networks: Hydrologically Driven Variations in Active Drainage Density, Flowing Channels and Stream Order. *Hydrological Processes* 28:5791–5803.
- Hafen, K.C., K.W. Blasch, A. Rea, R. Sando, and P.E. Gessler, 2020. The Influence of Climate Variability on the Accuracy of NHD Perennial and Nonperennial Stream Classifications. *JAWRA Journal of the American Water Resources Association* 56:903–916.
- Jaeger, K.L., R. Sando, R.R. McShane, J.B. Dunham, D.P. Hockman-Wert, K.E. Kaiser, K. Hafen, J.C. Risley, and K.W. Blasch, 2019. Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest. *Journal of Hydrology X* 2:100005.
- McCabe, G.J. and D.M. Wolock, 2008. Joint Variability of Global Runoff and Global Sea Surface Temperatures. *Journal of Hydrometeorology* 9:816–824.
- McCabe, G.J. and D.M. Wolock, 2011. Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States. *Water Resources Research* 47. doi:10.1029/2011WR010630.
- McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012. NHDPlus Version 2: User Guide. US Environmental Protection Agency.
- Miller, M.P., D.M. Carlisle, D.M. Wolock, and M. Wieczorek, 2018. A Database of Natural Monthly Streamflow Estimates from 1950 to 2015 for the Conterminous United States. *JAWRA Journal of the American Water Resources Association* 54:1258–1269.
- Moore, R.B., L.D. McKay, A.H. Rea, T.R. Bondelid, C.V. Price, T.G. Dewald, and C.M. Johnston, 2019. User's Guide for the National Hydrography Dataset plus (NHDPlus) High Resolution. USGS Numbered Series, U.S. Geological Survey, Reston, VA.
- Sando, R. and W.K. Blasch, 2015. Predicting Alpine Headwater Stream Intermittency: A Case Study in the Northern Rocky Mountains. *Ecohydrology & Hydrobiology* 15. <https://linkinghub.elsevier.com/retrieve/pii/S1642359315000154>.
- Ward, A.S., N.M. Schmadel, and S.M. Wondzell, 2018. Simulation of Dynamic Expansion, Contraction, and Connectivity in a Mountain Stream Network. *Advances in Water Resources* 114:64–82.
- Williamson, T.N., C.T. Agouridis, C.D. Barton, J.A. Villines, and J.G. Lant, 2015. Classification of Ephemeral, Intermittent, and Perennial Stream Reaches Using a TOPMODEL-Based Approach. *JAWRA Journal of the American Water Resources Association* 51:1739–1759.