Development and Evaluation of Streamflow Duration Assessment Methods for Regions of the Coterminous United States

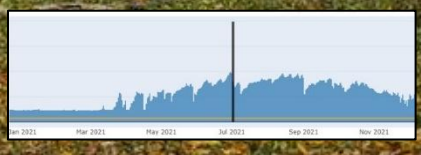
2026











EPA-843-R-25001

United States
Environmental Protection
Agency

Development and Evaluation of Streamflow Duration Assessment Methods for Regions of the Coterminous United States

Raphael D. Mazor
Southern California Coastal Water Research Project
Costa Mesa, CA 92626

Shannon Gross
RTI International
Fort Collins, CO 80528

Ken M. Fritz
USEPA Office of Wetlands, Oceans, and Watersheds
Cincinnati, OH 45268

Amy James
Ecosystem Planning and Restoration
Cary, NC 27511

Brian Topping
USEPA Office of Wetlands, Oceans, and Watersheds
Washington, D.C. 20004

Rachel Fertik Edgerton (retired)
USEPA Office of Wetlands, Oceans, and Watersheds
Washington, D.C. 20004

Kristina Nicholas, (former) ORISE Fellow
USEPA Office of Wetlands, Oceans, and Watersheds
Washington, D.C. 20004

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. This report fulfills EPA QA requirements. The research for the data was conducted under the Office of Water approved Quality Assurance Project Plans "Streamflow Duration Assessment Method (SDAM) Development" which was given an ORD ID of J-WECD-0033408-QP-1-0. Funding was provided to Ecosystem Planning and Restoration, RTI, and SCCWRP under contracts EP-C-17-001 and 68HERC21D0008 for data management and analysis, and TRC companies, PG Environmental and Normandeau under EP-C-16-006 and 68HERC22D0002 for data collection.

Disclaimer:

Any mention of trade names, manufacturers or products does not imply an endorsement by the United States Government or the U.S. Environmental Protection Agency. EPA and its employees do not endorse any commercial products, services, or enterprises.

Suggested citation:

Mazor, R.D., Gross, S., Fritz, K.M., Topping, B., James, A., Fertik Edgerton, R., Nicholas, K. 2026. Development and Evaluation of Streamflow Duration Assessment Methods for Regions of the Coterminous United States. Report No. EPA-843-R-25001.

Table of Contents

Contents

1. Introduction 1
1.1 Streamflow duration classes.....2
1.2 Development of the beta methods 2
2. Development of the final methods: Approach 4
2.1 Assembly of regional steering committees 4
2.2 Reach selection 4
2.3 Candidate indicator and metric selection.....6
2.4 Data collection 10
2.5 True streamflow duration class determination 11
2.6 Data preparation 12
2.6.1 Metric calculation.....12
2.6.2 Partition of the dataset into training and testing subsets.....20
2.6.3 Rebalancing uneven reach visits 21
2.7 Metric screening.....21
2.8 Preliminary model development.....22
2.8.1 Select optimal geospatial metrics and perennial indicator metrics.....25
2.8.2 Recursive feature elimination.....25
2.8.3 Assignment to streamflow duration outcomes 26
2.8.4 Performance assessment of preliminary models and model selection.....26
2.9 Model refinement.....29
2.9.1 Stepwise modification of metrics.....29
2.9.2 Single indicators 30
2.10 Performance assessment of the final models.....32
2.10.1 Accuracy at disturbed reach visits 32
2.10.2 Sensitivity to variation in indicators.....33
2.10.3 Sensitivity to time of sampling.....34
2.10.4 Comparison to other SDAMs 36
3. Development of the AW and WM SDAMs 37
3.1 Study areas 37
3.1.1: The Arid West (AW) 37
3.1.2: The Western Mountains (WM) 38

3.2 Data collection 39
3.3 Metric screening 42
3.4 Data stratification and preparation 47
3.5 Model development 50
3.6 Model refinement 57
3.6.1 AW model refinement 57
3.6.2 WM model refinement 61
3.6.3 Single indicators 67
3.7 Performance assessment of the final models 68
3.7.1 Accuracy at disturbed reaches 72
3.7.2 Sensitivity to variation in indicator values 73
3.7.3 Sensitivity to time of sampling 80
3.7.4 Comparison to other SDAMs 85
4. Development of the GP SDAM 89
4.1 Study area 89
4.2 Data collection 90
4.3 Metric screening 91
4.4 Data stratification and preparation 95
4.5 Model development 97
4.6 Model refinement 101
4.6.1 Single indicators 106
4.7 Performance assessment of final models 107
4.7.1 Accuracy at disturbed reaches 109
4.7.2 Sensitivity to variation in indicator values 110
4.7.3 Sensitivity to time of sampling 115
4.7.4 Comparison to other SDAMs 120
5. Development of the NE and SE SDAMs 123
5.1 Study area 123
5.2 Data collection 124
5.3 Metric screening 125
5.3.1 Data corrections 125
5.3.2 Metric Screening 125
5.4 Data stratification and preparation 130

5.5 Model development 132
5.6 Model refinement 137
5.6.1 Northeast Model Refinement 138
5.6.2 Southeast Model Refinement 141
5.6.3 Single indicators 143
5.7 Performance assessment of the final models 145
5.7.1 Accuracy at disturbed sites 147
5.7.2 Sensitivity to variation in indicator values 147
5.7.3 Sensitivity to time of sampling 157
5.7.4 Comparison to other SDAMs 162
6. Data and code availability 165
7. Acknowledgements 165
8. References 168
Appendix A. Glossary of terms 175
Appendix B: Indicators and metrics evaluated for SDAM development 183

1. Introduction

Streamflow duration assessment methods (SDAMs) are rapid, field-based methods to determine flow duration class at the reach scale. The development of SDAMs for the Arid West (AW), Western Mountains (WM), Great Plains (GP), Northeast (NE), and Southeast (SE) regions followed a conceptual framework and set process steps to integrate the three key components of an SDAM development study: hydrological data, indicators, and study reaches (Fritz et al. 2020). These regions are shown in Figure 1. The AW, WM, and the western boundary of the GP are based primarily on boundaries used for wetland delineation found in the Arid West and Western Mountains, Valleys and Coast regional supplements to the U.S. Army Corps of Engineers wetland delineation manual (U.S. Army Corps of Engineers 2008, 2010). The NE, SE, and eastern boundary of the GP are based on boundaries established for ordinary high water mark identification (David et al. 2025).

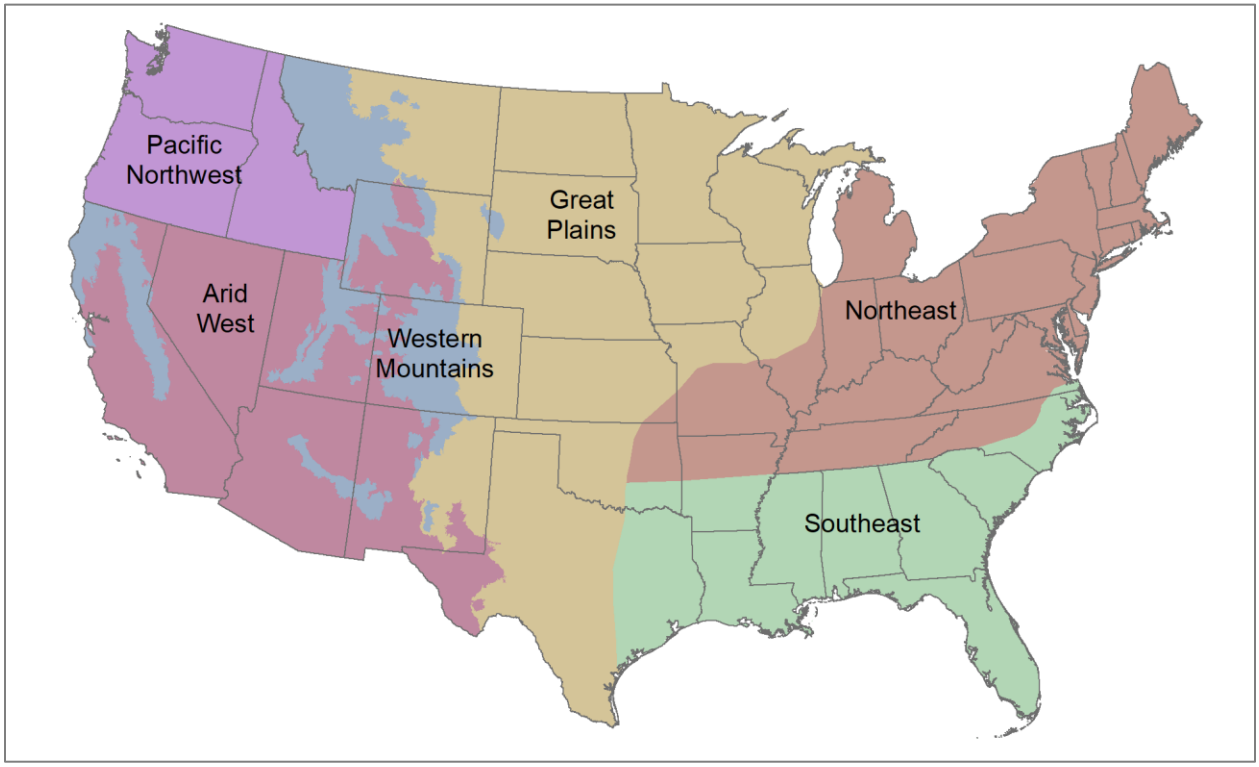


Figure 1. Map of streamflow duration assessment methods regions. Throughout this document, regional abbreviations are used. AW: Arid West. GP: Great Plains. NE: Northeast. PNW: Pacific Northwest. SE: Southeast.

This supplemental document describes the data collection, data analysis, and evaluation steps that resulted in the SDAMs for the AW, WM, GP, NE, and SE regions. Development of regional SDAMs began with the Pacific Northwest (PNW), as described in Nadeau et al. (2015), and that SDAM has been in use since 2011 to inform state and federal regulatory decisions and other resource management decisions in the region. Following the same framework, EPA and the

Corps have developed Regional SDAMs to cover the remaining coterminous United States which are also in use.

This document is a companion to the regional SDAMs for the AW, WM, GP, NE and SE, for those interested in more details on the development of these methods and the underlying data. For a complete description of how to use an SDAM, please see the appropriate Regional User Manual (PNW: Nadeau 2015, AW and WM: Mazor et al. 2024, GP: James et al. 2024, NE and SE: James et al. 2025). The data used to develop all SDAMs can be found on the SDAM website. For more information on the collaborative effort between the U.S. Environmental Protection Agency (EPA) and the U.S. Army Corps of Engineers (Corps) to develop regional SDAMs for nationwide coverage see the Development of Regional SDAMs portion of the website.

1.1 Streamflow duration classes

Streamflow duration governs important ecosystem functions (such as support for aquatic life, sediment transport, and biogeochemical processing rates), and streamflow duration classes are often used to guide watershed management decisions, including assessing the applicability of water quality standards. Our definitions of streamflow duration classes follow those used by Nadeau (2015):

- *Ephemeral reaches* flow only in direct response to precipitation. Water typically flows only during and/or shortly after large precipitation events, the streambed is always above the water table, and stormwater runoff is the primary water source.
- *Intermittent reaches* contain sustained flowing water for only part of the year, typically during the wet season, where the streambed may be below the water table or where the snowmelt from surrounding uplands provides sustained flow. The flow may vary greatly with stormwater runoff.
- *Perennial reaches* contain flowing water continuously during a year of normal rainfall, often with the streambed located below the water table for most of the year. Groundwater typically supplies the baseflow for perennial reaches, but the baseflow may also be supplemented by stormwater runoff or snowmelt.

For these definitions, a reach is a section of stream or river channel along which similar hydrologic conditions exist (e.g., discharge, depth, velocity, or sediment transport dynamics) and consistent drivers of hydrology are evident (e.g., slope, substrate, geomorphology, and confinement). A channel is an area that is confined by banks and a bed and contains flowing water (continuously or not). For SDAMs, the assessment reach length is 40 times the average bankfull channel width, but not less than 40 m and not longer than 200 m.

1.2 Development of the beta methods

Prior to the release of final regional methods, candidate indicators and study reaches with hydrologic data were identified based on regional literature reviews and beta methods were developed for each region (Table 1, see Beta SDAMs website). These beta methods were developed from smaller datasets (i.e., fewer reaches and/or fewer visits per reach), but with the

same sampling methods and following the same general process used to develop the final SDAMs. The beta SDAMs were developed and published to allow practitioners to use SDAMs and provide feedback about both the SDAMs and the process used to develop them that could then inform the development of the final methods. With the release of final SDAMs, continued use of the beta SDAMs is not recommended.

Table 1. Summary of beta SDAMs. Data digital object identifiers (DOI) can be used to access data associated with beta method development at data.gov.

Region	Release date	User manual	Data analysis documentation	Data DOI	Literature review
AW	Mar 2021	Mazor et al. (2021a) (updated Nov 2023)	Mazor et al. (2021b)	10.23719/1523371	McCune and Mazor (2019)
WM	Nov 2021	Mazor et al. (2021c) (updated Nov 2023)	Mazor et al. (2022)	10.23719/1526066	Mazor and McCune (2021)
GP	Sep 2022	James et al. (2023a) (updated Nov 2023)	Eddy et al. (2022)	10.23719/1527943	James et al. (2022a)
NE and SE	Apr 2023	James et al. (2023b) (updated Nov 2023)	Gross et al. (2023)	10.23719/1528743	James et al. (2022b)

2. Development of the final methods: Approach

2.1 Assembly of regional steering committees

As recommended by Fritz et al. (2020) and implemented in the development of the beta methods, the same technical advisory committees were used to help review the development of the final SDAMs. At the outset of the project, we assembled five regional steering committees (RSCs) consisting of technical staff at Corps Districts and USEPA Regional Offices that manage programs where streamflow duration information is often needed (e.g., U.S. Clean Water Act programs, including permitting and enforcement). RSC members were selected based on their expertise in both scientific and programmatic elements relevant to streamflow duration classification needs. The RSCs served several functions in the development process, such as facilitating connections with local experts, identifying resources such as sources of hydrologic data, reviewing candidate indicators for field sampling, identifying potential field sampling locations and relevant literature, identifying and deciding between preliminary models for further refinement and between model refinement options, and reviewing technical products. Some members served on multiple RSCs; all members who participated throughout the development of beta and or final SDAMs are recognized in the Acknowledgements.

2.2 Reach selection

There were two objectives in selecting candidate reaches for this study: first, to include a sufficient number of reaches in each streamflow duration class to capture variability in indicator measurements; and second, to select reaches representing the range of key natural and disturbance gradients within the study areas to support applicability of the method across anticipated conditions.

To improve representativeness within each region, candidate reaches were identified in each state of the coterminous U.S. In the AW, WM, and GP regions, we further attempted to obtain equal representation of likely perennial, intermittent, and ephemeral reaches within major subregions (Figure 2). In addition, reaches were clustered for field sampling efficiency (2-6 reaches within short distances between each other) and to represent the range of streamflow duration characteristics in the local area. Likely streamflow duration class was assigned by examining available hydrologic data, published studies, consultation with local experts, or best professional judgment. Historical aerial imagery available in Google Earth was also evaluated to determine likely streamflow duration classes at candidate reaches (Gallart et al. 2016). Reaches were prioritized for study inclusion based on accessibility (e.g., on public property or with landowner permission), likelihood of being wadeable during baseflow, and the number and type of data sources available to determine true streamflow duration classification.

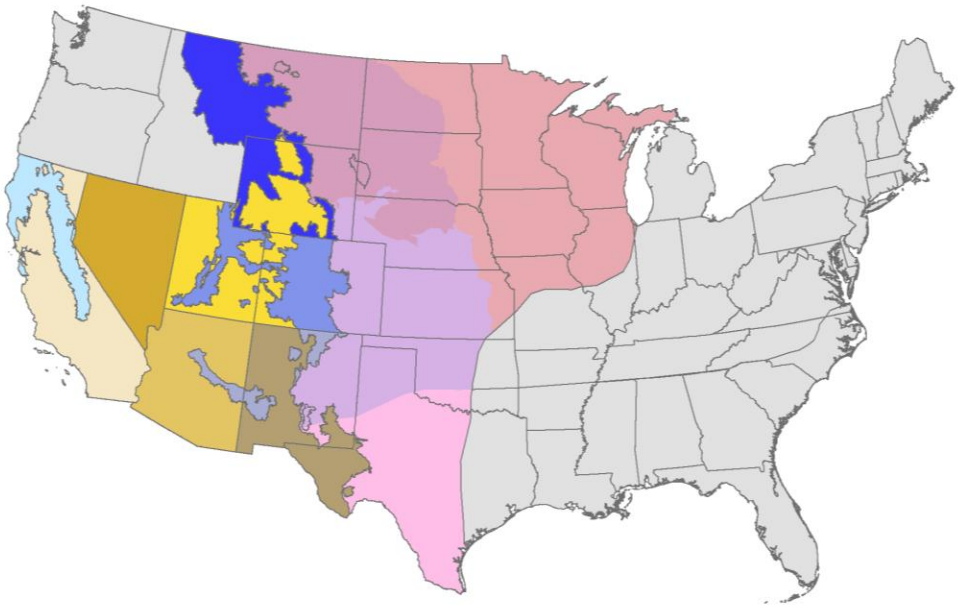


Figure 2. The AW, WM, and GP regions were divided into subregions to facilitate the distribution of candidate reaches for the environmental representativeness of the development datasets. The AW was divided into five subregions (California; Nevada; Arizona; Texas and New Mexico; Utah, Colorado, Wyoming, and Montana) shown in shades of yellow. The WM was divided into four subregions (California and Nevada; Arizona and New Mexico; Utah and Colorado; Wyoming, Montana, and South Dakota) shown in shades of blue. The GP was divided into four subregions (Southern Plains, Central Prairie, Northern Prairie, and Upper Midwest) shown in shades of pink. The NE and SE regions relied on placing clusters of sites in each state to capture the range of regional variation (rather than relying on subregions), and no sampling occurred in the PNW as part of this study; these regions are shown in gray.

Because this process resulted in low numbers of potential ephemeral reaches in most regions, we identified more candidate ephemeral reaches by generating high-resolution stream networks from 30-m digital elevation models over 55 x 55-km squares centered on selected candidate reaches with known accessibility. We then generated a stream network and assigned Strahler stream orders using the ESRI hydrology toolset (Tarboton et al. 1991, ESRI 2024). We selected additional candidate reaches with stream order of 1 with the assumption that these would likely be ephemeral. In addition, timeseries of available aerial imagery was reviewed to identify potential intermittent and ephemeral streams based on the absence of visible flow across most or all seasons.

This process resulted in a large dataset of candidate reaches with tentative determinations of streamflow duration class (Table 2).

Table 2. Candidate reaches identified from the USGS stream gage dataset and other sources (e.g., literature, local experts, other gauges, digital elevation models).

Region	Total candidate reaches	USGS candidate reaches				Other candidate reaches
		Total	Perennial	Intermittent	Ephemeral	
AW	717	520	301	169	50	197
WM	1175	857	804	51	2	318
GP	3565	2244	2298	550	96	621
NE	5494	4507	4300	155	22	987
SE	3199	2572	2216	345	11	627

2.3 Candidate indicator and metric selection

For each region, candidate indicators were identified that are supported by the scientific literature (AW: McCune and Mazor 2019; WM: Mazor and McCune 2021; GP: James et al. 2022a; NE and SE: James et al. 2022b) or other methods to distinguish between two or more streamflow duration classes in the region (Table 3).

Table 3. Methods in use in each region identified during literature review.

Region	Other methods used in the region
AW	Pacific Northwest (Nadeau 2015) New Mexico (NMED 2011)
WM	Pacific Northwest (Nadeau 2015) New Mexico (NMED 2011)
GP	New Mexico (NMED 2011)
NE	North Carolina (Dorney and Russell 2018) Tennessee (TDEC 2020) James City County, Virginia (James City County 2009) Ohio (Ohio EPA 2020)
SE	North Carolina (Dorney and Russell 2018)

Following input from the RSC, these candidate indicators were then screened using the criteria described by Fritz et al. (2020), including:

Primary criteria

- *Consistency*: Does the indicator consistently discriminate among streamflow duration classes (e.g., demonstrated in multiple studies)?

Indicators versus metrics

Indicators (Table 4) are measurable characteristics of a reach related to streamflow duration. Some are drivers (e.g., precipitation) or responses to streamflow duration (e.g., riparian vegetation), while others may be neither, but rather have a predictable association with streamflow duration (e.g., slope).

Some indicators must be transformed into numeric *metrics* (Appendix B: Indicators and metrics evaluated for SDAM development) to be used in an SDAM. For example, aquatic invertebrates are an indicator of streamflow duration, from which several metrics (e.g., total abundance, number of mayfly taxa) may be calculated.

Once a metric is selected for inclusion in an SDAM, it will be referred to as an *indicator*.

- *Repeatability:* Can different practitioners take similar measurements, given sufficient training and standardization?
- *Defensibility:* Does the indicator have a rational mechanistic relationship with streamflow duration, as either a response or a driver?
- *Rapidness:* Can the indicator be measured during a one-day reach visit (even if subsequent lab analyses are required)?
- *Objectivity:* Does the indicator rely on objective (often quantitative) measures, as opposed to subjective judgments of practitioners?

Secondary criteria

- *Robustness:* Can the indicator be measured or interpreted accurately in disturbed reaches (e.g., poor water quality may affect the expression of some biological indicators)?
- *Practicality:* Can practitioners realistically sample the indicator with typical capacity, skills, and resources?

Candidate indicators were included in the study (Table 4) if they met all the primary criteria and at least one of the secondary criteria. In addition, all indicators from existing methods used in a region were collected by default. Desktop geospatial indicators (derived using a geographic information system and applicable spatial datasets) that characterize mechanisms affecting streamflow duration and have been explored in other streamflow duration classification tools (Eng et al. 2016, Jaeger et al. 2019, Grantham et al. 2022) were also included in the analysis. Indicators that were direct measures of water presence were collected but not included among candidate indicators because some of these indicators were used to confirm the true streamflow duration classification of some study reaches (see Section 2.5 True streamflow duration class determination).

Table 4. Candidate indicators evaluated in the present study grouped by indicator type. The Origin column indicates the source of the indicator and its measurement process: NM: New Mexico Hydrology Protocol (NMED 2011); PNW SDAM: Pacific Northwest (Nadeau 2015); NC: North Carolina (Dorney and Russell 2018); OH: Ohio (Ohio EPA 2020); OTH: indicator identified through one of the regional literature reviews in Table 1 but not previously used in an SDAM. A single asterisk (*) designates hydrologic indicators that are considered direct measures of water presence. ^{NESE} designates indicators that were only measured in the NE and SE regions. ^{AWWMGP} designated indicators that were only measured in the AW, WM, and GP regions.

Candidate indicator	Description	Origin
Geomorphic indicators		
Sinuosity	Scored estimate of the curviness of the stream channel	NM/NC
Slope	Channel slope measured with a handheld clinometer	PNW SDAM
Bankfull width	Width of the channel at bankfull height	PNW SDAM

Candidate indicator	Description	Origin
Floodplain channel dimensions	Scored estimate of the extent of channel entrenchment and connectivity to the floodplain (scored or directly measured as entrenchment ratio—i.e., the ratio of floodprone area width to bankfull channel width).	NM/OH
Particle size/stream substrate sorting	Scored estimate of the extent of evidence of substrate sorting within the channel	NM/NC
In-channel structure/riffle pool sequence	Scored estimate of the diversity and distinctiveness of riffles, pools, and other flow-based microhabitats	NM/NC
Sediment deposition on plants and debris	Scored estimate of the extent of evidence of sediment deposition on plants and on debris within the floodplain	NM/NC
Alluvial deposits [NESE]	Scored estimate of abundance of recent alluvial deposits within the stream or on the floodplain	NC
Active or relict floodplain [NESE]	Scored estimate of the strength of active or relic floodplain characteristics surrounding the reach	NC
Continuity of channel bed and bank [NESE]	Scored estimate of the continuity of bed and banks along the channel	NC
Depositional bars and benches [NESE]	Scored estimate of the prevalence of depositional bars and benches within the channel	NC
Grade control [NESE]	Scored estimate of extent of grade controls (e.g., boulders, bedrock, woody jams) within the reach	NC
Headcuts [NESE]	Scored estimate of abundance and size of headcuts within the reach	NC
Natural valley [NESE]	Scored estimate of the strength of a natural valley characteristics surrounding the reach	NC
Streambed substrate characteristics [NESE]	Dominant and significant substrate types and size classes within the reach	OH
Hydrologic indicators		
Surface and subsurface flow*	Estimate of the percent of the reach-length with surface and subsurface flow	PNW SDAM
Isolated pools*	Number of pools in the channel without any connection to flowing surface water	PNW SDAM
Water in channel*	Scored estimate of the extent of surface flow in the channel	NM/NC
Seeps and springs*	Presence/absence of springs or seeps within one-half channel width of the channel	NM

Candidate indicator	Description	Origin
Hydric soils	Presence/absence of hydric soils within the channel, measured at up to three locations	NM/NC
Soil moisture and texture*	Extent of soil saturation and texture measured at three locations in the channel	OTH
Woody jams	Number of woody jams within the channel	OTH
Leaf litter [NESE]	Scored estimate of the abundance of leaf litter in the channel	NC
Maximum pool depth [NESE]*	Maximum dept of pools within the channel	OH
Organic drift lines [NESE]	Scored estimate of the prevalence of organic drift lines	NC
Biological indicators		
Algae	Visual estimate of the percent of streambed covered by live or dead algal growth	NM/NC/OTH
Stream shading	Percent shade-providing cover above the streambed measured with a densiometer at three locations	OTH
Hydrophytic plant species	Number of obligate (OBL) or facultative wet (FACW)-rated plants (as listed in Lichvar et al. 2016) growing within the channel or a half-channel width from the channel	PNW SDAM/NC
Fish	Estimate of the overall abundance of fish (other than non-native mosquitofish) in the channel	NM/NC
Aquatic invertebrates	Abundance and/or composition of aquatic invertebrates collected from the channel	PNW SDAM/NC/NM
Amphibians	Estimate of the overall abundance of amphibians within the channel. In the East, composition was also measured.	PNW SDAM/NC
Mosses and liverworts	Visual estimate of the percent of streambed and banks covered by live or dead bryophytes or liverworts	OTH
Differences in vegetation (riparian corridor)	Scored estimate of the distinctiveness of vegetation in the riparian corridor compared to surrounding upland vegetation	NM
Absence of rooted upland plants in the streambed	Scored estimate of the extent of rooted upland plants growing within the streambed	NM/NC
Iron-oxidizing bacteria or fungi	Presence of oily sheens indicative of iron-oxidizing bacteria or fungi within the assessment reach. In the East, abundance was visually estimated.	NM/NC

Candidate indicator	Description	Origin
Presence of aquatic or semi-aquatic snakes	Presence of aquatic or semi-aquatic snakes (e.g., most garter snake species) in the channel. Not measured in the Eat.	PNW SDAM
Fibrous rooted plants	Scored estimate of the abundance of fibrous roots in the streambed. Only measured in the East	NC
Geospatial indicators		
Location and watershed characteristics	Latitude, longitude, stream order, drainage area and elevation	OTH/NC/OH
Precipitation	30-y normal mean annual and monthly precipitation (PRISM climate data; Hart and Bell 2015).	OTH
Temperature	30-y normal mean, maximum, and minimum annual temperature (PRISM climate data; Hart and Bell 2015).	OTH
Snow persistence	Long-term mean snow persistence between January 1 and July 3 (Hammond et al. 2017)	OTH

2.4 Data collection

During each field visit to a study reach, a suite of candidate indicators (Table 4) was measured following the standard protocols. Zimmerman and Olson (2019) was used for initial data collection in the AW and subsequent updates (USEPA 2019, 2020) were used for the remaining data collection. The USEPA (2019, 2020) protocols included all indicators measured in Zimmerman and Olson (2019) with some modifications to indicator scoring to facilitate comparisons with methods used in the eastern US, as well as a few new indicators from methods used in the eastern US, as designated in Table 4.

The number of visits per reach varied among and within regions (Table 5). Note: the 23 reaches in the Caribbean (CB; specifically, Puerto Rico and the U.S. Virgin Islands), were originally sampled as part of the SE region but during development of the beta methods were found to be distinct from the rest of the SE and partitioned off as a separate region. There was not enough data collected in the CB to follow the same method development process, so a method was not produced.

Table 5. Number of reaches with a given number of reach visits by region. CB: Caribbean.

Region	1 visit	2 visits	3 visits	4 visits	5 visits	6 or more visits	Max # visits	Total visits
AW	62	21	88	16	2	5	8	480
WM	75	14	52	17	5	38	9	641
GP	100	13	7	2	5	160	7	1141
NE	73	12	0	3	121	0	5	714
SE	50	4	1	0	108	0	5	571
CB	13	0	0	0	10	0	5	63

2.5 True streamflow duration class determination

True streamflow duration class was determined from multiple sources whenever possible. Continuous timeseries data was obtained from nearby USGS stream gages, if available. Wherever possible, pairs of Stream Temperature, Intermittence, and Conductivity (STIC) loggers (Chapin et al. 2014) were deployed at reaches where repeat visits were planned. Loggers were deployed in riffles or other areas expected to experience drying first; pools were avoided because loggers deployed there could erroneously indicate continuous flow in streams with discontinuous surface water. STIC loggers were calibrated with distilled water following the process described by Schumacher and Fritz (2019) to identify a minimum threshold conductance for detecting inundated conditions.

Reaches where true streamflow duration class could be determined based on multiple data sources (e.g., hydrologic data and expert knowledge) were categorized as “preferred” for study inclusion. Reaches classified based solely on interpretation of USGS stream gage data without consultation of a local expert were categorized as “USGS gage” reaches. Reaches classified through local expertise alone were categorized as “acceptable” and included in the study to fill gaps in study regions where an insufficient number of “preferred” and “USGS gage” reaches classified as intermittent or ephemeral could be identified.

Timeseries data from gages or STIC loggers were analyzed by applying the flowchart developed by Kelso and Fritz (2021, Figure 3), which was informed by existing definitions (Hedman and Osterkamp 1982, Hewlett 1982). This flowchart results in three “epicenter” classes that most closely correspond to existing definitions of perennial, intermittent, or ephemeral streams. Additionally, several “fuzzy” classes reflecting intermediate conditions could also be identified. The flowchart was applied independently to data from gages or both loggers from a reach. If the flowchart resulted in fuzzy or ambiguous classifications, or if the two loggers had conflicting classifications, records were visually inspected in combination with temperature data from the loggers as well as additional information to inform the determination of true streamflow duration class. In addition to the timeseries data, field observations (see direct hydrologic indicators in Table 4), reach photos, historical aerial imagery, published studies, and consultation with local experts were used to determine the true streamflow duration class (i.e.,

perennial, intermittent, or ephemeral). Across the five SDAM regions, data from 35 reaches where no true streamflow duration class could be determined were withheld from all analyses.

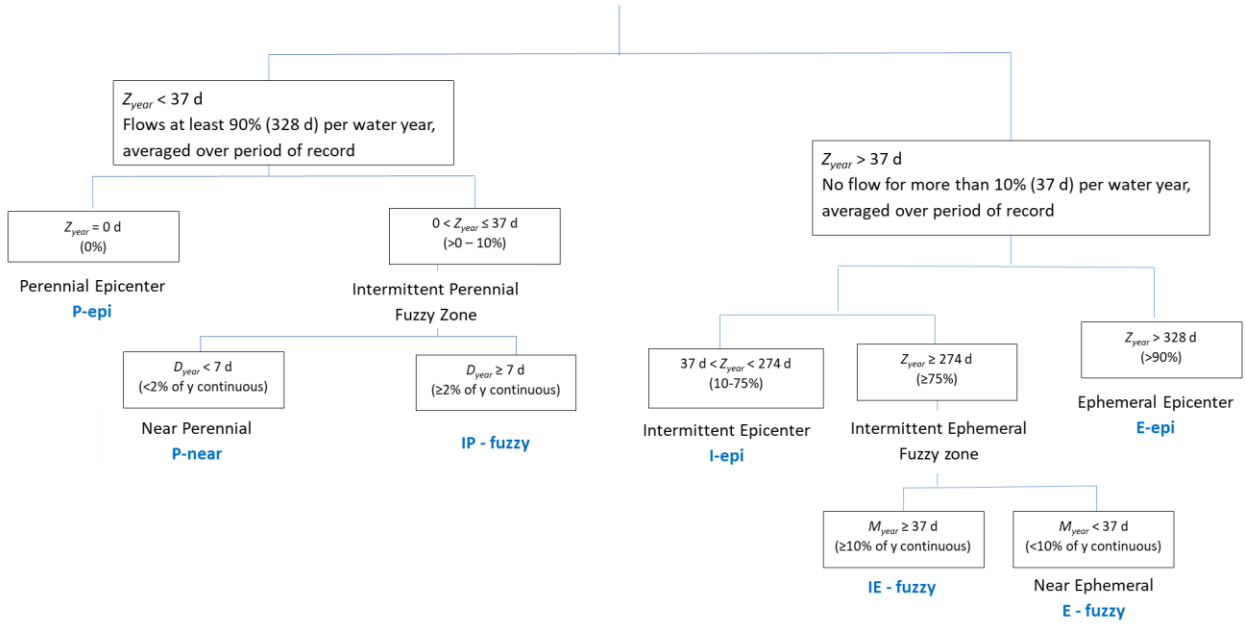


Figure 3. Flowchart used to determine the true streamflow duration class of reaches based on continuous datasets of water presence (e.g., USGS stream gages). Zyear: Total number of zero-flow days per water year, averaged over the period of record. Myear: Maximum continuous duration of flow (in days) per water year, averaged over the period of record. Dyear: Maximum continuous duration of no-flow (in days) per water year, averaged over the period of record. For USGS gage datasets, at least 20 years of data were analyzed whenever possible (Kelso and Fritz 2021), but datasets of 1-2 years were used for STIC logger datasets.

2.6 Data preparation

Data preparation varied by region due to differences in data collection effort, and in particular, the frequency and number of revisits. In general, data preparation consisted of steps described in Sections 2.5 True streamflow duration class determination through 2.6.1 Metric calculation (more details are provided in regional sections of this document).

2.6.1 Metric calculation

Measurements of indicators shown in Table 4 were transformed into quantitative metrics for further analysis (Appendix B). A total of 129 metrics were calculated: 66 biological metrics; 17 geomorphological metrics; 4 hydrological metrics; and 42 geospatial metrics. Because slightly different sets of indicators were measured in different regions, and because some metrics were only considered in regions where they were used in pre-existing SDAMs within the region, the number of candidate metrics varied among regions as well: 89 metrics were candidates in the AW and WM, 95 in the GP, and 113 in the NE and SE.

2.6.1.1 Indicator species analysis

Stream drying is a strong factor controlling not only the distribution and abundance of macroinvertebrates, but also their taxonomic composition. Different kinds of macroinvertebrates have adaptations to withstand, avoid, or recover quickly following stream drying, whereas other macroinvertebrates lack such adaptations. Some SDAMs use indicator metrics based on expert knowledge of the life histories and adaptations of macroinvertebrate taxa that are indicative of a particular streamflow duration class in different regions (NCDWQ 2010, Nadeau 2015). Here, we use an analytical approach called indicator species analysis (ISA) to identify groups of different macroinvertebrate taxonomic groups that occur more often in one or two streamflow duration classes compared to others within this study's different SDAM regional datasets. ISA determines if sets of samples differ in the relative abundance and occurrence frequencies of different taxa (Cáceres and Legendre 2009). For our purposes, the different sets of samples we are interested in are those collected from different streamflow duration classes (ephemeral, intermittent, and perennial). There are two quantities used in ISA that describe the affinity of macroinvertebrate taxa for a streamflow duration class. Specificity (A) is the probability that a reach belongs to a streamflow duration class or pair of classes given that a macroinvertebrate taxon occurred in a sample collected from the reach. Sensitivity (B) is the probability of finding the species in reaches belonging to a streamflow duration class or pair of classes. The square root of the product of specificity and sensitivity is the indicator value (IV) index. The ISA then uses the difference between the maximum observed IV for a streamflow duration class and the mean IV generated from random permutations (n = 9999) to derive a significance test for each taxon. The ISA was ran using the multiplatt function within the *indicspecies* package (version 1.7.14; De Cáceres et al. 2010). IndVal.g was used to correct for unequal sample sizes among streamflow duration classes. We designated ISA indicator taxa to be those identified as having a significant association with a streamflow duration class or pair of classes and were observed in at least 5% of the reach visits with macroinvertebrates for a regional, aggregate regional, or subregional dataset. We only considered ISA macroinvertebrate taxa for pairs of streamflow duration classes if they were adjacent in streamflow duration (i.e., intermittent + perennial or intermittent + ephemeral) and derived ISA macroinvertebrate taxa metrics by combining taxa identified from class pairs (e.g., intermittent and perennial indicator taxa) with any taxa identified for the individual classes within the pair (e.g., intermittent indicator taxa and perennial indicator taxa). Separate ISA were run for datasets collected within each of the SDAM regions, select subregions, and their aggregate combinations (Table 6).

Table 6. The number reach visits with macroinvertebrates, records (taxon occurrences, up to 11 per visit, across all reach visits), and number of macroinvertebrate taxa included in the Indicator Species Analysis (ISA) completed for individual and combinations of original SDAM regions. AW = Arid West region, WM = Western Mountains region, NE = Northeast region, SE = Southeast region, NGP = Central Prairie + Northern Prairie + and Upper Midwest subregions, SGP = Southern Plains subregion.

Region/subregion	Reach visits	Records	Taxa
AW	213	2202	98
WM	416	4742	133
West	629	6944	138
NGP	523	2703	100
SGP	162	1057	77
GP	685	3760	105
NE	514	3013	93
SE	424	2273	96
East	938	5286	105

The total dataset included 15,990 records across 149 taxa. Taxa abundance ranged from 0 to 11, where 11 represented 11 or more individuals recorded during a reach visit. The dataset only included those aquatic life stages represented by living or non-living individuals, where non-living individuals could be represented by shells, exuviae, or cases. Taxonomic level was family for mollusks and insects (except Tipulidae, Pediciidae, and Limonidae were combined into Tipuloidea), order for crustaceans and mites, and class or phylum for cnidarians, annelids, nematodes, horsehair worms, proboscis worms, and flatworms. A total of 1207 reach visits were excluded from the ISA analyses because they did not have aquatic macroinvertebrates recorded. An additional 438 records of macroinvertebrates were excluded because they were identified at higher taxonomic levels than used in the ISA (Table 7).

Table 7. Macroinvertebrate records (reach-visit occurrences) excluded from Indicator Species Analysis (ISA) because taxonomic identifications were at higher levels than used by the ISA and the number of associated taxonomic groups and records included in ISA.

Taxonomic ID	Excluded records	Number of groups included	Included records
Bivalvia	9	5 families	314
Gastropoda	68	13 families	812
Arthropoda	1	116 families & 9 orders	14,607
Crustacea	9	8 orders	1,485
Coleoptera	42	19 families	1,469
Diptera	46	23 families	3,396
Ephemeroptera	40	13 families	2,564
Hemiptera	10	14 families	733
Lepidoptera	1	1 family	24
Megaloptera	2	2 families	165
Odonata	29	10 families	900
Plecoptera	40	9 families	1,311
Trichoptera	141	24 families	2,220

A total of 46 taxa were identified as indicator taxa of an SDAM class or pair of classes within at least one region or regional aggregation (Table 8). Perennial indicator taxa were identified for each region and pairs of regions, ranging from two (NGP dataset and SGP dataset) to 15 taxa (AW and WM aggregated dataset). No intermittent indicator taxa were identified for any regions; however, there were intermittent + perennial indicator taxa identified in the AW, WM, AW+WM, SGP, NE, SE, and NE+SE invertebrate datasets. Ephemeral and ephemeral + intermittent indicator taxa were identified in the NGP, SGP, NGP+SGP, NE, SE, and NE+SE datasets. The caddisfly family, Hydropsychidae, and the mayfly family, Heptageniidae, were the most widely identified perennial indicator taxa by representing in seven of the nine regional datasets (Table 9). The mayfly family, Baetidae, the beetle family, Dytiscidae, and the mosquito family, Culicidae were the most widely identified intermittent + perennial, ephemeral + intermittent, and ephemeral indicator taxa, respectively (Table 8). Eight of the 46 taxa were identified as indicators of different streamflow-duration classes or pairs of classes across regions and regional combinations. Five of these taxa (Brachycentridae, Leptostomatidae, Hydroptilidae, Tipuloidea, and Lymnaeidae) were identified as perennial and intermittent + perennial indicators in different regions or regional aggregations. The snail family, Physidae was identified as an intermittent + perennial indicator taxon in WM and AW+WM datasets, but as an ephemeral + intermittent indicator in NGP, SGP, and NGP+SGP datasets. Similarly, the crustacean order, Amphipoda, was identified as an intermittent + perennial indicator taxon in WM and AW+WM datasets but as an ephemeral + intermittent indicator in the NE+SE dataset. Lastly, the flatworm group, Turbellaria was identified as a perennial indicator taxon in the NGP+SGP dataset, an intermittent + perennial indicator taxon in the WM and AW+WM datasets, and an ephemeral + intermittent indicator taxon in the SE dataset.

Table 8. Aquatic invertebrate taxa identified using Indicator Species Analysis (ISA) as indicators for individual or pairs of streamflow-duration classes for individual or aggregations of SDAM regions. The percent frequency (F) that taxa were collected across reach visits, indicator values (IV) and p-values. AW = Arid West region, WM = Western Mountains region, NE = Northeast region, SE = Southeast region, NGP = Central Prairie + Northern Prairie + and Upper Midwest subregions, SGP = Southern Plains subregion.

Region/subregion	SDAM class	Taxon	F (%)	IV	*p*-value
AW	Perennial	Hydropsychidae	34.3	0.677	0.0019
		Elmidae	34.3	0.668	0.0019
		Leptophlebiidae	18.3	0.517	0.0169
		Heptageniidae	16.0	0.483	0.0244
		Brachycentridae	15.0	0.469	0.0326
		Lepidostomatidae	13.6	0.434	0.0410
		Leptoceridae	10.3	0.426	0.0337
		Pteronarcyidae	8.4	0.397	0.0360
		Perlidae	8.0	0.386	0.0372
	Perennial + Intermittent	Baetidae	66.2	0.835	0.0001
		Ceratopogonidae	29.1	0.554	0.0403
WM	Perennial	Elmidae	37.9	0.663	0.0001
		Hydropsychidae	28.3	0.644	0.0001
		Heptageniidae	35.7	0.634	0.0001

Region/subregion	SDAM class	Taxon	F (%)	IV	*p*-value
		Ephemerellidae	28.8	0.594	0.0001
		Leptophlebiidae	27.8	0.575	0.0001
		Chloroperlidae	24.2	0.527	0.0003
		Rhyacophilidae	15.8	0.468	0.0005
		Empididae	15.6	0.436	0.0015
		Perlidae	15.3	0.429	0.0011
		Leptohyphidae	13.2	0.428	0.0010
		Hydroptilidae	16.3	0.425	0.0047
WM	Perennial	Glossosomatidae	14.4	0.418	0.0027
		Pteronarcyidae	7.9	0.348	0.0035
		Uenoidae	7.2	0.322	0.0135
	Perennial + Intermittent	Baetidae	60.9	0.810	0.0001
		Simuliidae	49.2	0.720	0.0001
		Tipuloidea	34.5	0.599	0.0001
		Perlodidae	24.5	0.523	0.0003
		Brachycentridae	24.5	0.519	0.0005
		Sphaeriidae	22.5	0.502	0.0006
		Lepidostomatidae	18.7	0.452	0.0023
		Physidae	18.5	0.437	0.0073
		Psychodidae	13.4	0.387	0.0098
		Ameletidae	13.2	0.384	0.0091
		Amphipoda	12.9	0.380	0.0108
		Turbellaria	11.8	0.355	0.0256
		Lymnaeidae	11.0	0.351	0.0258
AW+WM	Perennial	Hydropsychidae	30.4	0.657	0.0001
		Elmidae	36.7	0.657	0.0001
		Heptageniidae	29.1	0.570	0.0001
		Leptophlebiidae	24.6	0.542	0.0001
		Ephemerellidae	23.5	0.524	0.0001
		Brachycentridae	21.3	0.483	0.0002
		Chloroperlidae	19.6	0.466	0.0003
		Leptohyphidae	17.8	0.445	0.0007
		Perlidae	12.9	0.400	0.0006
		Glossosomatidae	12.4	0.396	0.0005
		Empididae	14.0	0.393	0.0035
		Rhyacophilidae	12.6	0.381	0.0018
		Pteronarcyidae	8.1	0.360	0.0010
		Leptoceridae	9.0	0.345	0.0029
		Uenoidae	5.2	0.268	0.0155
	Perennial + Intermittent	Baetidae	62.8	0.817	0.0001
		Simuliidae	48.0	0.692	0.0001
		Physidae	22.7	0.478	0.0005
		Sphaeriidae	20.3	0.467	0.0003
		Perlodidae	19.4	0.461	0.0003
		Hydroptilidae	18.6	0.443	0.0011
		Lepidostomatidae	17.0	0.427	0.0009
		Amphipoda	16.46	0.423	0.0011

Region/subregion	SDAM class	Taxon	F (%)	IV	*p*-value
		Turbellaria	13.2	0.375	0.0068
		Psychodidae	12.1	0.364	0.0064
		Lymnaeidae	9.8	0.328	0.0141
		Ameletidae	9.5	0.323	0.0201
NGP	Perennial	Hydropsychidae	24.7	0.584	0.0026
		Heptageniidae	10.9	0.384	0.0459
	Ephemeral	Corixidae	24.1	0.589	0.0039
	Ephemeral + Intermittent	Physidae	27.2	0.548	0.0120
NGP	Ephemeral + Intermittent	Dytiscidae	17.4	0.489	0.0133
		Hydrophilidae	12.2	0.393	0.0427
SGP	Perennial	Psephenidae	18.5	0.521	0.0078
		Lymnaeidae	8.6	0.384	0.041
	Perennial + Intermittent	Coenagrionidae	45.7	0.694	0.0020
	Ephemeral	Culicidae	8.0	0.608	0.0002
		Haliplidae	9.9	0.416	0.0253
	Ephemeral + Intermittent	Physidae	34.0	0.611	0.0086
NGP+SGP	Perennial	Heptageniidae	12.0	0.396	0.0100
		Turbellaria	8.5	0.318	0.0377
		Leptohyphidae	5.7	0.304	0.0219
	Ephemeral	Culicidae	5.5	0.502	0.0001
		Haliplidae	5.5	0.291	0.0377
	Ephemeral + Intermittent	Physidae	28.8	0.563	0.0003
		Dytiscidae	15.0	0.451	0.0046
		Hydrophilidae	10.7	0.371	0.0143
NE	Perennial	Hydropsychidae	24.7	0.571	0.0001
		Perlidae	14.0	0.486	0.0001
		Philopotamidae	13.6	0.404	0.0090
		Elmidae	10.5	0.336	0.0420
		Gomphidae	5.4	0.317	0.0113
	Perennial + Intermittent	Simuliidae	12.3	0.361	0.0444
	Ephemeral + Intermittent	Dytiscidae	16.2	0.419	0.0124
		Hydrophilidae	7.0	0.325	0.0175
SE	Perennial	Heptageniidae	29.2	0.658	0.0001
		Hydropsychidae	15.3	0.544	0.0001
		Tipuloidea	17.2	0.460	0.0048
		Perlidae	13.0	0.371	0.0080
		Ephemerellidae	7.5	0.379	0.0076
		Cordulegastridae	6.8	0.363	0.0071
		Philopotamidae	8.2	0.355	0.0141
		Calopterygidae	9.4	0.328	0.0440
	Perennial + Intermittent	Baetidae	15.6	0.408	0.0294
	Ephemeral	Culicidae	12.0	0.489	0.0004
	Ephemeral + Intermittent	Amphipoda	35.8	0.641	0.0001
		Isopoda	25.2	0.586	0.0001
		Dytiscidae	20.3	0.475	0.0064
NE+SE	Perennial	Heptageniidae	28.4	0.583	0.0001

Region/subregion	SDAM class	Taxon	F (%)	IV	*p*-value
		Hydropsychidae	20.5	0.560	0.0001
		Perlidae	13.5	0.445	0.0001
		Ephemerellidae	10.4	0.391	0.0003
		Philopotamidae	11.2	0.385	0.0004
		Gomphidae	9.4	0.346	0.0014
		Corydalidae	7.7	0.304	0.0089
		Elmidae	8.4	0.300	0.0177
		Calopterygidae	6.3	0.293	0.0077
NE+SE	Perennial	Cordulegastridae	6.1	0.291	0.0090
		Peltoperlidae	5.6	0.270	0.0141
	Perennial + Intermittent	Capniidae	14.1	0.377	0.0054
		Simuliidae	11.5	0.351	0.0052
		Perlodidae	9.4	0.316	0.0140
NE+SE	Perennial + Intermittent	Nemouridae	8.1	0.288	0.0407
		Gerridae	6.9	0.272	0.0356
		Ameletidae	6.6	0.266	0.0444
	Ephemeral + Intermittent	Amphipoda	30.7	0.562	0.0001
		Isopoda	27.3	0.541	0.0001
		Dytiscidae	18.0	0.445	0.0005
		Hydrophilidae	5.0	0.275	0.0072
	Ephemeral	Culicidae	7.6	0.366	0.0006

The AW, WM, and AW+WM perennial indicator taxa had 4 of 9, 6 of 14, and 6 of 15 taxa, respectively, in common with the 16 perennial indicator taxa used in the Pacific Northwest SDAM (Nadeau 2015; Table 9). The AW, WM, and AW+WM perennial indicator had 5 of 9, 6 of 14, and 6 of 15 taxa, respectively, in common among the 29 perennial indicator taxa used in the New Mexico Level 2 Hydrology Protocol for the Determination of uses Supported by Ephemeral, Intermittent, and Perennial Waters (NMED 2011). The NE, SE, and NE+SE perennial indicator taxa were all in common with the 35 perennial indicator taxa used in the North Carolina Methodology for Identification of Intermittent and Perennial Streams and Their Origins (NCDWQ 2010). Only one or two perennial indicator taxa identified in the NGP, SGP, and NGP+SGP were in common with the Pacific Northwest, New Mexico, or North Carolina perennial taxa lists but there were also only two or three perennial indicator taxa identified in the NGP, SGP, and NGP+SGP datasets (Table 9, Table 10).

Table 9. Aquatic macroinvertebrate indicator lists for existing methods (NC: NCDWQ 2010, NM: NMED 2011, PNW: Nadeau 2015) and from Indicator Species Analysis (ISA) for SDAM regions, aggregate regions, and select subregions. Indicator aquatic invertebrate taxa for perennial (P), perennial and intermittent (P+I), ephemeral and intermittent (E+I), or ephemeral (E) reaches. Indicator taxa are organized by higher taxa representing orders, classes or phyla.

Taxon	NC	PNW	NM	AW	WM	West	NGP	SGP	GP	NE	SE	East
Ephemeroptera												
Ameletidae					P+I	P+I					P+I	P+I
Baetidae	P			P+I	P+I	P+I						
Caenidae	P		P									
Ephemerellidae	P		P		P	P					P	P

Taxon	NC	PNW	NM	AW	WM	West	NGP	SGP	GP	NE	SE	East
Ephemeridae	P		P									
Heptageniidae	P		P	P	P	P	P		P		P	P
Leptohyphidae					P	P			P			
Leptophlebiidae	P			P	P	P						
Siphlonuridae	P											
Plecoptera												
Capniidae												P+I
Chloroperlidae					P	P						
Nemouridae												P+I
Peltoperlidae	P		P									P
Perlidae	P	P	P	P	P	P				P	P	P
Perlodidae	P		P		P+I	P+I						P+I
Pteronarcyidae		P		P	P	P						
Trichoptera												
Brachycentridae				P	P+I	P						
Glossosomatidae		P			P	P						
Hydropsychidae	P	P	P	P	P	P	P			P	P	P
Hydroptilidae					P	P+I						
Lepidostomatidae	P		P	P	P+I	P+I						
Leptoceridae				P		P						
Limnephilidae	P											
Molannidae	P		P									
Odontoceridae	P		P									
Philopotamidae	P	P	P							P	P	P
Polycentropodidae	P		P									
Psychomyiidae	P		P									
Rhyacophilidae	P	P	P		P	P						
Uenoidae					P	P						
Hemiptera												
Corixidae							E					
Gerridae												P+I
Odonata												
Aeshnidae	P		P									
Calopterygidae	P	P	P								P	P
Coenagrionidae								P+I				
Cordulegastridae	P	P	P								P	P
Gomphidae	P	P	P							P		P
Libellulidae	P											
Megaloptera												
Corydalidae	P	P	P									P
Sialidae	P		P									
Coleoptera												
Dytiscidae							E+I		E+I	E+I	E+I	E+I
Elmidae	P	P	P	P	P	P				P		P
Haliplidae								E	E			
Helichus	P		P									
Hydrophilidae							E+I		E+I	E+I		E+I

Taxon	NC	PNW	NM	AW	WM	West	NGP	SGP	GP	NE	SE	East
Psephenidae	P	P	P					P				
Diptera												
Ceratopogonidae				P+I								
Culicidae								E	E		E	
Empididae					P	P						
Psychodidae					P+I	P+I						
Ptychopteridae	P		P									
Simuliidae						P+I				P+I		P+I
Tipula	P		P		P+I							
Tipuloidea					P+I						P	
Crustacea												
Amphipoda					P+I	P+I					E+I	E+I
Isopoda											E+I	E+I
Planaria												
Turbellaria					P+I	P+I			P			
Ancylidae	P	P	P									
Gastropoda												
Hydrobiidae+Tateidae		P										
Lymnaeidae					P+I			P				
Physidae					P+I	P+I	E+I	E+I	E+I			
Pleuroceridae	P	P	P									
Planorbidae	P											
Bivalvia												
Margaritiferidae		P										
Sphaeriidae					P+I	P+I						
Unionidae	P		P									

Table 10. Total number of indicator taxa for each streamflow duration class. NA = Not Applicable.

Streamflow duration class	NC	PNW	NM	AW	WM	West	NGP	SGP	GP	NE	SE	East
P	35	16	29	9	14	15	2	2	3	5	8	11
P+I	NA	NA	NA	2	13	12	0	1	0	1	1	6
E+I	NA	NA	NA	0	0	0	3	1	3	2	3	4
E	NA	NA	NA	0	0	0	1	2	2	0	1	1

For each reach visit, total abundance and richness metrics were calculated for aquatic invertebrate identified using ISA as indicator taxa for a single flow class (i.e., perennial or ephemeral) and pairs of adjacent flow classes (i.e., perennial + intermittent + perennial or ephemeral + ephemeral + intermittent). These metrics were calculated for individual regions (e.g., ephinteph_ISAsubregion_taxa) and aggregated regions (e.g., perennial_ISA_taxa; see Appendix B: Indicators and metrics evaluated for SDAM development).

2.6.2 Partition of the dataset into training and testing subsets.

Reaches were split into training (80%) and testing (20%) subsets. Assignment to each subset was stratified by streamflow duration class, as well as geographical subregions or factors in the AW,

WM, and GP (in the East, only streamflow duration class was used to stratify the split into training and testing datasets). The training dataset was used to calibrate models, and the testing dataset was used for final model performance evaluation.

2.6.3 Rebalancing uneven reach visits

Because the number of visits varied among reaches, the training dataset was rebalanced so that all reaches had more equal representation (Chawla et al. 2004, Yap et al. 2014, Buda et al. 2018). First, the target number of visits per reach was determined based on the observed frequency in the training dataset. For reaches with fewer visits than this target, visit data were randomly selected for replication to create an "augmented" or "oversampled" dataset. For example, a reach with two visits might have data from one visit selected at random and duplicated to obtain a target number of 3 visits per reach. For reaches with more visits than this target number, the target number of visit(s) were randomly selected to retain for model calibration; withheld visit data was used to assess some model performance characteristics (e.g., precision), as described below.

Only the training dataset was rebalanced. The testing dataset was not rebalanced.

2.7 Metric screening

The full list of candidate indicators was screened to identify a subset of metrics that would be useful for further analysis. The rebalanced training data set was used for metric screening.

As an initial data exploration step, we visualized the relationships between true streamflow duration class and indicators by ordinating all of the region's candidate metrics for all samples in the dataset in a nonmetric multidimensional scaling using Gowers' distance (Gower 1971). Two-axis ordinations were constructed with the metaMDS function in the vegan package in R (Oksanen et al. 2019, R Core Team 2025). Convex hulls were drawn around each flow class to help visualize their distributions in ordination space. The extent of overlap between the hulls classes can indicate which streamflow duration classes will be more difficult to distinguish.

After the initial data exploration, candidate metrics were evaluated using criteria for inclusion in the final SDAMs (Table 11) characterizing distribution of data and responsiveness:

- Distribution statistic criterion: calculated as percent dominance of the most common value (which was typically zero); all metrics had to meet this criterion.
- Criteria for measuring the responsiveness of metrics (i.e., ability to discriminate across flow classes) included:
    - A set of statistical comparisons of mean values at different subsets of reaches (e.g., t-statistic from a comparison of metric values at perennial and non-perennial reaches), as has been used in other studies (Hawkins et al. 2010, Cao and Hawkins 2011, Mazor et al. 2016).
    - A responsiveness statistic based on metric importance (specifically, mean decrease in accuracy of simple-majority classifications) from a random forest

model to predict flow class from all candidate metrics; the model was calibrated using the the randomForest function in the *randomForest* package in R (Liaw and Wiener 2002), with 1500 trees and all other arguments set to default settings.

The distribution statistic criterion was calculated on the full regional dataset, using data from all reach visits. Responsiveness criteria were evaluated using data from a single randomly selected reach visit for reaches where multiple visits were conducted. Candidate metrics had to meet at least one responsiveness criterion, in addition to the distribution criterion, to be considered in further analyses.

Table 11. Metric screening criteria. Metrics had to meet the distribution criterion and at least one responsiveness criterion to be considered screened for further analysis.

Criterion		Definition
Distribution criterion		
% dominance of most common value	<95%	Frequency of most common value (typically, zero) in the development dataset
Responsiveness criteria		
PvIvE	F>2	F-statistic in a comparison of values at perennial versus intermittent versus ephemeral reaches (PvIvE)
EvALI	t>2	t-statistic in a comparison of values at ephemeral versus at least intermittent reaches (EvALI)
PvLTP	t>2	t-statistic in a comparison of values at perennial versus non-perennial reaches (PvLTP)
PvIwet	t>2	t-statistic in a comparison of values at perennial versus flowing intermittent reaches
EvIdry	t>2	t-statistic in a comparison of values at ephemeral versus dry intermittent reaches
rf_MDA	Top quartile	Mean decrease accuracy (MDA) in a random forest model to predict perennial, intermittent, or ephemeral streamflow duration class (see Figure 4)

2.8 Preliminary model development

Several options for developing Regional SDAMs were explored. First, different stratifications were considered (e.g., developing separate models for different regions or subregions, or developing one model for aggregated regions). Second, we considered models with and without geospatial metrics. For each geographic stratification with or without geospatial metrics, we conducted a similar set of steps to identify an optimal subset of metrics that should be used in the final SDAMs. All steps used the training dataset that included rebalanced data (see 2.6.3 Rebalancing uneven reach visits). Random forest models were trained using the randomForest package in R (Liaw and Wiener 2002) with 1500 trees and all other arguments left at default settings.

We evaluated the performance of these models in both the training and testing datasets and selected a subset of these models for further refinement. Testing datasets, used in performance

assessment, were not rebalanced. A primer on random forest modeling specific to the SDAM development process is below.

Random Forest Modeling and SDAMs

At several steps of the SDAM development process, random forest models were used to associate field-measured indicators with streamflow duration classes. Random forest is a machine learning model that works by utilizing many decision trees, collectively called a "forest," to make a prediction (Breiman 2001, Cutler et al. 2007). Decision trees can be thought of as a flow chart of conditions that can be used to most accurately classify or predict outcomes. The decision trees yield decisions or model predictions in aggregate by utilizing the collective outcomes of all the trees in the forest (Figure 4).

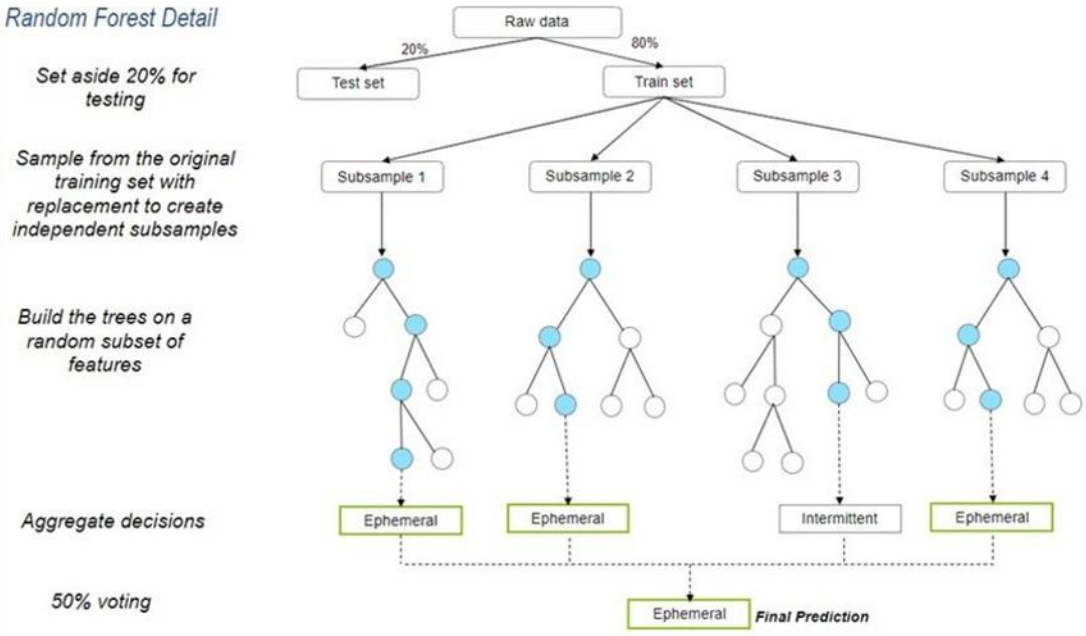


Figure 4. Conceptual illustration of a random forest model showing how a streamflow duration classification may be determined.

In Figure 4, each tree in the random forest represents an independent subsample, comprised of a subset of training visits and subset of candidate metrics selected randomly. By repeatedly sampling data with replacement from the original training dataset, each decision tree was trained on a different random sample of the dataset. Having decision trees that are uncorrelated, or different from one another, reduced the variance of predictions and improved performance. Each tree has nodes, represented as blue circles, that connect to one another through branches (sometimes referred to as descendants). Each split (from node to node) represents an indicator and associated split value. Most nodes have two descendants. For simplicity, the diagram only shows binary, e.g., presence/absence, splits, although some SDAM indicators have multiple discrete values or a continuous spectrum of possible values. The final prediction or classification is based on majority voting from the classifications of all 1500 trees, the "aggregate decisions."

Note, since the 1500 trees in the random forest model are not using the exact same subset of indicators, the relationship between any one indicator value and the resulting classification can be quite complex and dependent on the relationship between indicators represented in each of the 1500 decision trees.

Early in the SDAM development process, random forest models were used to identify which of the many potential metrics under evaluation would be most useful indicators in a final model. This property, known as variable importance, is determined by comparing the difference in classification accuracy of a model containing all metrics to models that exclude one metric; a large decrease in accuracy suggests that the metric has high importance (Figure 5).

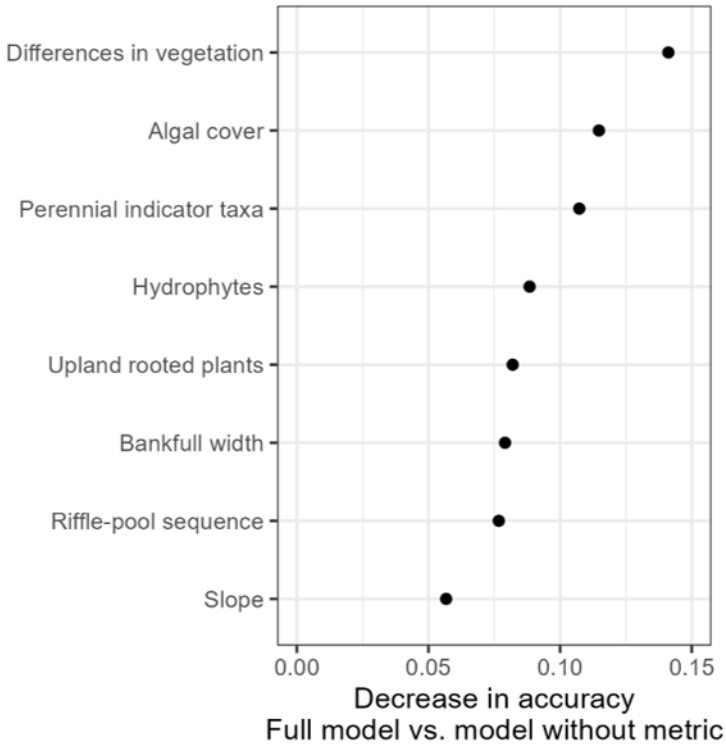


Figure 5. Using a random forest model to evaluate indicator importance. In this example, differences in vegetation and algae metrics have greater importance than slope or riffle-pool sequence because a model that excludes these metrics have much lower accuracy than a model that includes them.

In each final Regional SDAM model, each decision tree is a subset of the geomorphological, hydrological, and biological metrics identified by the modeling as most important for making consistent and accurate classifications of streamflow duration. Each regional model uses a forest of 1500 decision trees based on training dataset to make predictions. When the model is run with indicator data (i.e., field data is entered into the SDAM web application), each tree in the forest votes on the streamflow duration classification, and the relative number of votes provides the basis for predicting a streamflow duration class. If 50% or more of the trees agree on a single class – perennial, intermittent, or ephemeral – then that class is the final prediction. If no single class is 50% or more of the trees and perennial is greater than ephemeral, the outcome is

"at least intermittent." If no single class is 50% or more of the trees and ephemeral is greater than perennial, the outcome is "less than perennial." Finally, if no class is 50% or more, and perennial is equal to ephemeral, the outcome is "needs more information." The adjustment of the decision threshold for increased confidence for the final Regional SDAM classifications resulted in six possible streamflow duration classification outcomes: perennial, intermittent, ephemeral, "at least intermittent," "less than perennial," or "needs more information."

2.8.1 Select optimal geospatial metrics and perennial indicator metrics

Analyses for beta SDAM development showed that, for model development approaches that allowed geospatial metrics, the metric selection process highly favored geospatial metrics over other types of metrics, yielding a final subset that largely excluded other types of metrics (Mazor et al. 2022, Edy et al. 2022, Gross et al. 2023). Because this result produced models that performed poorly with testing datasets and would be unsuitable for our goal of developing a field-based SDAM, we constrained the metric selection process by first creating an initial random forest model with all possible metrics, and then removing all but the two most important geospatial metrics from further consideration. For the AW, WM, and GP – mean decrease in accuracy was used to determine the most important geospatial metrics. For the NE and SE, highly correlated metrics were dropped from similar groupings of geospatial metrics (e.g. precipitation, temperature) before selecting the top metrics via mean decrease in accuracy. This optimal geospatial selection step resulted in a subset of metrics that included no more than two geospatial metrics.

2.8.2 Recursive feature elimination

To select the optimal subset of metrics that remained after the previous steps, we used recursive feature elimination (RFE) using the caret package in R (Kuhn 2020), which is based on variable importance in random forest models (see Figure 4). Briefly, RFE is a form of stepwise selection where complex models (i.e., those based on many metrics) are calibrated and simpler models are considered incrementally by eliminating the least important metrics. Here, the most complex model of each type (stratified or unstratified, with or without geospatial metrics) was first considered. Then for each model, the five least important metrics were eliminated based on their relative performance in the random forest model. This process was iterated until a 20-metric model was identified, after which only one metric was eliminated in each successive step, to a minimum size of a 10-metric model. The best-performing model (i.e., highest accuracy in predicting true streamflow duration class on the training dataset) was identified. Then, the simplest model (i.e., the one with the fewest metrics) with accuracy within 1% of the model with the best accuracy was selected to identify the final set of metrics. If the simplest model selected by this approach had more than 20 metrics, the 20-metric model was selected.

For this analysis, accuracy on the training dataset was measured with Cohen's Kappa statistic—a measure of accuracy that accounts for uneven distribution among the three streamflow duration classes. Note that the Kappa statistic varies from 0 to 1, where 0 equals agreement

equivalent to chance and 1 equates to perfect agreement between the predicted and true classification. Due to the use of random forest models, the Out-of-Bag (OOB) error rate is provided. This means that the prediction error measure for the model is computed through bootstrap or bagging, where subsampling with replacement creates a set of training samples for the model to learn from and the OOB error is the mean prediction error on each training sample for all flow classes.

A final random forest model was recalibrated using the optimal subset of metrics identified by RFE for each stratum, with and without geospatial metrics.

2.8.3 Assignment to streamflow duration outcomes

The process describes how streamflow duration classes are assigned based on the probabilities predicted by a random forest model. Random forest models used for classification calculate the probability that a new observation belongs to each class used to calibrate the model (in this case, perennial, intermittent, or ephemeral). Typically, random forest models assign a final class based on the simple majority, which would be anything greater than 33.3% since three classes were used to calibrate the model. However, greater confidence in this assignment was chosen to avoid making assignments when the probability was less than 50%. Thus, we used model output probabilities to assign streamflow duration classes as shown in Table 12. These assignments serve to communicate the level of confidence with which streamflow duration classifications can be determined by the model.

Table 12. Assignment of streamflow duration classes from model outputs.

Model output	**Classification**
Probability of perennial ≥50%	Perennial (P)
Probability of intermittent ≥ 50%	Intermittent (I)
Probability of ephemeral ≥ 50%	Ephemeral (E)
If no individual class is ≥ 50% see below:	
Probability of perennial > probability of ephemeral	At least intermittent (ALI)
Probability of perennial < probability of ephemeral	Less than perennial (LTP)
Probability of perennial = probability of ephemeral	Need more information (NMI)

2.8.4 Performance assessment of preliminary models and model selection

To select among the different model options (e.g., with vs. without geospatial variables, stratified vs. unstratified models), we compared the performance of the preliminary models in terms of accuracy and precision, and selected a subset for further analysis. To facilitate comparisons across models calibrated for different strata, performance statistics were calculated for the same assessment strata whenever appropriate. For example, to compare the accuracy of a model calibrated for the entire Great Plains with a model calibrated for just the Southern Great Plains, performance statistics of both models were calculated for reaches in the Southern Great Plains by themselves.

2.8.4.1 Accuracy

Accuracy was defined as the percent of correct classifications and was calculated for different subsets of the multi-class classification (Table 13), similar to the comparisons of metric values used in metric screening criteria (Table 11). To select one of the approaches to modeling evaluated by recursive feature elimination (e.g., select a stratified vs. unstratified approach, or to select a model with vs. without geospatial indicators), we sought to maximize three measures of accuracy:

- PvIvE: Accuracy in classifying a reach as perennial, intermittent, or ephemeral.
- EvALI: Accuracy in classifying a reach as ephemeral or at least intermittent
- EvIdry: Accuracy in classifying dry ephemeral or intermittent reaches (perennial and flowing reaches excluded)

These three measures were the focus of model selection because of their importance in management decisions. The first emphasizes the model's overall ability to distinguish among the three classifications. The second focuses on the model's ability to specifically distinguish ephemeral from intermittent or perennial reaches. The third measure only considers the models' performance in distinguishing ephemeral and intermittent reaches when many indicators are expected to be most difficult to characterize or be less discriminating.

Two additional measures of accuracy were calculated, but given less emphasis at this stage of model development and instead were only evaluated for the final methods:

- PvLTP: Accuracy in classifying a reach as perennial or less than perennial.
- PvIwet: Accuracy in classifying flowing intermittent or perennial reaches (ephemeral and dry reaches excluded)

Table 13 provides details about how these five measures of accuracy were calculated.

Table 13. Accuracy measurements. Classifications that are considered correct vary for each measurement. E: Ephemeral. I: Intermittent. P: Perennial. ALI: At least intermittent. LTP: Less than perennial. The classification of need more information (NMI) is considered incorrect for all accuracy measurements.

Accuracy measurement	Description	Correct classifications for E reaches	Correct classifications for I reaches	Correct classifications for P reaches
PvIvE	Percent of reaches correctly classified as perennial, intermittent, or ephemeral. Ambiguous classifications (i.e., ALI, LTP) are incorrect.	E	I	P
EvALI	Percent of reaches correctly classified as ephemeral or at least intermittent. LTP classifications are incorrect.	E	I, P, ALI	I, P, ALI
PvLTP	Percent of reaches correctly classified as perennial or less than perennial. ALI classifications are incorrect.	E, I, LTP	E, I, LTP	P
EvIdry	Percent of dry ephemeral or intermittent reach visits that were correctly classified. ALI or LTP classifications are incorrect.	E	I	Excluded
PvIwet	Percent of flowing intermittent or perennial reach visits that were correctly classified. ALI or LTP classifications are incorrect.	Excluded	I	P

In addition to the accuracy measures in Table 13, we calculated the frequency of two kinds of "big errors" or severe misclassifications:

- EnotP: The number of ephemeral reach visits classified as perennial.
- PnotE: The number of perennial reach visits classified as ephemeral.

2.8.4.2 Precision

Precision was defined as the repeatability of classifications. For each reach with multiple visits, we calculated a consistency score for the percent of visits with the most frequent classification:

$$\frac{(\#visits\ with\ the\ most\ common\ classification) - 1}{(total\ \#visits) - 1}$$

Subtracting the numerator and denominator by one accounts for the differences in the number of visits per reach. Exact matches in classification were considered. For example, if two visits were classified as P and ALI, they would not be considered consistent. No oversampled data were used to calculate consistency scores. Precision was then calculated as the average consistency score across all reaches with multiple visits.

2.9 Model refinement

Once an optimal subset of metrics was identified by RFE, these metrics are referred to as indicators. These indicators were subjected to further manual refinements. Two kinds of model refinements were evaluated: 1) Stepwise modification of the metrics used by the model; and 2) Use of single indicators of at least intermittent status.

2.9.1 Stepwise modification of metrics

The metric selection process described above does not consider the difficulties in measuring each metric, or the overall effort required to measure all metrics. For example, RFE may have selected a metric based on the total abundance of aquatic invertebrates, even if there is little new information gained after the first 10 individuals are documented. Thus, capping this metric at 10 could expedite measurement by allowing practitioners to stop once 10 individuals are documented. Refinement of metrics aims to increase SDAM application efficiency and facilitate method use and transparency. Increasing SDAM application efficiency also helps ensure that an SDAM can be applied during a single reach visit.

At each step, we evaluated PvIvE, EvALI, and EvIdry accuracy as shown in Table 13, as well as precision as described above in Section 2.8.4.2 Precision.

The following tactics were considered during the modification of metrics:

2.9.1.1 Eliminating or substituting metrics

We eliminated redundant metrics, or metrics that were particularly difficult to calculate. In some cases, we substituted related metrics that were easier to measure in the field or required less expertise to consistently measure. For example, a metric based on total richness of aquatic invertebrates requires more taxonomic expertise (and thus is more difficult to measure) than one based on the richness of just a few invertebrate orders (e.g., mayflies, stoneflies and caddisflies).

2.9.1.2 Binning or capping metric values

We explored converting continuous metrics (e.g., invertebrate abundance) into bins or capping them at a maximum value above which all metric values would be treated the same. For ordinal metrics (e.g., scored indicators derived from the New Mexico hydrology protocol, NMED 2011), we explored aggregating scores into fewer ordinal scoring levels. In addition, we explored converting ordinal or continuous metrics into binary forms, such as presence/absence.

2.9.2 Single indicators

Subsequent to the stepwise modifications described in the previous section, single indicators were evaluated. Single indicators are evidence of reaches being at least intermittent, and observation of single indicators may supersede model classifications, as used in some SDAMs (e.g., NMED 2011, Nadeau 2015, Mazor et al. 2022). Single indicators potentially provide technical benefits by improving the overall accuracy of an SDAM, as well as non-technical benefits, such as promoting acceptance of the method, given the public's understanding of the role of streamflow duration in supporting wildlife.

We evaluated several potential single indicators based on their use in other SDAMs, as well recommendations from the RSCs (Table 14).

Table 14. Single indicators evaluated after model refinement and associated descriptions that would indicate reaches were at least intermittent (perennial or intermittent).

Single indicator	Description
SI_Algae_10pct	Algal (live and dead) cover at least 10%
SI_BMI_PA	Presence of any aquatic invertebrates
SI_BMI_10	Presence of at least 10 aquatic invertebrates
SI_BMI_Peren_PA	Presence of any perennial indicator taxa of aquatic invertebrates for aggregate regions, Table 8
SI_BMI_Peren_10	Presence of at least 10 individuals of perennial indicator taxa of aquatic invertebrates for aggregate regions, Table 8
SI_BMI_PerenInt_PA	Presence of any perennial or intermittent indicator taxa of aquatic invertebrates for aggregate regions, Table 8
SI_BMI_PerenInt_10	Presence of at least 10 individuals of perennial or intermittent indicator taxa of aquatic invertebrates only for West aggregate region, Table 8.
SI_BMI_PerenS1_PA	Presence of any perennial or intermittent indicator taxa of aquatic invertebrates only for West aggregate region, Table 8 (excluding Empididae)
SI_BMI_PerenS1_10	Presence of at least 10 individuals of perennial or intermittent indicator taxa of aquatic invertebrates only for West aggregate region, Table 8 (excluding Empididae)
SI_BMI_PerenS2_PA	Presence of any perennial or intermittent indicator taxa of aquatic invertebrates only for West aggregate region, Table 9 (excluding Empididae, Pteronarcyidae, Leptoceridae, and Uenoidae)

Single indicator	Description
SI_BMI_PerenS2_10	Presence of at least 10 individuals of perennial or intermittent ISA indicator taxa of aquatic invertebrates only for West aggregate region, Table 9 (excluding Empididae, Pteronarcyidae, Leptoceridae, and Uenoidae)
SI_BMI_Peren_NC_PA	Presence of any perennial indicator taxa of aquatic invertebrates used in NCDWQ 2010 (East only, Table 9)
SI_BMI_Peren_NC_10	Presence of at least 10 individuals of perennial indicator taxa of aquatic invertebrates used in NCDWQ 2010 (East only, Table 9)
SI_EPT_PA	Presence of any Ephemeroptera, Plecoptera, or Trichoptera taxa
SI_EPT_10	Presence of at least 10 individuals of Ephemeroptera, Plecoptera, or Trichoptera taxa
SI_Fish_NoMosq_PA	Presence of fish other than mosquitofish
SI_HydricSoil_PA	Presence of hydric soils
SI_Hydrophytes_PA	Presence of hydrophytes (FACW or OBL) within the channel or a half-channel width's distance from the channel
SI_Hydrophytes_OBL_PA	[hydrophytes_obl_present] Presence of OBL hydrophytes within the channel or a half-channel width's distance from the channel (East only)
SI_Hydrophytes_Channel_PA	[hydrophytes_present_inchannel] Presence of FACW or OBL hydrophytes within the channel (East only)
SI_Hydrophytes_OBL_Channel_PA	[hydrophytes_obl_present_inchannel] Presence of OBL hydrophytes within the channel (East only)
SI_IOFB_PA	Presence of iron-oxidizing fungi or bacteria
SI_Vertebrates_PA	Presence of aquatic lifestages of fish, amphibians, or reptiles (other than mosquitofish)

To evaluate the impacts of single indicators on performance, we updated classifications so that if a single indicator was observed, any ephemeral, less than perennial, or need more information classifications would be updated to at least intermittent. We then evaluated the number of errors corrected and the numbers of errors introduced for each single indicator, in terms of ephemeral versus at least intermittent accuracy (Table 15).

Table 15. Effect of single indicators (SI) on assessing ephemeral vs. at least intermittent (EvALI) accuracy.

True Class	Predicted Class	EvALI	Predicted class (if SI is present)	EvALI (if SI is present)
E	E	Correct	ALI	Incorrect, error introduced
E	I	Incorrect	I	Incorrect
E	P	Incorrect	P	Incorrect (big error unaffected)
E	ALI	Incorrect	ALI	Incorrect

True Class	Predicted Class	EvALI	Predicted class (if SI is present)	EvALI (if SI is present)
E	LTP	Incorrect	ALI	Incorrect
E	NMI	Incorrect	ALI	Incorrect
I	E	Incorrect	ALI	Correct, error corrected
I	I	Correct	I	Correct
I	P	Correct	P	Correct
I	ALI	Correct	ALI	Correct
I	LTP	Incorrect	ALI	Correct, error corrected
I	NMI	Incorrect	ALI	Correct, error corrected
P	E	Incorrect	ALI	Correct, big error corrected
P	I	Correct	I	Correct
P	P	Correct	P	Correct
P	ALI	Correct	ALI	Correct
P	LTP	Incorrect	ALI	Correct, error corrected
P	NMI	Incorrect	ALI	Correct, error corrected

A correction occurs when a single indicator is present during a reach visit at a reach whose true class is perennial or intermittent, and the predicted class is incorrect for EvALI accuracy measures (i.e., ephemeral, less than perennial, or need more information).

A mistake (i.e., the introduction of an error) occurs when a single indicator is present at a reach whose true class is ephemeral, and the predicted class is ephemeral.

A miss occurs when a single indicator is absent from a reach whose true class is perennial or intermittent, and the predicted class is incorrect for EvALI accuracy measures (i.e., ephemeral, less than perennial, or need more information).

We noted if single indicators were able to fix "big errors" where perennial reaches were misclassified as ephemeral. Because single indicators can only modify classifications of ephemeral or need more information, they have no impact on big errors where ephemeral reaches were misclassified as perennial.

2.10 Performance assessment of the final models

2.10.1 Accuracy at disturbed reach visits

Disturbed conditions (e.g., urban runoff, channel clearing, discharge of effluent, etc.) have the potential to affect the measurement of SDAM indicators. In order to evaluate this potential impact, we compared the accuracy (using measures shown in Table 13) for subsets of reach visits where field crews recorded a disturbance, compared to reach visits where no disturbances were recorded. For reach visits where no disturbance was noted, we presumed the reach visit to be undisturbed. These disturbances focused primarily on the surrounding land use. If the

dominant or secondary land use was noted as urban, agricultural, or developed open space, the reach visit was considered disturbed, and if it was noted as forested or other natural, it was considered undisturbed. If the land use was described as other (but not natural), we classified the reach visit as disturbed or undisturbed based on a review of field notes and aerial imagery. If no land use was noted, we presumed that the reach was undisturbed. Because disturbed sites were included in training and testing data for all methods, the potential impact of disturbance on SDAM performance is not expected to be large.

2.10.2 Sensitivity to variation in indicators

To evaluate the sensitivity of the final models, we re-ran models on the testing dataset after permuting one indicator at a time over a range of possible values. Changes in indicator values were classified as small or large based on best professional judgment. For most binned or scored ordinal variables, a change of one full score or one bin was generally considered small. For continuous measures (such as bankfull width), a change that was between 50% and 150% of the original value was generally considered small (specific thresholds for large vs small changes are presented for each indicator in its appropriate section). Any change to a binary indicator was considered large. We then calculated the percent of reach visits where the outcome changed from the outcome resulting from the original values: If more than 90% of reach visits had unchanged outcomes then the impact was considered small; if between 75% and 90% were unchanged, the impact was considered medium; and if fewer than 75% were unchanged, the impact was considered large. Sensitivity for each indicator was then classified based on the size of the change and the impact, as shown in Figure 6.

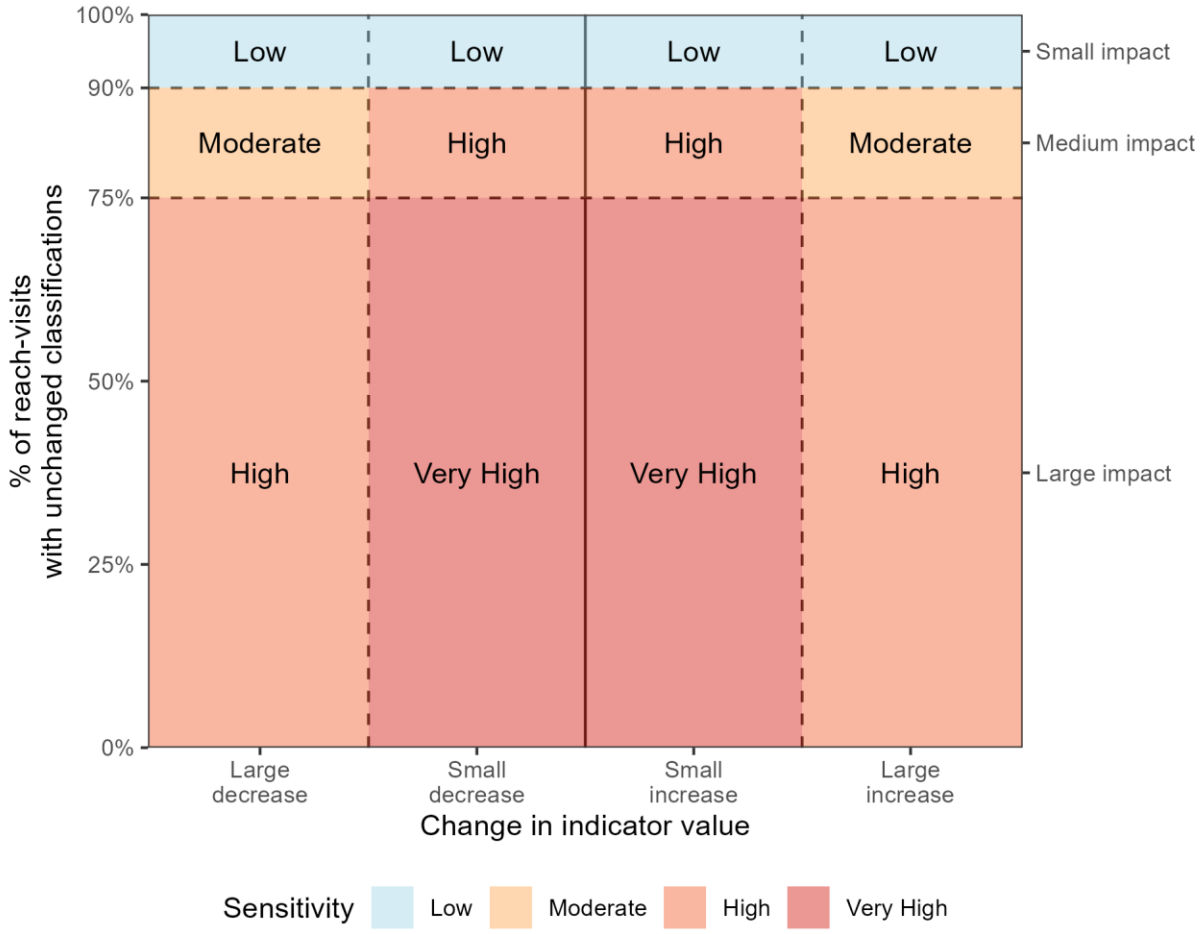


Figure 6. Classification of indicator sensitivity based on the size of change in indicator value and the percent of reach visits with unchanged classifications.

2.10.3 Sensitivity to time of sampling

We compared the accuracy of assessments that occurred in different seasons. Seasons were defined in regionally appropriate ways, as described in each region's section using temperature and precipitation data directly.

We then compared several accuracy statistics for reaches sampled in each season: PvIvE, EvALI, and PvLTP accuracy. Accuracy statistics were calculated separately for training and testing datasets.

We also defined wet and dry seasons and antecedent precipitation conditions (APC) using US Army Corps of Engineers' Antecedent Precipitation Tool (APT) Version 3.0 (Sparrow et al. 2025). Specifically, the APT uses the Web-based, Water-Budget, Interactive, Modelling Program (WebWIMP or WW) to classify dates for locations as occuring in the dry season or wet season based on climatic water balance estimates (0.5º gridded raster). Dates for locations are identified as being in the dry season when the estimated historical (1950-1999) monthly average potential evapotranspiration (PET) is greater than the historical monthly average

precipitation plus snowmelt. Months for locations are identified as being in the wet season when the historical monthly average precipitation plus snowmelt is greater than the historical monthly average PET. Because WW estimates climatic water balance for land, season classifications could not be made for 21 reach visits (at 4 Northeast reaches) located within 0.5º raster cells with mostly water (i.e., near coast).

APC is the relative wetness condition at a particular location based on the antecedent 90-d period (three rolling 30-d periods) to that of the previous 30-y precipitation climatology. Estimated precipitation amounts are scored for each rolling 30-d period to represent Normal (between 30th-70th%), Dry (<30th%), and Wet (>70th%) and given more weight based on the imminency of the rolling 30-d period relative to the reach visit. The APC is the combined score (range 6 – 18) across the three rolling 30-d periods and is assigned an APC class as either Drier than Normal (APC score < 10), Normal (10 ≤ APC score < 15), or Wetter than normal (≥ 15 APC score) classes.

We used single-point APT analyses for the field collected coordinates and the dates for each reach visit to acquire season and APC classes for each reach visit. For most reach visits we were able to use station-based analyses to acquire APC class, but because of limited weather data available for 44 reach visits we needed to use gridded precipitation analyses to acquire APC class (see Sparrow et al. 2025). Accuracy statistics by season and APC were calculated separately for training and testing datasets for each SDAM region (except PNW). Original and augmented site visits were included to calculate accuracy of the training dataset. In addition to Regional SDAM accuracy comparisons between seasons and APCs, we characterized direction of misclassification errors (inaccuracy) across the seasons and APCs. Regional SDAM misclassifications of reach visits were either predicted to be a drier or a wetter class than the actual class (e.g., SDAM predicted intermittent classification for a perennial reach)

The approach used to develop Regional SDAMs attempted to balance the spatial representation within the training datasets (see 2.6.3 Rebalancing uneven reach visits) and testing datasets were composed of randomly selected subsets of the study reach visits. Balancing of the training dataset was done so that study reaches with more visits did not have substantial influence on classification performance than study reaches with fewer visits. Because indicators in all study reaches within a Region could not be simultaneously measured, the timing of visits for data collection varied among study reaches within each Regional dataset and because the SDAM Regions represent large geographic areas, seasonality of weather and associated antecedent weather extremes vary among reaches for a given date. A limitation in the approach used for developing Regional SDAMs was the inability to simultaneously balance reach visits for space and time within the training and testing datasets. The distribution of the reach visits between seasons and APCs is reported for each Regional SDAM and the reader should consider these when interpreting the results characterizing the SDAMs' sensitivity to time of sampling.

2.10.4 Comparison to other SDAMs

We compared the accuracy (using measures shown in Table 13) of the final regional SDAMs to other SDAMs that have been used or could be considered for use in each region. First, we compared the final SDAMs to SDAMs used within the region or in adjacent regions (Table 16). In addition, we compared the final SDAMs to the beta methods developed in each region. Finally, we compared final SDAMs to state-level SDAMs developed within their regions but restricting analyses to training and test reaches within that state.

Table 16. Methods for stream duration classification used within the same region or adjacent regions for each final SDAM.

Final SDAM	**Methods used within the region or in adjacent regions**
AW SDAM	WM SDAM GP SDAM PNW SDAM NM protocol
WM SDAM	AW SDAM GP SDAM PNW SDAM NM protocol
GP SDAM[1]	AW SDAM WM SDAM NM protocol
NE SDAM	GP SDAM SE SDAM NC protocol OH protocol
SE SDAM	NE SDAM GP SDAM NC protocol OH protocol

[1] Comparison to NE or SE SDAMs not possible because not all required indicators were measured at reaches in the GP.

3. Development of the AW and WM SDAMs

3.1 Study areas

The Western US comprises two major regions: the Arid West and the Western Mountains (Figure 7). For the purposes of SDAM development, the regions do not include states in the Pacific Northwest, which are covered by the SDAM described in Nadeau (2015). Throughout this document, the terms “Arid West,” “Western Mountains” and “the West” refer only to the highlighted subregions in Figure 7.

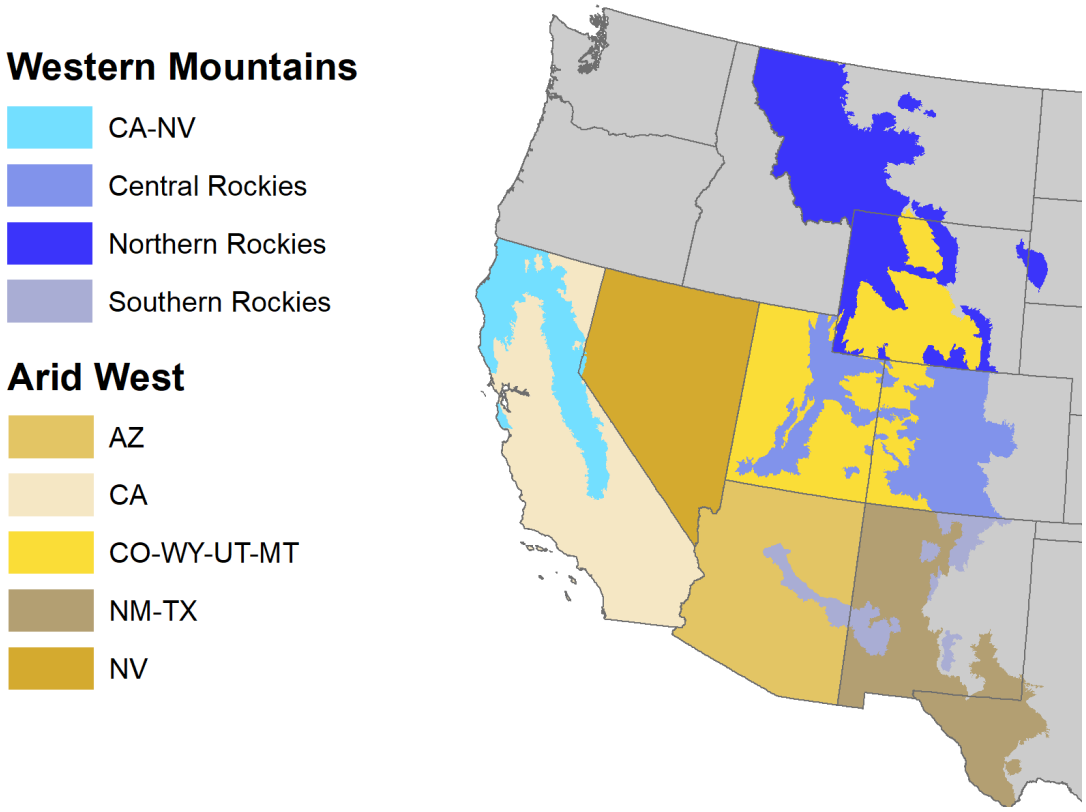


Figure 7. Western regions for SDAM development. Subregions of the WM are shown in shades of blue, and subregions of the AW are shown in shades of yellow.

3.1.1: The Arid West (AW)

The AW encompasses over 1.4 million km^2 in the western United States, covering portions of states from California to Texas and Montana. The region is defined by a combination of variables related to climatic, landcover, vegetation, and soil conditions (U.S. Army Corps of Engineers 2008). The AW includes deserts, grasslands, steppes, shrublands, and woodlands, and is characterized by relatively low rainfall (i.e., <15 inches per year) (U.S. Army Corps of Engineers 2008). In the Mediterranean climate of coastal California, rainfall is restricted to mild and wet winter months; many stream-reaches cease to flow in the dry summer (Gasith and Resh 1999).

In contrast, other portions of the AW are characterized by desert climates, where rainfall is less predictable. The Sonoran Desert is characterized by both a wet and mild winter, as well as summer monsoons, while the Great Basin, Mojave, and Chihuahuan Deserts regularly experience freezing winters. Snowmelt has minimal influence on stream hydrology in the Arid West, except in high-elevation areas (U.S. Army Corps of Engineers 2008). Hydrologic models suggest that the vast majority of stream-miles in the region are non-perennial (Goodrich et al. 2018, Messager et al. 2021), although the relative extent of ephemeral and intermittent reaches is not well documented. Non-perennial reaches dominate the region (Nadeau and Rains 2007), and (in contrast to other regions) often occur in the middle or lower portions of watersheds in addition to headwaters (U.S. Army Corps of Engineers 2008).

Many of the largest metropolitan areas in the USA are located within the AW, and the region continues to experience rapid population growth with expanding urbanization (e.g., Sleeter et al. 2011, Mackun 2019). Thus, the need for an SDAM in permitting and management programs is high in this region. Within the AW, at least two SDAMs are currently in use, applicable to only specific geographic areas: the Pacific Northwest (PNW) method (Nadeau 2015), and the New Mexico (NM) method (New Mexico Environment Department (NMED) 2011). However, prior to the current study, the rest of the region lacked any rapid tool to assess streamflow duration. Our effort focused on the portion of the AW outside the PNW (Figure 7).

3.1.2: The Western Mountains (WM)

The WM encompasses nearly 1 million km^2 in the western United States, covering portions of twelve western states. The region is defined by a combination of variables related to climatic, landcover, vegetation, and soil conditions (U.S. Army Corps of Engineers 2010). The WM includes low-elevation temperate rainforests along the coast that rarely freeze, although much of the region is characterized by high-elevation snow-dominated mountain ranges, including the Sierra Nevada, Rocky Mountains, and Cascades. Typical vegetation is coniferous forests, although higher elevations are characterized by grassland and tundra. Total annual rainfall typically exceeds 20 inches. Ephemeral and intermittent reaches may be found at any position within a watershed but are more common in smaller headwaters, where flow accumulation is insufficient to sustain longer-duration flows.

Although few large cities are found within the WM, several growing metropolitan areas are found near the border the AW or Great Plains, such as Denver, Reno, and Salt Lake City. Thus, the need for an SDAM in permitting and management programs is high in this region. Within the WM, at least two SDAMs are currently in use but are applicable to only specific geographic areas: the Pacific Northwest (PNW) method (Nadeau 2015), and the New Mexico (NM) method (NMED 2011). However, prior to the current study, the rest of the region lacked any rapid tool to classify streamflow duration. Our effort focused on the portion of the WM outside the PNW (Figure 7). This method applies to WM region of the United States as defined in the National Wetland Plant list (U.S. Army Corps of Engineers 2010) excluding the WM region that overlaps with the states of Washington, Oregon, and Idaho covered by the method by Nadeau (2015).

For reaches near regional borders or for reaches in atypical (e.g., arid) conditions within the WM, consult the Western Mountains regional supplement (U.S. Army Corps of Engineers 2010) to determine whether this method is appropriate.

3.2 Data collection

Data were collected at a total of 194 reaches in the AW and 201 reaches in the WM (Table 17, Figure 8). Every state within each study area contained at least 1 reach. Of these, 4 reaches in the AW and 1 reach in the WM could not be confidently assigned to a streamflow duration class and thus were withheld from further data analysis. Of these, 64% of AW reaches and 68% of WM reaches were classified based on a single data source ("Acceptable" in Table 17), whereas multiple data sources were available for classifying the remaining reaches ("Preferred" in Table 17). Direct hydrologic data from gages or loggers deployed as part of this study were available at 89% of AW reaches and 82% of WM reaches; the remaining reaches were classified based on published studies, hydrologic data collected under earlier efforts, and local expertise. Field data collection was incomplete at three single-visit reaches in the AW (an intermittent reach in Nevada, a perennial reach in Nevada, and a perennial reach in California), and these reaches were dropped from further analysis, leaving 187 reaches for method development in the AW and 200 in the WM.

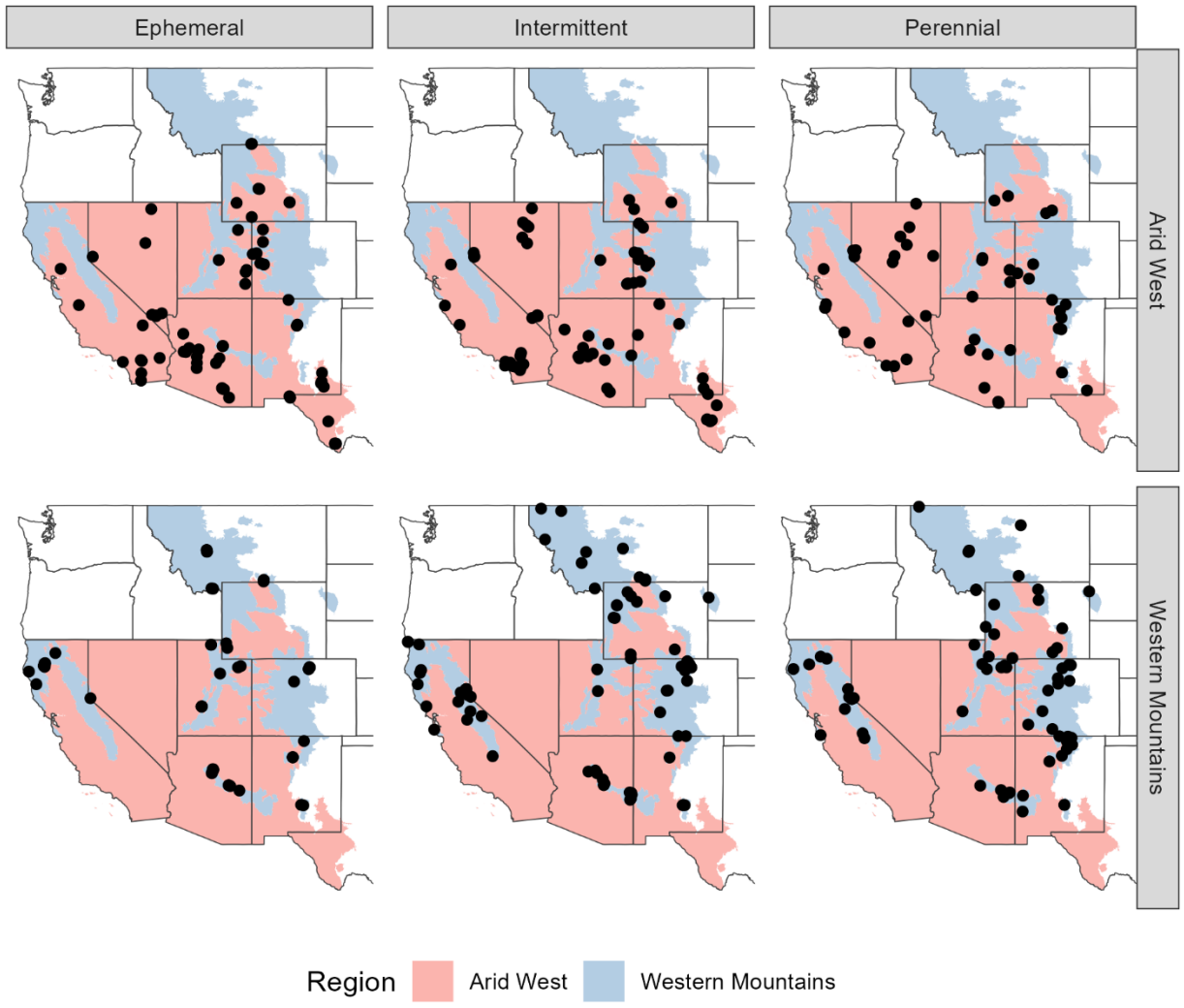


Figure 8. Locations of reaches that were used to develop the final AW and WM SDAMs.

Table 17. Tally of reaches used to develop the AW and WM SDAMs. This table does not include sampled sites that were excluded from analysis: 4 reaches in the AW and 1 reach in the WM where the true streamflow duration class could not be determined, and 3 reaches in the AW where data collection was incomplete.

Region	Class	Sampling stratum	Single visit, acceptable	Single visit, preferred	Repeat visit, acceptable	Repeat visit, preferred	Total acceptable or preferred
AW	E	AZ	4	6	6	1	17
AW	E	CA	3	0	3	3	9
AW	E	CO-UT-WY-MT	6	0	13	2	21
AW	E	NM-TX	2	0	9	3	14
AW	E	NV	1	0	6	0	7
AW	***E***	***Total***	***16***	***6***	***37***	***9***	***68***
AW	I	AZ	1	4	8	0	13
AW	I	CA	3	3	8	3	17
AW	I	CO-UT-WY-MT	4	0	1	12	17
AW	I	NM-TX	2	0	7	2	11
AW	I	NV	2	0	7	4	13
AW	***I***	***Total***	***12***	***7***	***31***	***21***	***71***
AW	P	AZ	0	3	3	1	7
AW	P	CA	2	2	4	2	10
AW	P	CO-UT-WY-MT	2	0	1	10	13
AW	P	NM-TX	5	0	1	2	8
AW	P	NV	3	0	3	4	10
AW	***P***	***Total***	***12***	***5***	***12***	***19***	***48***
WM	E	CA-NV	3	0	7	2	12
WM	E	Central Rockies	5	0	2	9	16
WM	E	Northern Rockies	2	0	6	4	12
WM	E	Southern Rockies	8	0	6	2	16
WM	***E***	***Total***	***18***	***0***	***21***	***17***	***56***
WM	I	CA-NV	6	2	12	2	22
WM	I	Central Rockies	5	0	2	9	16
WM	I	Northern Rockies	11	0	5	6	22

Region	Class	Sampling stratum	Single visit, acceptable	Single visit, preferred	Repeat visit, acceptable	Repeat visit, preferred	Total acceptable or preferred
WM	I	Southern Rockies	1	3	12	2	18
WM	***I***	***Total***	***23***	***5***	***31***	***19***	***78***
WM	P	CA-NV	7	0	3	3	13
WM	P	Central Rockies	9	0	1	7	17
WM	P	Northern Rockies	6	0	6	7	19
WM	P	Southern Rockies	7	0	5	5	17
WM	***P***	***Total***	***29***	***0***	***15***	***22***	***66***

3.3 Metric screening

Ordinations showed that metric values at intermittent and perennial reaches were fairly similar to each other, but substantially different from metric values at ephemeral reaches (Figure 9). This pattern was consistent within the AW and WM regions.

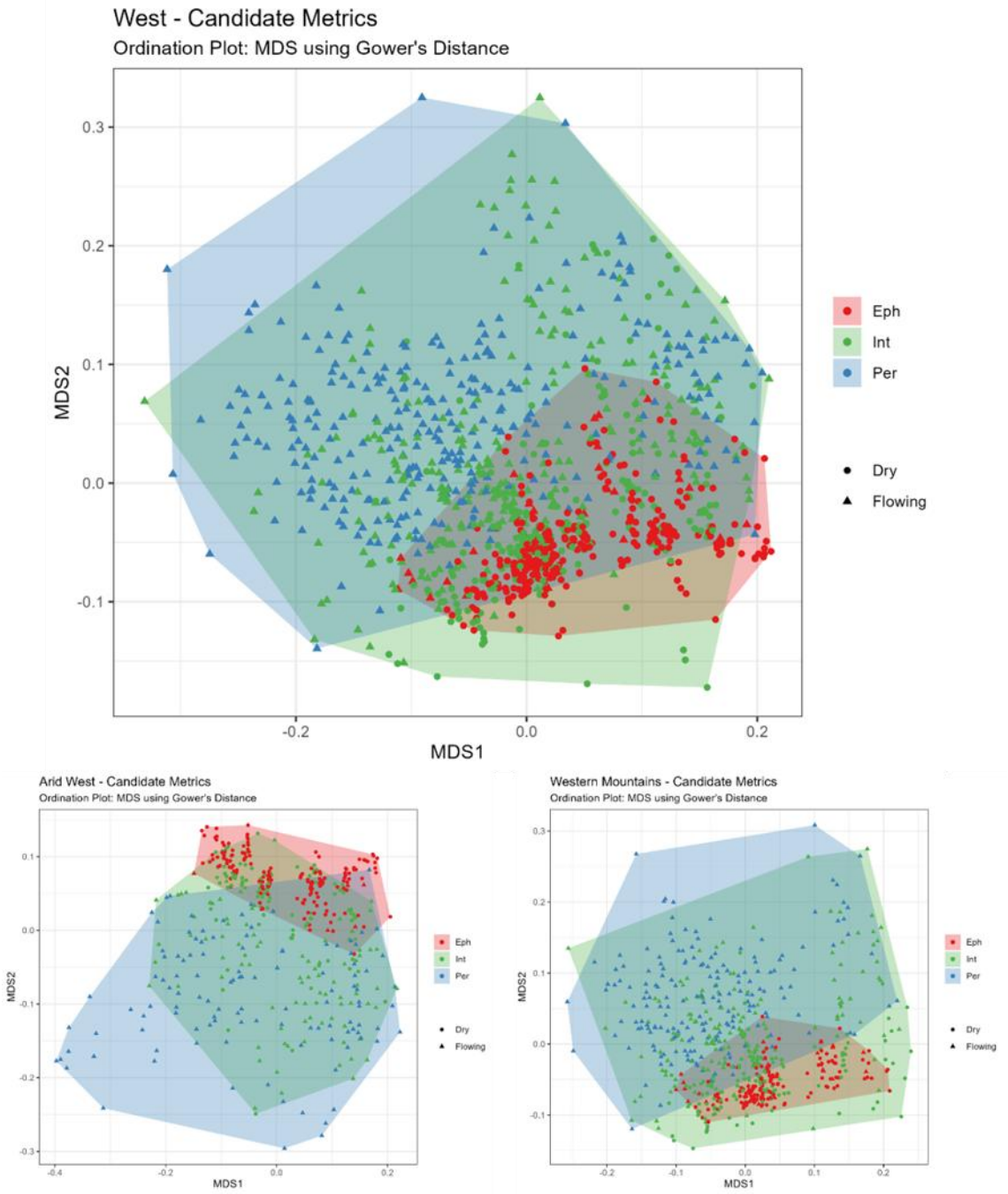


Figure 9. Nonmetric multidimensional scaling (MDS) plot of metric values at perennial (Per), intermittent (Int), and ephemeral (Eph) reaches in the West (top), AW (bottom row, left) and WM (bottom row, right).

Screening only eliminated 4 of the 89 possible metrics (Table 18). One metric (i.e., the abundance of snakes) failed the percent dominance screen. Three additional metrics failed to meet any responsiveness criteria (i.e., precipitation in May through July, precipitation in July, and the number of woody jams in the reach). These metrics were eliminated from further consideration.

Table 18. Screening criteria for metrics in Western regions (AW + WM). Metrics are described in Appendix B.

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Algae	AlgalCover_Live_NoUpstream	0.58	0	4	156.4	18.8	13.3	1.8	3.3	0.012	Pass
Biological	Algae	AlgalCover_LiveOrDead_NoUpstream	0.52	0	4	157.7	19.8	12.6	2.1	6.0	0.021	Pass
Biological	Algae	ironox_bfscore_NM	0.90	0	1.5	16.3	7.7	2.5	0.8	3.4	0.000	Pass
Biological	Aquatic invertebrates	BMI_presence	0.64	0	1	266.6	20.4	23.3	2.1	5.6	0.011	Pass
Biological	Aquatic invertebrates	Ephemeroptera_abundance	0.59	0	62	265.4	18.4	16.5	7.8	2.6	0.019	Pass
Biological	Aquatic invertebrates	Ephemeroptera_taxa	0.59	0	7	279.4	19.9	17.3	7.4	3.3	0.016	Pass
Biological	Aquatic invertebrates	EPT_abundance	0.50	0	226	224.5	16.5	15.0	7.7	2.8	0.016	Pass
Biological	Aquatic invertebrates	EPT_taxa	0.50	0	20	244.4	17.8	15.8	7.5	3.5	0.009	Pass
Biological	Aquatic invertebrates	GOLD_abundance	0.45	0	91	144.9	18.4	11.7	2.8	4.6	0.006	Pass
Biological	Aquatic invertebrates	Noninsect_abundance	0.54	0	85	93.2	13.8	9.2	4.1	4.5	0.004	Pass
Biological	Aquatic invertebrates	OCH_abundance	0.57	0	66	78.2	12.6	8.6	3.6	3.1	0.003	Pass
Biological	Aquatic invertebrates	OCH_taxa	0.57	0	9	72.1	13.9	8.0	0.9	4.0	0.002	Pass
Biological	Aquatic invertebrates	perennial_ISA_abundance	0.63	0	176	183.1	13.4	12.9	8.5	2.1	0.020	Pass
Biological	Aquatic invertebrates	perennial_ISA_taxa	0.63	0	13	227.2	15.5	14.7	8.8	2.8	0.008	Pass
Biological	Aquatic invertebrates	perennial_ISAsubregion_abundance	0.54	0	266	200.4	15.7	13.7	7.7	3.2	0.017	Pass
Biological	Aquatic invertebrates	perennial_ISAsubregion_taxa	0.54	0	24	224.0	17.5	14.9	7.3	3.9	0.012	Pass
Biological	Aquatic invertebrates	perennial_PNW_abundance	0.68	0	84	153.2	12.6	12.0	7.8	2.4	0.011	Pass
Biological	Aquatic invertebrates	perennial_PNW_live_abundance	0.68	0	84	154.2	12.5	12.0	7.8	2.2	0.011	Pass
Biological	Aquatic invertebrates	perennial_PNW_taxa	0.68	0	10	189.5	15.2	14.0	7.4	3.5	0.007	Pass
Biological	Aquatic invertebrates	perintper_ISA_abundance	0.51	0	250	257.2	17.2	15.4	8.8	3.6	0.024	Pass
Biological	Aquatic invertebrates	perintper_ISA_taxa	0.51	0	23	293.9	19.0	16.9	8.7	4.4	0.016	Pass
Biological	Aquatic invertebrates	perintper_ISAsubregion_abundance	0.54	0	266	200.4	15.7	13.7	7.7	3.2	0.018	Pass
Biological	Aquatic invertebrates	perintper_ISAsubregion_taxa	0.54	0	24	224.0	17.5	14.9	7.3	3.9	0.012	Pass
Biological	Aquatic invertebrates	Plecoptera_abundance	0.73	0	48	70.7	10.5	8.5	3.7	1.1	0.002	Pass
Biological	Aquatic invertebrates	Plecoptera_taxa	0.73	0	7	110.3	12.2	10.5	5.0	1.1	0.001	Pass
Biological	Aquatic invertebrates	Richness	0.40	0	38	274.8	21.5	17.2	6.2	5.0	0.011	Pass
Biological	Aquatic invertebrates	TotalAbundance	0.40	0	366	245.0	19.4	15.5	6.9	4.6	0.014	Pass
Biological	Aquatic invertebrates	Trichoptera_abundance	0.58	0	116	143.4	12.9	11.8	7.1	2.4	0.006	Pass
Biological	Aquatic invertebrates	Trichoptera_taxa	0.58	0	11	173.7	15.1	13.5	6.7	3.0	0.003	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Vegetation	DifferencesInVegetation_score	0.27	0	3	301.8	23.0	18.2	4.7	9.7	0.040	Pass
Biological	Vegetation	hydrophytes_present	0.31	0	13	154.2	18.5	12.0	4.5	6.7	0.030	Pass
Biological	Vegetation	Liverwort_cover	0.93	0	3	4.3	3.7	1.4	0.0	1.7	0.000	Pass
Biological	Vegetation	Moss_cover	0.84	0	3	5.9	3.7	2.1	0.2	1.7	0.000	Pass
Biological	Vegetation	PctShading	0.21	0	100	17.2	5.4	0.1	2.6	4.4	0.006	Pass
Biological	Vegetation	UplandRootedPlants_score	0.45	0	3	132.4	13.5	14.5	2.8	4.6	0.005	Pass
Biological	Vertebrates	Amphibian_presence	0.95	0	1	6.7	3.5	0.7	2.9	1.1	0.000	Pass
Biological	Vertebrates	AmphSnake_PA	0.94	0	1	6.0	3.9	0.3	2.2	1.1	0.000	Pass
Biological	Vertebrates	Fish_PA	0.81	0	1	112.7	13.9	10.9	5.0	2.9	0.004	Pass
Biological	Vertebrates	snake_score	0.98	0	1	1.6	1.3	1.4	0.4	1.0	0.000	Fail
Geomorph	Geomorph	BankWidthMean	0.01	0.13	149.0	4.5	1.0	3.5	3.0	0.3	0.016	Pass
Geomorph	Geomorph	ChannelDimensions_score_NM	0.56	0	3	3.1	2.2	0.4	0.3	2.2	0.001	Pass
Geomorph	Geomorph	RiffflePoolSeq_score	0.33	0	3	112.3	12.3	13.9	5.1	4.0	0.004	Pass
Geomorph	Geomorph	SedimentOnPlantsDebris_score	0.53	0	1.5	18.8	5.1	5.0	2.5	1.8	0.001	Pass
Geomorph	Geomorph	Sinuosity_score	0.56	0	3	9.2	4.2	2.7	0.1	1.7	0.001	Pass
Geomorph	Geomorph	Slope	0.19	0	57	21.2	5.1	5.5	1.8	4.3	0.004	Pass
Geomorph	Geomorph	SubstrateSorting_score	0.41	0	3	77.8	10.3	11.5	5.1	4.8	0.003	Pass
Geospatial	Snow persistence	MeanSnowPersistence_01	0.01	0.0	85.1	5.3	2.5	3.0	1.5	0.7	0.009	Pass
Geospatial	Snow persistence	MeanSnowPersistence_05	0.01	0.0	86.6	7.8	2.9	3.6	2.0	0.9	0.008	Pass
Geospatial	Snow persistence	MeanSnowPersistence_10	0.01	0.1	82.6	8.6	2.7	3.9	2.4	0.7	0.008	Pass
Geospatial	Precipitation	ppt	0.02	84.5	1867.4	11.1	3.8	1.3	3.0	2.9	0.008	Pass
Geospatial	Precipitation	ppt.11121	0.02	6.4	312.0	8.8	2.9	1.9	3.3	2.1	0.011	Pass
Geospatial	Precipitation	ppt.234	0.02	8.7	240.7	9.9	3.8	1.0	3.0	2.5	0.007	Pass
Geospatial	Precipitation	ppt.567	0.02	1.3	102.9	0.7	1.2	0.6	0.6	1.3	0.011	Fail
Geospatial	Precipitation	ppt.8910	0.02	4.7	89.8	4.2	2.7	0.1	0.7	2.8	0.010	Pass
Geospatial	Precipitation	ppt.m01	0.02	5.7	344.9	9.2	2.9	2.0	3.5	2.2	0.010	Pass
Geospatial	Precipitation	ppt.m02	0.02	5.4	300.5	11.4	3.7	1.6	3.5	2.5	0.013	Pass
Geospatial	Precipitation	ppt.m03	0.02	6.9	266.8	9.1	3.6	1.0	2.8	2.5	0.009	Pass
Geospatial	Precipitation	ppt.m04	0.02	1.3	154.7	5.4	3.3	0.2	1.5	2.0	0.011	Pass
Geospatial	Precipitation	ppt.m05	0.02	0.7	116.8	0.7	1.0	1.1	0.6	0.5	0.014	Pass
Geospatial	Precipitation	ppt.m06	0.02	0.3	132.9	0.3	0.4	0.8	0.9	0.5	0.011	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Geospatial	Precipitation	ppt.m07	0.02	0.1	124.6	1.2	1.3	0.3	0.1	1.9	0.009	Fail
Geospatial	Precipitation	ppt.m08	0.02	0.6	133.8	1.5	1.1	0.8	0.4	2.0	0.008	Pass
Geospatial	Precipitation	ppt.m09	0.02	3.5	85.3	2.0	2.1	0.6	0.0	1.9	0.010	Pass
Geospatial	Precipitation	ppt.m10	0.02	3.9	116.2	8.7	4.3	0.5	1.4	2.9	0.009	Pass
Geospatial	Precipitation	ppt.m11	0.02	4.3	257.9	9.0	3.4	1.4	3.0	2.3	0.011	Pass
Geospatial	Precipitation	ppt.m12	0.02	6.3	391.7	7.8	2.5	2.1	3.3	1.9	0.011	Pass
Geospatial	Temperature	temp.11121	0.02	-8.5	15.2	3.7	0.3	2.6	2.7	0.1	0.006	Pass
Geospatial	Temperature	temp.234	0.02	-5.2	18.6	4.1	1.8	2.8	2.2	1.0	0.006	Pass
Geospatial	Temperature	temp.567	0.02	7.1	30.6	9.2	4.0	3.1	1.1	2.7	0.007	Pass
Geospatial	Temperature	temp.8910	0.02	6.0	29.2	5.8	3.0	2.8	1.6	2.0	0.006	Pass
Geospatial	Temperature	temp.m01	0.02	-10.8	14.2	3.7	0.2	2.5	2.7	0.2	0.006	Pass
Geospatial	Temperature	temp.m02	0.02	-9.1	15.4	3.5	0.7	2.6	2.5	0.2	0.005	Pass
Geospatial	Temperature	temp.m03	0.02	-5.3	18.3	4.1	1.9	2.8	2.1	1.0	0.006	Pass
Geospatial	Temperature	temp.m04	0.02	-1.9	22.0	5.9	2.8	3.0	1.7	1.9	0.005	Pass
Geospatial	Temperature	temp.m05	0.02	2.7	26.5	7.8	3.5	3.1	1.4	2.3	0.006	Pass
Geospatial	Temperature	temp.m06	0.02	7.5	31.0	9.0	3.9	3.1	1.1	2.6	0.007	Pass
Geospatial	Temperature	temp.m07	0.02	11.1	34.2	10.9	4.5	3.0	0.9	3.1	0.008	Pass
Geospatial	Temperature	temp.m08	0.02	10.3	33.7	9.4	4.2	2.9	1.0	3.0	0.009	Pass
Geospatial	Temperature	temp.m09	0.02	6.5	30.1	5.9	3.0	2.8	1.6	2.0	0.007	Pass
Geospatial	Temperature	temp.m10	0.02	1.1	24.5	3.7	1.8	2.6	2.0	1.1	0.005	Pass
Geospatial	Temperature	temp.m11	0.02	-5.4	17.9	3.5	0.8	2.6	2.4	0.2	0.005	Pass
Geospatial	Temperature	temp.m12	0.02	-10.6	14.2	4.1	0.1	2.6	2.9	0.4	0.005	Pass
Geospatial	Temperature	tmax	0.02	6.3	31.1	4.5	2.3	2.7	1.8	1.4	0.005	Pass
Geospatial	Temperature	tmean	0.02	-0.1	23.2	4.7	2.2	2.9	2.0	1.3	0.006	Pass
Geospatial	Temperature	tmin	0.02	-7.0	15.3	4.7	2.1	2.9	2.2	1.2	0.007	Pass
Geospatial	Watershed	DRNAREA_mi2	0.03	0.01	2030	27.2	2.9	5.3	5.0	1.2	0.018	Pass
Geospatial	Watershed	Elev_m	0.02	8	3260	6.1	3.4	2.6	1.5	2.4	0.009	Pass
Hydro	Indirect	HydricSoils_score	0.65	0	3	113.1	14.2	12.0	3.1	3.0	0.004	Pass
Hydro	Indirect	WoodyJams_number	0.84	0	100	0.6	1.3	0.1	1.4	1.0	0.000	Fail

3.4 Data stratification and preparation

Reaches within the West, which here and throughout refers to the portions of the AW and WM included in this study and excludes the PNW (Figure 7), were split into training (80%) and testing (20%) subsets within each region (AW vs. WM) and streamflow duration class. Within the WM, reaches were additionally stratified based on snow influence. Snow influence has previously been shown to be an important driver of streamflow duration in the West (Hammond et al. 2018, 2021), and was used as a stratification variable in the beta SDAM WM (Mazor et al. 2022). Snow influence was measured as the mean snow persistence within a 10-km radius of the assessment reach, using Hammond et al. (2017) as a source of snow persistence data for the West. Snow persistence is the fraction of time that snow is present on the ground between January 1 and July 3, averaged across the years between 2000 and 2020. Assessment reaches where the mean snow persistence is greater than 25% were classified as snow influenced, as this threshold differentiates areas where snow is minimal from areas where snow is intermittent, transitional, or persistent (Hammond et al. 2018). Snow influenced areas are prevalent in the Rocky Mountains, as well as at higher elevations in Arizona and the Sierra Nevada of California. Areas that are not snow-influenced are prevalent in the coastal mountains and valleys of northern California, the Sierra Nevada Foothills, and the mountains of southern New Mexico, but they are also found throughout other regions of the WM (Figure 10).

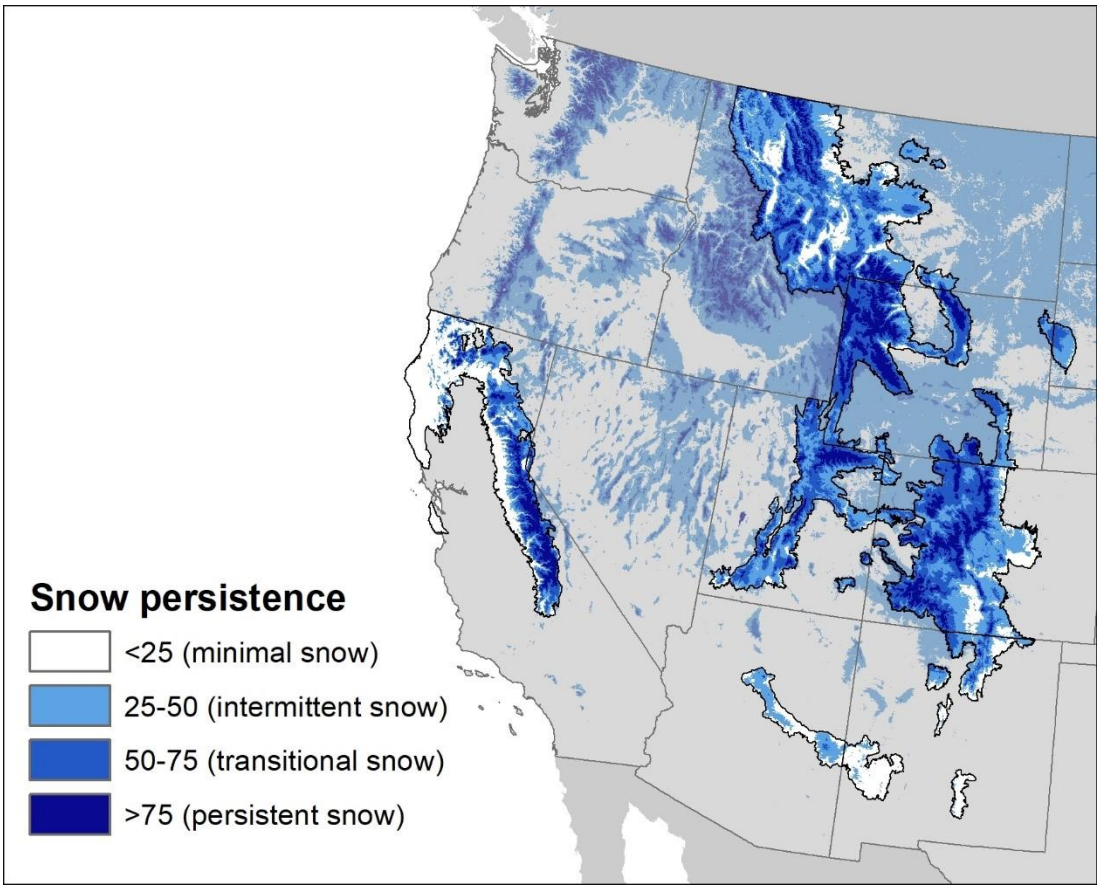


Figure 10. Average snow persistence in the western United States, highlighting the WM region. Data accessed from Hammond et al. (2017). Snow influenced areas are defined as those with mean snow persistence greater than 25 (i.e., on average, snow is on the ground more than 25% of the time between January 1 and July 3). Portions of the west outside the WM region are presented with a gray overlay.

Within the WM, 137 reaches were classified as snow influenced, and 64 as not snow influenced. Although snow influence affects some parts of the AW (i.e., 51 out of 194 reaches), we did not explore this as a stratification approach within the AW following preliminary analyses. The distribution of reaches in testing and training datasets is shown in Table 19.

Table 19. Number of reaches and reach visits in training and testing datasets for each region, streamflow duration class, and snow influence category. Snow influence was used to stratify reaches within the WM but not the AW.

Region	Class	Snow influence	Training Reaches	Training Visits	Testing Reaches	Testing Visits
AW	E	Not snow influenced	39	88	11	29
AW	E	Snow influenced	14	38	4	10
AW	I	Not snow influenced	45	116	11	25
AW	I	Snow influenced	10	32	5	13

Region	Class	Snow influence	Training Reaches	Training Visits	Testing Reaches	Testing Visits
AW	P	Not snow influenced	25	59	6	13
AW	P	Snow influenced	13	33	4	11
WM	E	Not snow influenced	21	74	5	21
WM	E	Snow influenced	25	65	5	13
WM	I	Not snow influenced	20	102	2	3
WM	I	Snow influenced	44	126	12	48
WM	P	Not snow influenced	13	35	3	5
WM	P	Snow influenced	39	112	11	33

The number of visits per reach varied considerably across sites in both the AW and WM (Figure 11), underscoring the need for rebalancing the dataset. The goal of rebalancing is to ensure that each reach used to train the model has equal representation. Thus, reaches with many visits would be down sampled (that is, a random subset of visits would be retained for further analysis and the other visits would be excluded), and reaches with few visits would be oversampled (that is, data from these visits would be replicated), resulting in a training dataset where each reach is represented by an equal number of reaches (see Section 2.6.3 Rebalancing uneven reach visits).

Based on a review of the frequency of revisits by region, a target of 3 visits per reach was selected as a target number for rebalancing the training dataset (Figure 11). Thus, the following steps were taken to rebalance the training dataset:

- If a reach had one visit, it was repeated three times
- If a reach had two visits, one visit was selected at random and replicated
- If a reach had three visits, it was left as-is.
- If a reach had four or more visits, three visits were selected at random and retained in the dataset.

This rebalancing resulted in a training dataset that included 438 samples within the AW, of which 113 were oversampled (i.e., replicates of other reach visits), and 486 samples in the WM, of which 131 were oversampled.

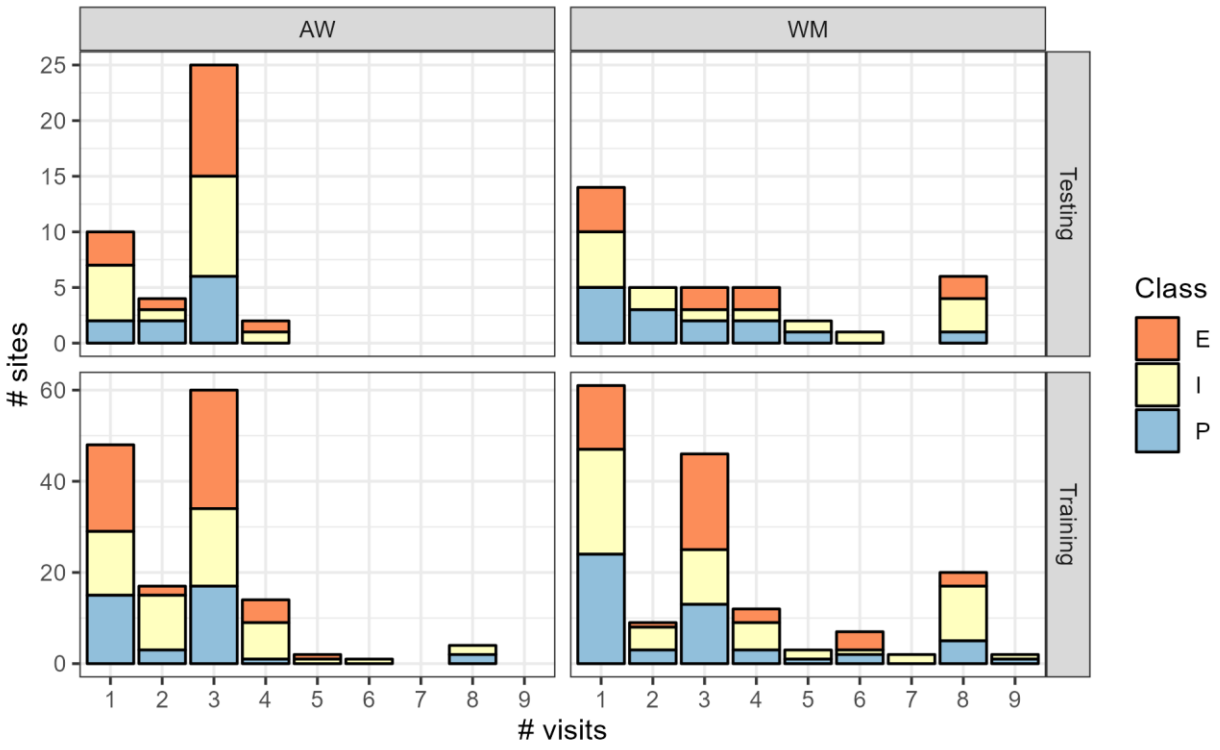


Figure 11. Histograms showing the frequency of reaches by number of visits per reach prior to rebalancing.

3.5 Model development

Similar to the problems noted above concerning geospatial metrics (Section 2.8.1 Select optimal geospatial metrics and perennial indicator metrics), we observed that recursive feature elimination tended to select numerous closely related metrics, to the exclusion of other metrics. Thus, we implemented a similar solution: we trained a random forest model consisting of all possible metrics (including the two previously identified geospatial metrics, if applicable), and then identified the two most important perennial indicator metrics to retain for further consideration.

We explored several different models for developing SDAMs for the West. Specifically, we explored models that either included or excluded geospatial variables, and we explored three stratifications:

- One West model that combined the AW and WM into an aggregated region
- Two regional models, one for the AW and one for the WM
- Within the WM, one model for each stratum based on snow influence. (No stratification within the AW was evaluated)

Thus, there were five possible strata, each requiring its own model: West, AW, WM, WM-snow influenced, and WM-not snow influenced. For each stratum, two models were developed—one

with, and one without geospatial metrics. Accordingly, 10 models were developed for the West. Table 20 summarizes the number of reaches and visits used to calibrate each model.

Table 20. Number of reaches and visits used to train or test each model.

	Training # reaches			Testing # reaches			Training # visits			Testing # visits		
Model	**P**	**I**	**E**	**P**	**I**	**E**	**P**	**I**	**E**	**P**	**I**	**E**
West	90	119	99	24	30	25	239	376	265	62	89	73
AW	38	55	53	10	16	15	92	148	126	24	38	39
WM	52	64	46	14	14	10	147	228	139	38	51	34
WM-Snow influenced	39	44	25	11	12	5	112	126	65	33	48	13
WM-Not snow influenced	13	20	21	3	2	5	35	102	74	5	3	21

Note: Three visits to one ephemeral site in Utah (i.e., UTWM1363) was erroneously assigned to the AW, and was included in the training dataset for the AW model; one visit to this site was included the testing dataset for the WM model. Thus, one site is counted twice in Table 20. This site was identified as snow-influenced.

Recursive feature elimination resulted in models requiring from 10 to 20 predictors, and Cohen's Kappa statistics ranging from 0.77 to 0.88 (Table 21). Of the 85 candidate metrics that passed screens, 36 were selected for at least one of the 10 models (Figure 12). Three were selected for all 10 models: bankfull width mean, percent shading, and differences in vegetation. Among the geospatial metrics, drainage area was the most frequently selected metric. Metrics based on perennial indicator taxa of aquatic invertebrates were important for all models that excluded geospatial metrics; however, they were less frequently selected for the geospatial models.

Table 21. Number of predictors selected and Cohen's kappa statistic by recursive feature elimination

Model	**# Predictors**	**Cohen's Kappa**
West	20	0.80
West GIS	11	0.83
AW	10	0.81
AW GIS	11	0.85
WM	12	0.77
WM GIS	20	0.85
WM-Snow influenced	17	0.83
WM-Snow influenced GIS	10	0.88
WM-Not snow influenced	11	0.79
WM-Not snow influenced GIS	13	0.85



Figure 12. Metrics selected by Recursive Feature Elimination for each model in the West.

For these western models, we evaluated performance statistics for five assessment strata based on the same strata used to define modeling approaches i.e., (West, AW, WM, WM-snow influenced, and WM-not snow influenced). We calculated performance statistics for every assessment stratum that was geographically within the range of the calibration datasets. For example, the West model was assessed for all assessment strata, but the WM model was only assessed for three assessment strata (i.e., WM, WM-snow influenced, and WM-not snow influenced).

Based on a comparison of the performance of the different models, the AW and WM models without geospatial metrics were selected for further analysis (Figure 13, Figure 14). Models with

geospatial metrics had comparable or somewhat better performance than their counterparts for training data yet often showed substantially worse performance in testing data.

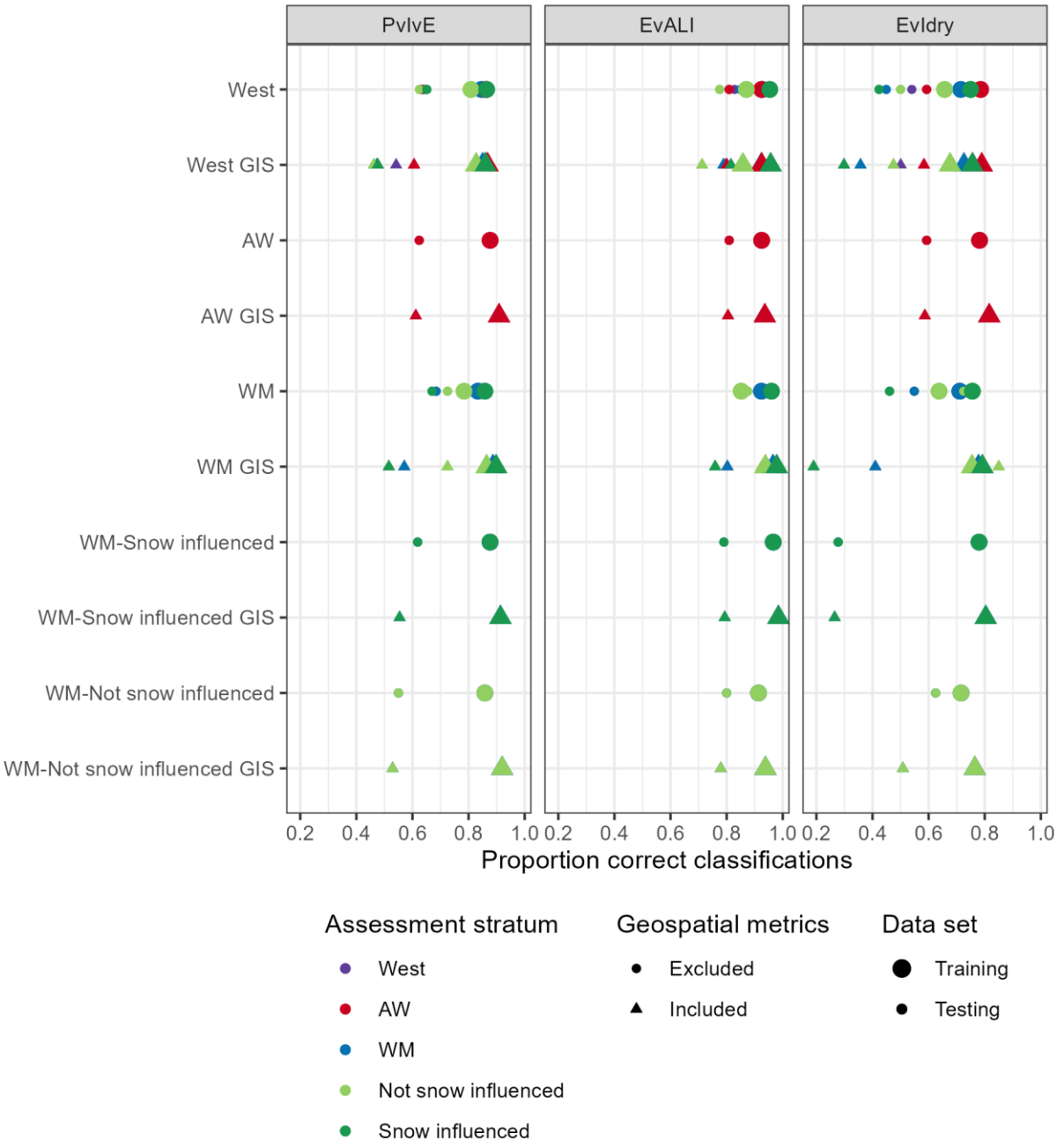


Figure 13. Accuracy of the preliminary models.

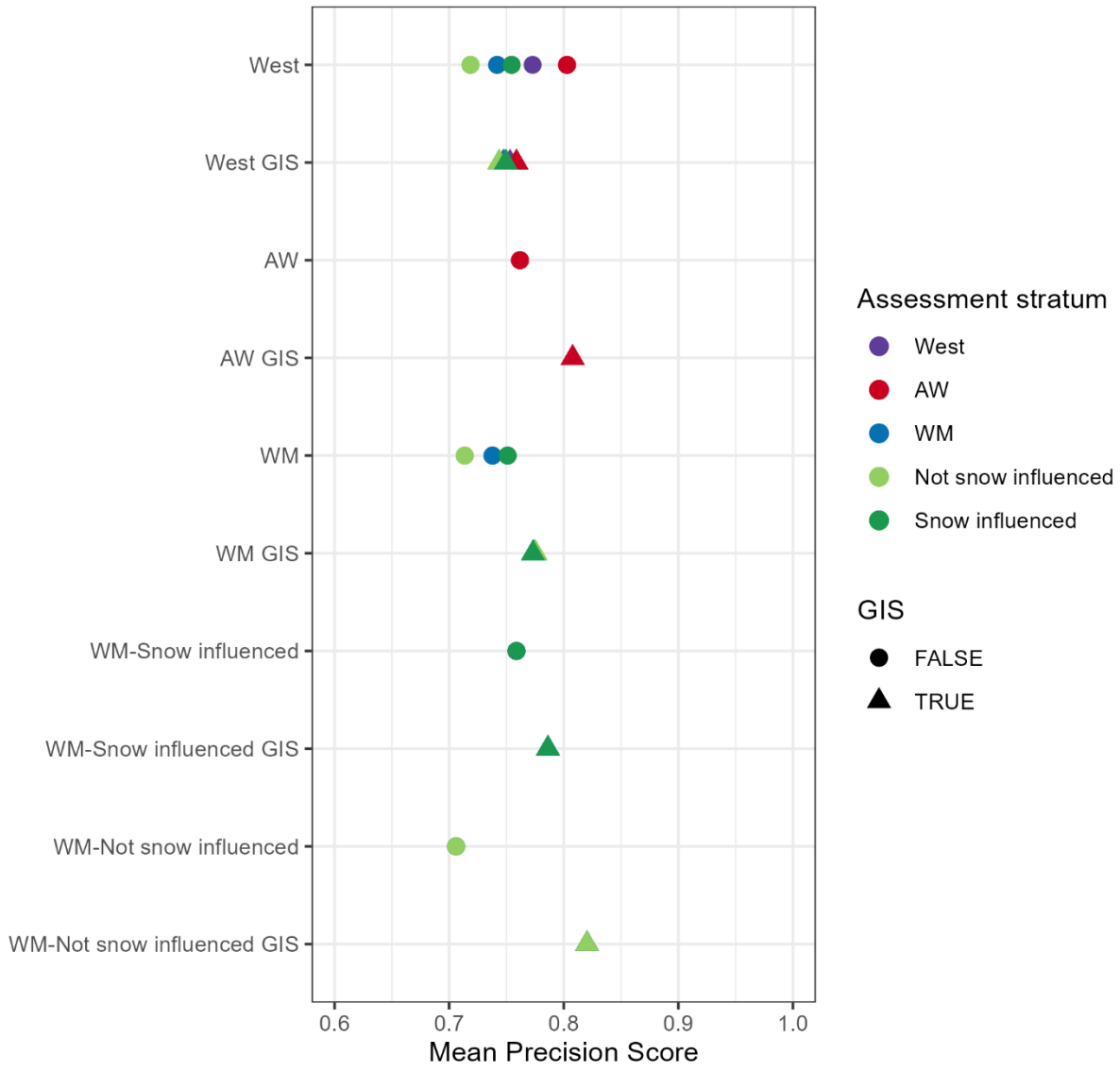


Figure 14. Precision of preliminary models.

Based on the performance of the preliminary models, the AW and WM models without geospatial metrics were selected for further refinement (Table 22). These models had relatively high accuracy for all measures (Figure 13). One “big error” was noted in the testing dataset for the WM, and none in the AW.

Table 22. Accuracy of the preliminary AW and WM models. Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral), and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures, but are not considered "big errors". Note Training data includes augmented data.

Arid West: Training

Predicted class	**Ephemeral**	**Intermittent**	**Perennial**
Ephemeral	146	19	0
Intermittent	12	134	10
Perennial	0	8	104
At least intermittent	0	3	0
Less than perennial	1	1	0
Need more information	0	0	0

Arid West: Testing

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	32	11	0
Intermittent	6	18	6
Perennial	0	6	17
At least intermittent	1	2	1
Less than perennial	0	1	0
Need more information	0	0	0

Western Mountains: Training

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	119	15	0
Intermittent	15	155	22
Perennial	1	17	131
At least intermittent	0	2	3
Less than perennial	3	3	0
Need more information	0	0	0

Western Mountains: Testing

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	20	9	1
Intermittent	12	25	9
Perennial	0	13	27
At least intermittent	0	2	1
Less than perennial	2	2	0
Need more information	0	0	0

The two selected preliminary models shared many metrics in common (Figure 15). For both models, the most important metric was the differences in vegetation score, and the two perennial indicator taxa metrics were nearly as important. However, algal cover was one of the top metrics for the AW model yet not selected at all for the WM model. Conversely, metrics based on Ephemeroptera or EPT abundance were important for the WM model but not selected for the AW model. Bankfull channel width was among the most important variables in the WM model, whereas it was relatively less important for the AW model (Figure 15).

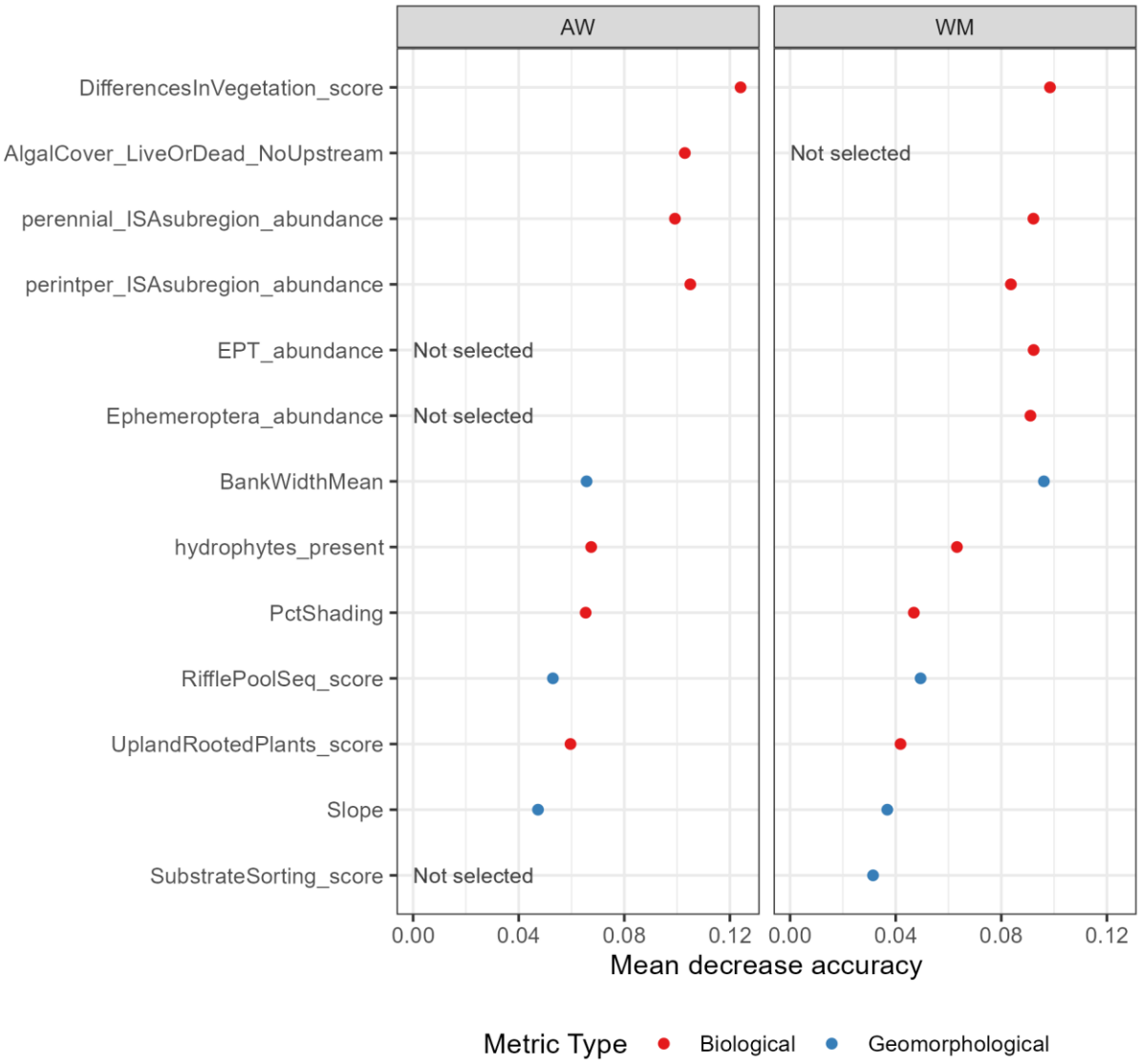


Figure 15. Importance as mean decrease accuracy of metrics selected for the preliminary AW and WM models.

3.6 Model refinement

3.6.1 AW model refinement

Each step of model refinement is described in Table 23. After each step, we assessed the performance in terms of accuracy (measured as PvIvE, EvALI, and EvIdry) in both training and testing datasets (Table 24, Figure 16). First, we sought to drop potentially redundant or complicated metrics, possibly replacing them with simpler or more comprehensive alternatives. For example, for the AW model, two closely related aquatic invertebrate metrics were selected: one based on perennial indicator taxa, and another based on perennial plus intermittent indicator taxa. We explored the impact of dropping each one separately (steps 1 and 2), and seeing that they had fairly comparable performance, we selected the one based on perennial indicator taxa alone. This metric is based on a smaller number of taxa (9 vs 11 taxa) and thus would be less challenging for most practitioners to measure. We explored dropping metrics that required special equipment (i.e., a clinometer or densiometer, steps 3 and 5), as well as other metrics with relatively low importance (e.g., the riffle-pool sequence score, step 4). After reducing the number of variables, we explored ways to bin hydrophyte and aquatic invertebrate metrics (steps 7 and 8) because doing so would greatly reduce the amount of effort required by field practitioners during data collection. At two points during refinement (steps 6 and 9), we explored replacing the metric based on the regional (AW) list of perennial indicator taxa with a more general list for the entire West. Even though this replacement increases the number of taxa practitioners in the AW need to learn (9 vs 15 taxa), we expect most practitioners within the AW would also need to use the WM SDAM, and having both methods use the same indicator taxa would simplify implementation for most users in the regions. In addition to these refinements, we explored the impacts of dropping all aquatic invertebrate metrics (step 10), all hydrophyte metrics (step 11), or both (step 12), because these indicators require more training. We also explored metrics based on a reduced list of indicators, excluding taxa that would be more challenging for most practitioners to identify in the field (e.g., Empididae, step 14) and taxa with frequency below 0.10 in the dataset (i.e., Pteronarcyidae, Leptoceridae, and Uenoidae, step 15). This last refinement (step 15) resulted in the final selected set of indicators:

- *DifferencesInVegetation_score*
- *AlgalCover_LiveOrDead_NoUpstream*
- *BankWidthMean*
- *UplandRootedPlants_score*
- *RifflePoolSeq_score*
- *Slope*
- *hydrophytes_present_0_1_2_3_4_5*
- *perennial_ISAsubregion_abundance_simp2_0_5_10_20*

Although most refinements had small impacts on performance, dropping all invertebrate or plant metrics was associated with large drops in classification accuracy based on the training and testing datasets (Table 24, Figure 16).

Table 23. Steps in refining the AW model. Step 15, marked with an asterisk (*) was the final selected refined model. Grey boxes identify indicators included in a refined model, B = binned modification, and S = simplified modification.

Step	Preceding step	DifferencesInVegetation_score	PctShading	BankWidthMean	UplandRootedPlants_score	Slope	AlgalCover_LiveOrDead_NoUpstream	hydrophytes_present	RiffflePoolSeq_score	perintper_ISAsubregion_abundance	perennial_ISAsubregion_abundance	perennial_ISA_abundance	Modification
0	NA												Preliminary model
1	0												Dropped perintper_ISAsubregion_abundance
2	0												Dropped perennial_ISAsubregion_abundance
3	1												Dropped slope
4	1												Dropped RiffflePoolSeq_score
5	1												Dropped PctShading
6	5												Replaced perennial_ISAsubregion_abundance with perennial_ISA_abundance
7	5										B		Binned perennial_ISAsubregion_abundance at 0, 5 and 10
8	7							B			B		Binned hydrophytes_present at 0, 1, 2, 3, 4, and 5
9	8							B				B	Binned perennial_ISA_abundance at 0, 5, 10 and 20
10	9							B					Dropped perennial_ISA_abundance binned
11	9											B	Dropped hydrophytes_present binned
12	9												Dropped hydrophytes_present binned and perennial_ISA_abundance binned
13	9							B			B		Replaced regional perennial indicators with subregional indicators that were binned perennial_ISAsubregion_abundance at 0, 5, 10 and 20
14	13							B			S		Simplify perennial_ISAsubregion_abundance by excluding Empididae; binned at 0, 5, 10, and 20.
15*	14							B			S		Simplify perennial_ISAsubregion_abundance by excluding Empididae and taxa with frequency <0.1; binned at 0, 5, 10, and 20.

Table 24. Training and testing accuracy of each step in the refinement of the AW model. Step 15, marked with an asterisk (*) was the final selected refined model. Statistics are calculated for the entire AW assessment stratum.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0	None	0.92	0.81	0.84	0.69	0.88	0.62
1	0	0.93	0.82	0.86	0.70	0.87	0.64
2	0	0.93	0.81	0.86	0.69	0.87	0.61
3	1	0.92	0.81	0.85	0.69	0.86	0.60
4	1	0.93	0.80	0.85	0.67	0.85	0.62
5	1	0.94	0.82	0.87	0.71	0.88	0.65
6	5	0.93	0.80	0.88	0.68	0.87	0.64
7	5	0.93	0.82	0.88	0.71	0.86	0.67
8	7	0.93	0.82	0.87	0.70	0.86	0.66
9	8	0.93	0.81	0.87	0.70	0.86	0.66
10	9	0.92	0.81	0.85	0.70	0.83	0.63
11	9	0.93	0.80	0.87	0.68	0.87	0.62
12	9	0.92	0.79	0.85	0.67	0.83	0.58
13	9	0.94	0.82	0.88	0.70	0.87	0.62
14	13	0.93	0.81	0.87	0.69	0.86	0.63
15*	14	0.92	0.82	0.87	0.71	0.86	0.64

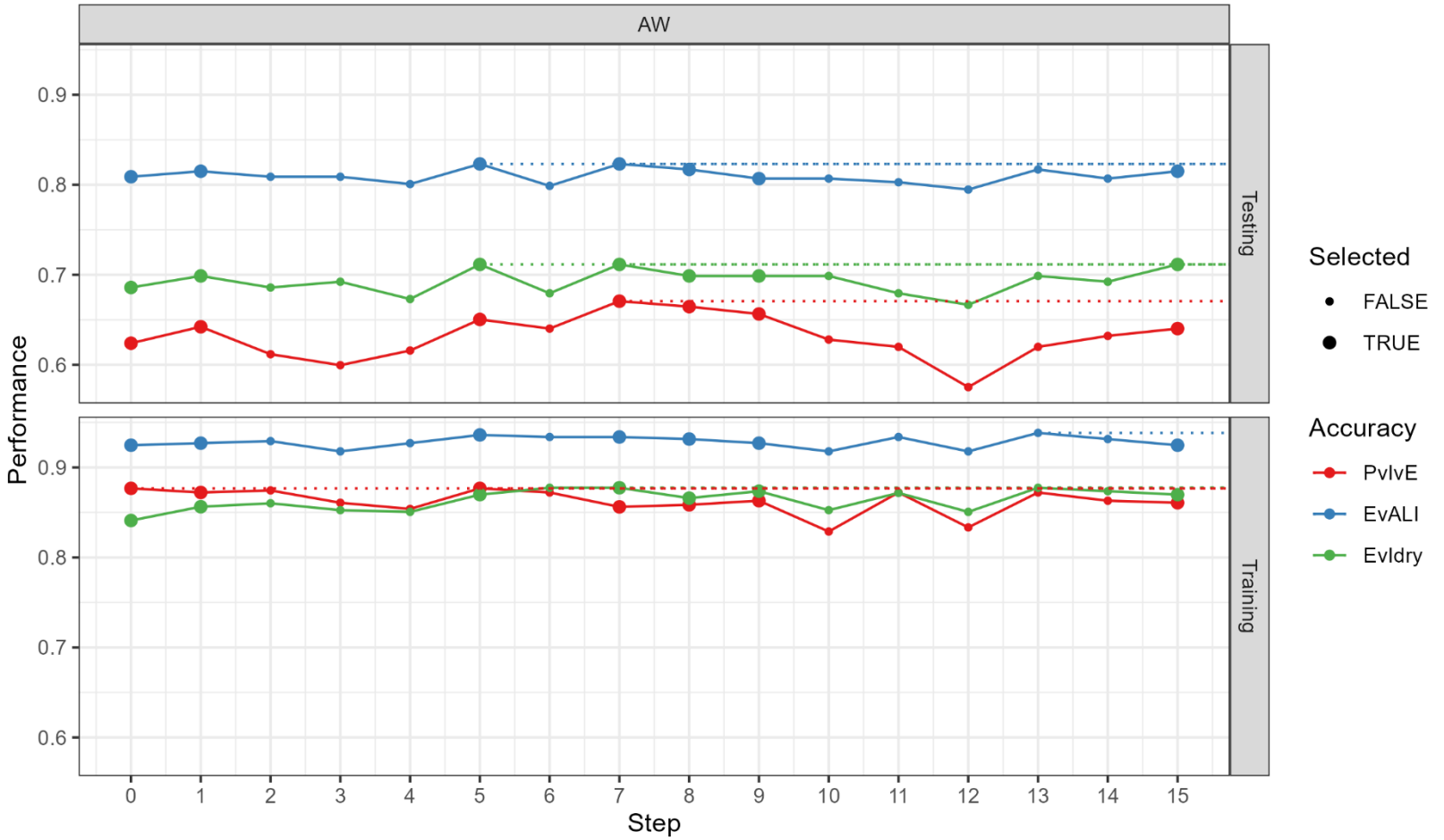


Figure 16. Accuracy of each step in the refinement of the AW model. Dotted horizontal lines indicate the peak level of accuracy observed over all refinement steps. Performance statistics were calculated for the entire AW assessment stratum.

3.6.2 WM model refinement

Key steps of model refinement are described in Table 25. After each step, we assessed the performance in terms of accuracy (measured as PvIvE, EvALI, and EvIdry) in both training and testing datasets (Table 26, Figure 17) for three assessment strata (i.e., the entire WM, the snow influenced portion, and the not snow influenced portion). First, we sought to drop potentially redundant or complicated metrics, possibly replacing them with simpler or more comprehensive alternatives. For example, for the WM model, two closely related aquatic invertebrate metrics were selected: one based on perennial indicator taxa, and another based on perennial plus intermittent indicator taxa. We explored dropping metrics that required special equipment (i.e., a clinometer or densiometer, steps 4 and 7), as well as other metrics with relatively low importance (e.g., the substrate sorting score, step 3). After reducing the number of variables, we explored ways to bin hydrophyte and aquatic invertebrate metrics (steps 8, 9, and 10).

At step 2, we explored replacing the metric based on the regional (WM) list of perennial indicator taxa with a more general list for the entire West. Doing only slightly increased the number of indicator taxa practitioners would need to learn (14 vs. 15), while having the advantage of eliminating the need for practitioners working in both the AW and WM to use different indicator taxa lists.

In addition to these refinements, we explored the impacts of dropping all aquatic invertebrate metrics (step 11), all hydrophyte metrics (step 12), or both (step 13), because indicators demand the greatest level of expertise. We also explored metrics based on a reduced list of indicators, excluding taxa that would be challenging for most practitioners to identify in the field (e.g., Empididae, step 17) and taxa with frequency below 0.10 in the dataset (i.e., Pteronarcyidae, Leptoceridae, and Uenoidae, step 18). This last refinement (step 18) resulted in the final selected set of indicators:

- *DifferencesInVegetation_score*
- *BankWidthMean*
- *RifflePoolSeq_score*
- *PctShading*
- *UplandRootedPlants_score*
- *Slope*
- *SubstrateSorting_score*
- *EPT_abundance_0_5_10_20*
- *hydrophytes_present_0_1_2_3_4_5*
- *perennial_ISAsubregion_abundance_simp2_0_5_10_20*

Although most refinements had small impacts on performance, dropping all invertebrate or plant metrics was associated with large drops in accuracy in both the training and testing datasets (Table 26, Figure 17).

Table 25. Steps in refining the WM model. Step 18, marked with an asterisk (*) was the final selected refined model. Grey boxes identify indicators included in a refined model, B = binned modification, and S = simplified modification.

Step	Preceding step	DifferencesInVegetation_score	BankWidthMean	hydrophytes_present	PctShading	RiffflePoolSeq_score	Slope	perintper_ISAsubregion_abundance	EPT_abundance	perennial_ISAsubregion_abundance	UplandRootedPlants_score	Ephemeroptera_abundance	SubstrateSorting_score	perennial_ISA_abundance	Modification
0	NA														Preliminary model
1	0														Drop *Ephemeroptera_abundance* and *perintper_ISAsubregion_abundance*
2	1														Replace WM indicators with Western indicators
3	2														Drop substrate sorting
4	2														Drop slope
5	2														Drop rooted upland plants
6	2														Drop riffle-pool sequence
7	2														Drop shading
8	2								B						Bin EPT abundance at 0, 5, 10 and 20
9	8								B					B	Bin perennial indicator abundance at 0, 5, 10, and 20
10	9			B					B					B	Bin hydrophytes at 0, 1, 2, 3, 4, and 5
11	10			B											Drop all BMI indicators
12	10								B					B	Drop hydrophytes
13	10														Drop BMI and hydrophyte indicators
14	10			B					B						Drop perennial BMI but keep EPT
15	10			B										B	Drop EPT but keep perennial BMI

Step	Preceding step	DifferencesInVegetation_score	BankWidthMean	hydrophytes_present	PctShading	RiffflePoolSeq_score	Slope	perintper_ISAsubregion_abundance	EPT_abundance	perennial_ISAsubregion_abundance	UplandRootedPlants_score	Ephemeroptera_abundance	SubstrateSorting_score	perennial_ISA_abundance	Modification
16	10			B					B	B					Replace western indicators with WM indicators
17	10			B					B					S	Simplify *perennial_ISAsubregion_abundance* by excluding Empididae; binned at 0, 5, 10, and 20.
18*	10			B					B					S	Simplify *perennial_ISAsubregion_abundance* by excluding Empididae and taxa with frequency <0.1; binned at 0, 5, 10, and 20.

Table 26a. Training and testing accuracy of each step in the refinement of the WM. Step 18, marked with an asterisk (*) was the final selected refined model. Statistics are calculated for the entire WM region.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0		0.92	0.86	0.86	0.68	0.83	0.68
1	0	0.93	0.84	0.86	0.62	0.84	0.67
2	1	0.93	0.85	0.86	0.60	0.85	0.69
3	2	0.93	0.85	0.87	0.64	0.85	0.67
4	2	0.92	0.85	0.84	0.64	0.83	0.67
5	2	0.93	0.84	0.88	0.61	0.84	0.67
6	2	0.93	0.83	0.87	0.59	0.85	0.65
7	2	0.93	0.83	0.87	0.56	0.84	0.64
8	2	0.93	0.82	0.87	0.63	0.85	0.70
9	8	0.92	0.83	0.86	0.65	0.84	0.67
10	9	0.92	0.85	0.85	0.63	0.84	0.69
11	10	0.90	0.79	0.84	0.56	0.80	0.53
12	10	0.92	0.85	0.85	0.64	0.84	0.69
13	10	0.90	0.79	0.85	0.60	0.82	0.56
14	10	0.92	0.81	0.87	0.57	0.84	0.65
15	10	0.93	0.79	0.89	0.58	0.85	0.65
16	10	0.93	0.86	0.87	0.64	0.85	0.71
17	10	0.93	0.85	0.87	0.60	0.85	0.69
18*	10	0.93	0.84	0.87	0.60	0.85	0.69

Table 26b. Training and testing accuracy of each step in the refinement of the WM model. Step 18, marked with an asterisk (*) was the final selected refined model. Statistics are calculated only using sites in WM-snow influenced stratum.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0		0.96	0.85	0.92	0.64	0.86	0.67
1	0	0.96	0.84	0.92	0.58	0.86	0.66
2	1	0.95	0.85	0.91	0.57	0.86	0.69
3	2	0.96	0.86	0.93	0.62	0.86	0.67
4	2	0.96	0.85	0.91	0.57	0.86	0.66
5	2	0.96	0.84	0.93	0.59	0.86	0.67
6	2	0.96	0.83	0.94	0.56	0.87	0.64
7	2	0.96	0.86	0.93	0.60	0.86	0.66
8	2	0.95	0.82	0.92	0.60	0.86	0.70
9	8	0.94	0.83	0.91	0.63	0.85	0.66
10	9	0.95	0.85	0.91	0.60	0.87	0.69
11	10	0.94	0.83	0.90	0.59	0.81	0.56

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
12	10	0.94	0.88	0.90	0.68	0.86	0.72
13	10	0.93	0.85	0.88	0.64	0.82	0.60
14	10	0.96	0.84	0.93	0.62	0.87	0.68
15	10	0.96	0.81	0.94	0.62	0.86	0.67
16	10	0.96	0.87	0.93	0.61	0.87	0.72
17	10	0.95	0.85	0.92	0.57	0.87	0.70
18*	10	0.95	0.84	0.93	0.55	0.86	0.69

Table 26c. Training and testing accuracy of each step in the refinement of the WM model. Step 18, marked with an asterisk (*) was the final selected refined model. Statistics are calculated using sites in the WM-not snow influenced stratum.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0		0.85	0.88	0.75	0.77	0.78	0.73
1	0	0.87	0.84	0.76	0.70	0.81	0.69
2	1	0.88	0.83	0.78	0.67	0.82	0.68
3	2	0.88	0.83	0.77	0.67	0.82	0.68
4	2	0.85	0.84	0.74	0.77	0.77	0.69
5	2	0.88	0.83	0.79	0.67	0.81	0.68
6	2	0.86	0.83	0.75	0.67	0.80	0.68
7	2	0.86	0.74	0.76	0.50	0.80	0.59
8	2	0.87	0.84	0.77	0.70	0.81	0.69
9	8	0.86	0.83	0.77	0.67	0.80	0.68
10	9	0.86	0.84	0.76	0.70	0.80	0.69
11	10	0.82	0.69	0.76	0.50	0.78	0.44
12	10	0.87	0.76	0.77	0.57	0.80	0.61
13	10	0.85	0.60	0.80	0.53	0.81	0.45
14	10	0.85	0.73	0.76	0.47	0.78	0.58
15	10	0.88	0.74	0.81	0.50	0.81	0.59
16	10	0.87	0.84	0.77	0.70	0.81	0.69
17	10	0.87	0.83	0.77	0.67	0.81	0.68
18*	10	0.88	0.84	0.78	0.70	0.81	0.69



Figure 17. Accuracy of each step in the refinement of the WM model. Dotted horizontal lines indicate the peak level of accuracy over all refinement steps. Performance statistics were calculated for the entire WM region, the snow influenced assessment stratum (WMsnow), and the not snow influenced assessment stratum (WMnosnow).

3.6.3 Single indicators

Most of the single indicators that were evaluated introduced more errors than they corrected (Figure 18). Although some single indicators based on aquatic invertebrates showed net benefits, especially in the AW, the benefits were small, and thus no single indicators were added to either SDAM. There were no big errors (i.e., perennial reaches misclassified as ephemeral) that could be fixed by single indicators in either region.

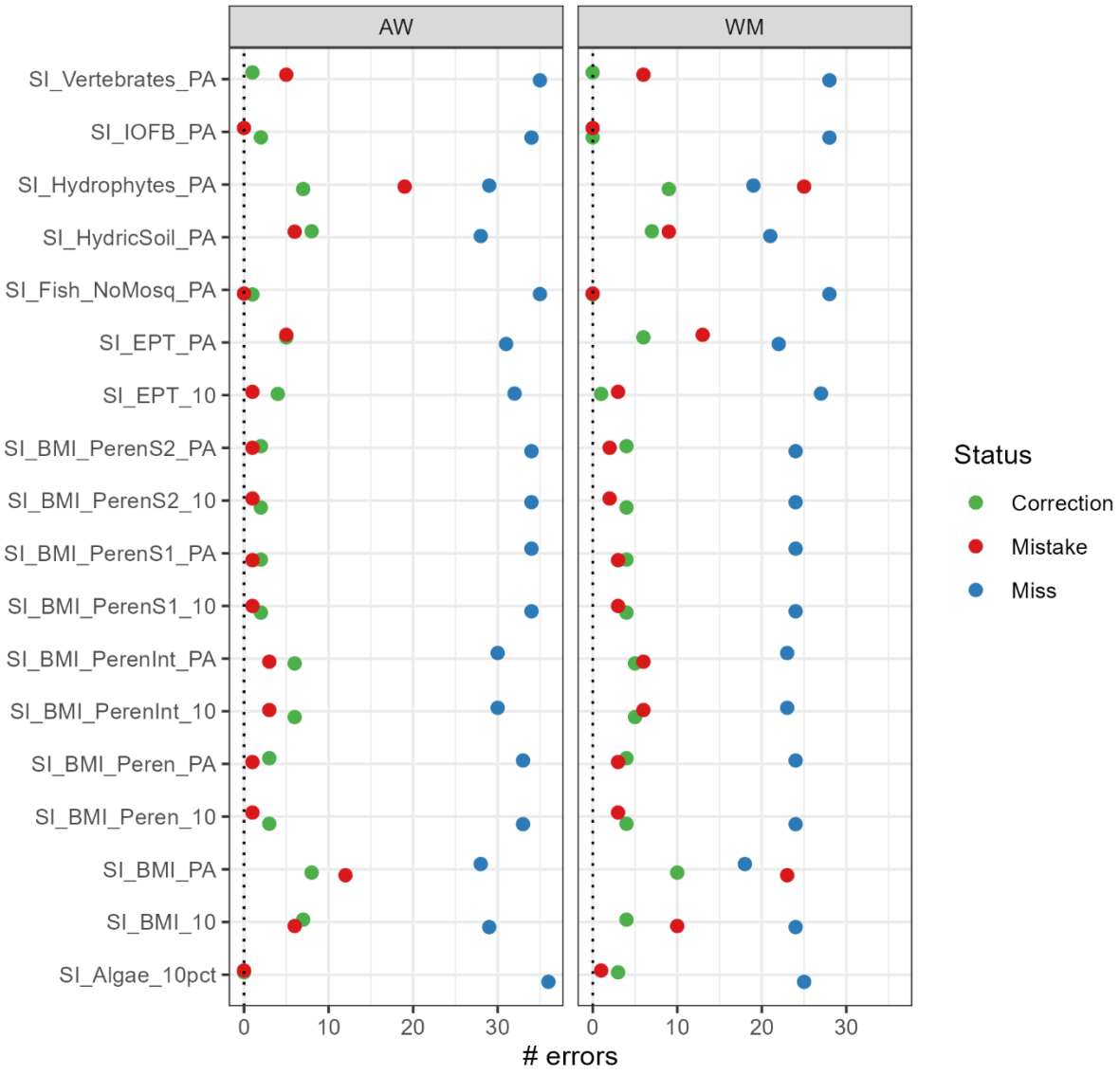


Figure 18. Effect of single indicators on the accuracy of the AW and WM SDAMs.

3.7 Performance assessment of the final models

When evaluated with testing data the final AW model had an overall PvIvE accuracy of 0.86, which dropped to 0.64 with testing data (Table 27, Figure 19). EvALI accuracy was substantially better, with measures of 0.93 and 0.82 in training and testing data respectively. PvLTP accuracy was comparable to EvALI (0.93 and 0.83 in training and testing data, respectively). Accuracy in distinguishing perennial from intermittent flowing reaches was at least 0.9 for both training and testing datasets (Table 28b).

For the WM model, numbers were overall comparable to those in the AW, with performance being somewhat better in the snow influenced portion than in the not snow influenced portion of the region; however, this difference largely disappeared in the testing data (Table 27, Figure 19).

Table 27a. Performance (accuracy) of the final AW and WM SDAMs. Note: These calculations are weighted to account for the varying number of visits to each reach.

Model	Dataset	Assessment Stratum	EvALI	PvLTP	PvIvE	EvIdry	PvIwet
AW	Training	AW	0.92	0.93	0.86	0.87	0.97
AW	Testing	AW	0.82	0.83	0.64	0.71	0.90
WM	Training	WM	0.93	0.92	0.85	0.87	0.97
WM	Testing	WM	0.84	0.84	0.69	0.60	0.92
WM	Training	WM-Not snow influenced	0.88	0.93	0.81	0.78	0.97
WM	Testing	WM-Not snow influenced	0.84	0.85	0.69	0.70	1.00
WM	Training	WM-Snow influenced	0.95	0.91	0.86	0.93	0.97
WM	Testing	WM-Snow influenced	0.84	0.84	0.69	0.55	0.91

Table 27b. Performance (precision) of the final AW and WM SDAMs. Note: These calculations are weighted to account for varying numbers of visits to each reach.

Model	Dataset	Assessment Stratum	EvALI	PvIvE
AW	Training	AW	0.90	0.80
AW	Testing	AW	0.84	0.72
WM	Training	WM	0.93	0.79
WM	Testing	WM	0.82	0.71
WM	Training	WM-Not snow influenced	0.91	0.80
WM	Testing	WM-Not snow influenced	0.82	0.65
WM	Training	WM-Snow influenced	0.94	0.79
WM	Testing	WM-Snow influenced	0.82	0.72

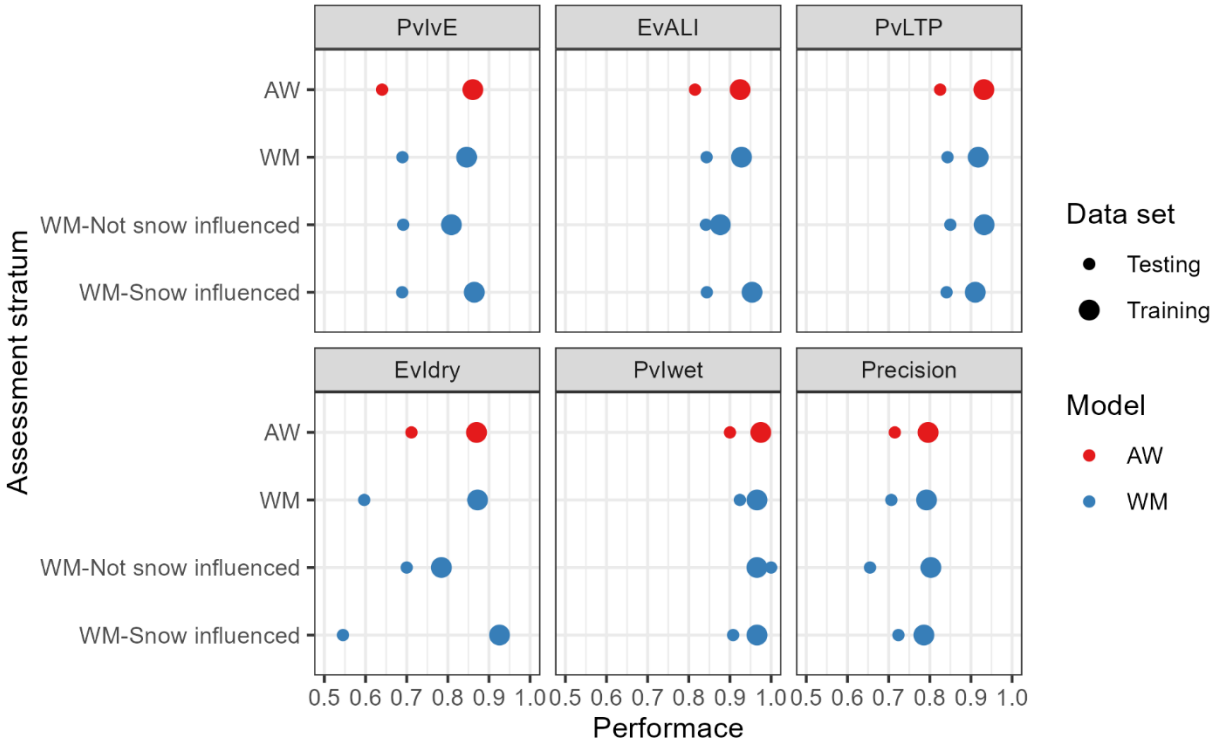


Figure 19. Performance (classification accuracy) of the final AW and WM SDAM models. Results for the WM are presented for the region as a whole, as well as for snow influenced and not snow influenced stratums separately.

Only 13 reach visits in the AW had ambiguous predictions (8 at least intermittent, and 5 less than perennial), as did only 20 in the WM (9 at least intermittent and 11 less than perennial). No reach visits in either region resulted in need more information (Table 28).

Table 28a. Comparison of predicted and true streamflow duration classes for the AW and WM models Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors." Note: The calculations in Table 27 are weighted to account for the varying number of visits to each reach and therefore do not match the numbers in this table.

Region	Dataset	Predicted class	E	I	P
AW	Training	E	111	21	0
AW	Training	I	7	88	11
AW	Training	P	0	11	66
AW	Training	ALI	0	4	2
AW	Training	LTP	1	1	2
AW	Training	NMI	0	0	0
AW	Testing	E	33	11	0
AW	Testing	I	6	17	5
AW	Testing	P	0	7	19
AW	Testing	ALI	0	2	0
AW	Testing	LTP	0	1	0
AW	Testing	NMI	0	0	0
WM	Training	E	91	15	0
WM	Training	I	15	108	20
WM	Training	P	0	13	81
WM	Training	ALI	0	3	4
WM	Training	LTP	3	2	0
WM	Testing	NMI	0	0	0
WM	Testing	E	17	7	0
WM	Testing	I	15	27	11
WM	Testing	P	0	12	26
WM	Testing	ALI	0	2	0
WM	Testing	LTP	2	3	1
WM	Testing	NMI	0	0	0

Table 28b. Comparison of predicted and true streamflow duration classes for the AW and WM models, with flowing and not flowing reach visits presented separately Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors". Note: The calculations in Table 27 are weighted to account for the varying number of visits to each reach and therefore do not match the numbers in this table.

Region	Dataset	Predicted class	E (flowing)	E (not flowing)	I (flowing)	I (not flowing)	P (flowing)	P (not flowing)
AW	Training	E	2	109	2	19	0	0
AW	Training	I	1	6	48	40	11	0
AW	Training	P	0	0	10	1	66	0
AW	Training	ALI	0	0	4	0	2	0
AW	Training	LTP	0	1	1	0	2	0
AW	Testing	E	1	32	3	8	0	0
AW	Testing	I	1	5	6	11	5	0
AW	Testing	P	0	0	6	1	19	0
AW	Testing	ALI	0	0	1	1	0	0
AW	Testing	LTP	0	0	0	1	0	0
WM	Training	E	6	85	3	12	0	0
WM	Training	I	1	14	54	54	20	0
WM	Training	P	0	0	10	3	81	0
WM	Training	ALI	0	0	2	1	4	0
WM	Training	LTP	0	3	1	1	0	0
WM	Testing	E	1	16	2	5	0	0
WM	Testing	I	3	12	20	7	11	0
WM	Testing	P	0	0	12	0	26	0
WM	Testing	ALI	0	0	2	0	0	0
WM	Testing	LTP	0	2	0	3	1	0

3.7.1 Accuracy at disturbed reaches

Disturbances were noted at 151 (35%) reach visits in the AW, and 212 (44%) reach visits in the WM. Within the AW, 69% of these disturbances were associated with urban or developed landuse, and 31% with agricultural landuse. Within the WM, 84% were associated with urban or developed land use, and 16% with agricultural landuse. Accuracy was slightly lower in disturbed reaches compared to undisturbed reaches, with a slightly larger effect in the WM than the AW. However, these effects were reduced or reversed in the testing dataset (Table 29).

Table 29. Comparison of accuracy at disturbed and undisturbed reach visits.

Stratum	Dataset	Measure	# disturbed reaches	# disturbed visits	Disturbed accuracy	# undisturbed reaches	# undisturbed visits	Undisturbed accuracy	Difference in accuracy (Disturbed minus undisturbed)
AW	Testing	EvALI	20	31	0.87	33	70	0.81	0.06
AW	Testing	PvLTP	20	31	0.80	33	70	0.84	-0.04
AW	Testing	PvIvE	20	31	0.67	33	70	0.65	0.02
AW	Testing	EvIdry	9	15	0.70	23	44	0.75	-0.04
AW	Testing	PvIwet	11	16	0.73	15	24	0.61	0.12
AW	Training	EvALI	73	120	0.90	105	205	0.93	-0.03
AW	Training	PvLTP	73	120	0.90	105	205	0.95	-0.05
AW	Training	PvIvE	73	120	0.81	105	205	0.88	-0.08
AW	Training	EvIdry	45	64	0.85	60	112	0.87	-0.02
AW	Training	PvIwet	33	54	0.75	52	92	0.89	-0.10
WM	Testing	EvALI	30	63	0.87	23	60	0.74	0.13
WM	Testing	PvLTP	30	63	0.81	23	60	0.83	-0.02
WM	Testing	PvIvE	30	63	0.69	23	60	0.57	0.11
WM	Testing	EvIdry	12	25	0.70	9	20	0.32	0.37
WM	Testing	PvIwet	21	36	0.71	16	38	0.71	-0.00
WM	Training	EvALI	98	149	0.88	111	206	0.93	-0.05
WM	Training	PvLTP	98	149	0.88	111	206	0.91	-0.04
WM	Training	PvIvE	98	149	0.76	111	206	0.85	-0.09
WM	Training	EvIdry	51	75	0.81	59	98	0.88	-0.06
WM	Training	PvIwet	56	71	0.74	63	104	0.85	-0.11

3.7.2 Sensitivity to variation in indicator values

Thresholds for small versus large changes were determined for every indicator in the final AW and WM models (Table 30).

Table 30. Thresholds for identifying small versus large changes from their original indicator values for the AW and WM SDAMs

Indicator	Decreases	Increases
BankWidthMean	<50%	>150%
Slope	<50%	>150%
UplandRootedPlants_score	-1.25	+1.25
DifferencesInVegetation_score	-1.25	+1.25
PctShading (WM only)	-25%	+25%
AlgalCover_LiveOrDead_NoUpstream (AW only)	-1.5 bins	+1.5 bins
EPT_abundance_0_5_10_20 (WM only)	-1.5 bins	+1.5 bins
Perennial_ISAsubregion_abundance_simp2_0_5_10_20	-1.5 bins	+1.5 bins
Hydrophytes_present_0_1_2_3_4_5	+1.5 taxa	-1.5 taxa
RifflePoolSeq_score	-1.25	+1.25
SubstrateSorting_score (WM only)	-1.25	+1.125

The AW SDAM had high sensitivity to 5 indicators and moderate sensitivity to 3 indicators (Table 31). The WM SDAM had high sensitivity to 7 indicators, moderate sensitivity to 2 indicators, and low sensitivity to one indicator. Neither model showed very high sensitivity to any metric. Sensitivity classifications were similar for the two methods, although the AW SDAM was more sensitive to slope and the WM model was more sensitive to the number of hydrophytes. For many indicators, an asymmetrical pattern was evident, in which changes in one direction—typically increases—had larger impacts than changes in the other direction (Table 31). For example, increasing the number of hydrophytes typically had a larger impact than decreasing them (Figure 20e). This pattern suggests that most of these indicators are treated as positive evidence of longer-duration flows. Thus, false positive (e.g., mistaking a FAC plant like mulefat, *Baccharis salicifolia*, for a FACW plant, like arroyo willow, *Salix lasiolepis*) will have larger consequences than false negatives. A notable exception was the rooted upland plant score indicator (Figure 20b), which showed different patterns in the two regions. In the WM, this indicator showed a greater sensitivity to increases, typical of other indicators. However, the opposite pattern was evident in the AW. This reversal may be due to the region's large ephemeral washes often having relatively sparse upland vegetation compared to the ephemeral headwaters in the WM.

Table 31. Sensitivity of indicators in the AW and WM SDAMs.

Indicator	Model	Large decrease	Small decrease	Small increase	Large increase	Sensitivity
BankWidthMean	AW	Medium	Small	Small	Small	Moderate
BankWidthMean	WM	Medium	Medium	Medium	Medium	Moderate
Slope	AW	Medium	Small	Small	Small	Moderate
Slope	WM	Small	Small	Small	Small	Low
UplandRootedPlants_score	AW	Large	Medium	Small	Medium	High
UplandRootedPlants_score	WM	Medium	Medium	Medium	Large	High
DifferencesInVegetation_score	AW	Large	Medium	Medium	Large	High
DifferencesInVegetation_score	WM	Large	Medium	Medium	Large	High
PctShading	WM	Medium	Small	Medium	Large	High
AlgalCover_LiveOrDead_NoUpstream	AW	Large*	Medium	Medium	Large	High
EPT_abundance_0_5_10_20	WM	Medium	Medium	Medium	Large	High
perennial_ISAsubregion_abundance_simp2_0_5_10_20	AW	Large	Medium	Small	Large	High
perennial_ISAsubregion_abundance_simp2_0_5_10_20	WM	Large	Medium	Medium	Large	High
hydrophytes_present_0_1_2_3_4_5	AW	Moderate	Small	Small	Medium	Moderate
hydrophytes_present_0_1_2_3_4_5	WM	Large*	Medium	Medium	Large	High
RifflePoolSeq_score	AW	Large	Medium	Small	Moderate	High
RifflePoolSeq_score	WM	Large	Medium	Medium	Large	High
SubstrateSorting_score	WM	Small	Small	Small	Medium	Moderate

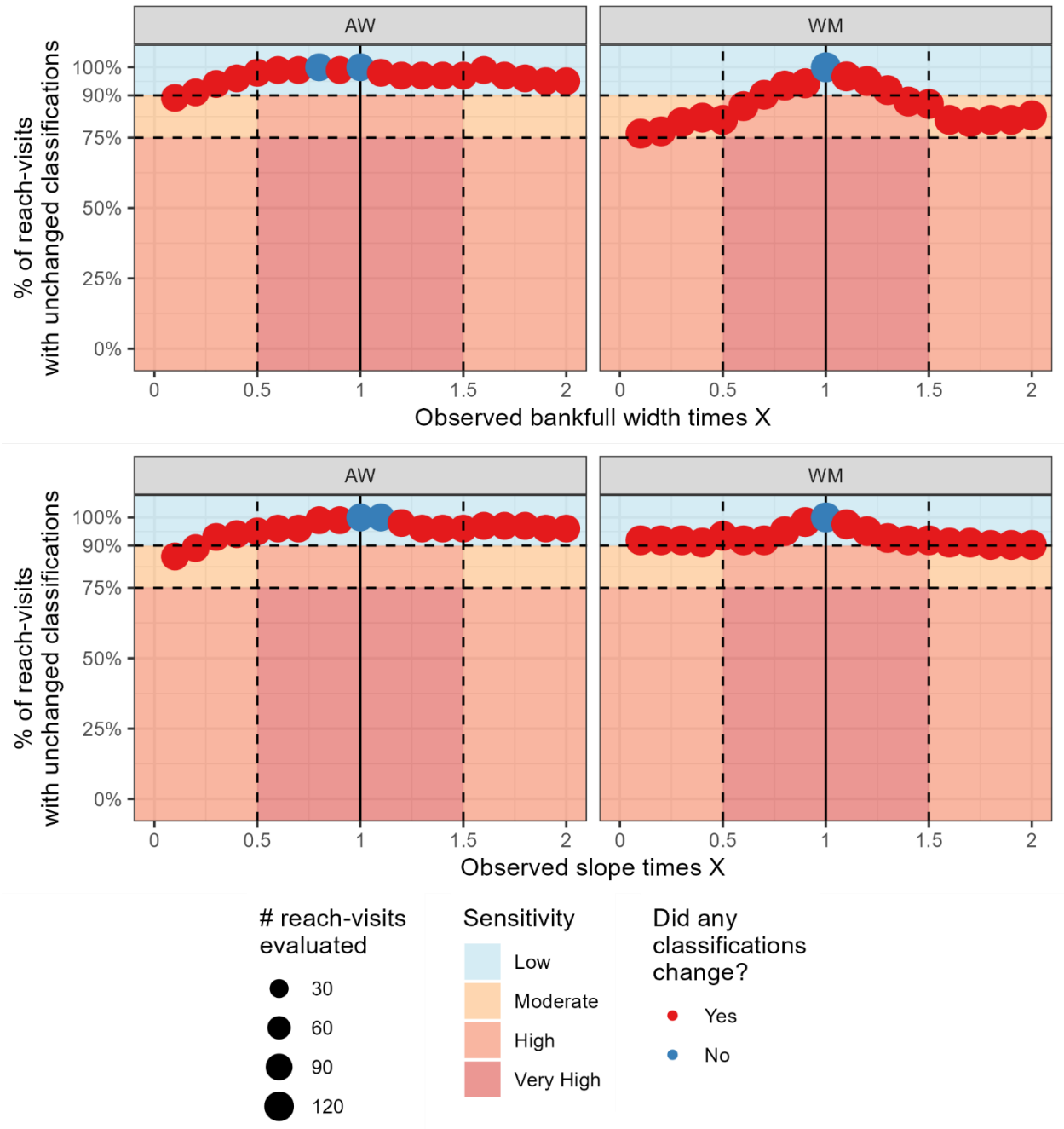


Figure 20a. Sensitivity of the AW and WM SDAMs to changes in indicator values: mean bankfull channel width (top) or slope (bottom). The x-axis represents the factor by which the original indicator value was multiplied, and the y-axis represents the percent of reach visits with unchanged classifications with the changed indicator value (all other indicator values held constant). The solid vertical lines indicate no change in the indicator value. Dashed vertical lines delineate the threshold between large and small changes in indicator value. Horizontal dashed lines indicate thresholds beteen small, moderate, and large impacts on the percent of reach visits with unchanged classifications.

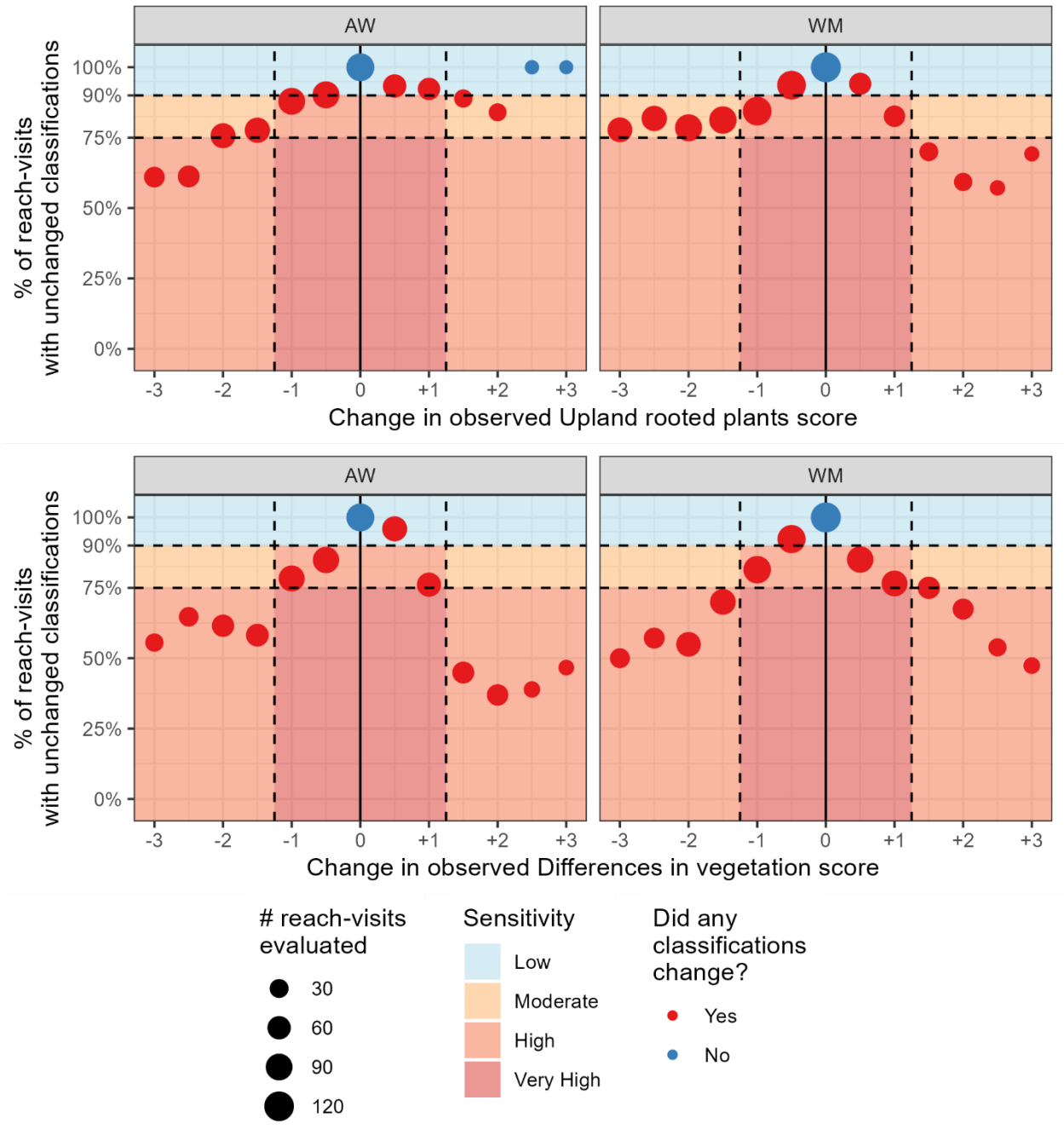


Figure 20b. Sensitivity of the AW and WM SDAMs to changes in indicator values: upland rooted plants score (top) or differences in vegetation score (bottom).

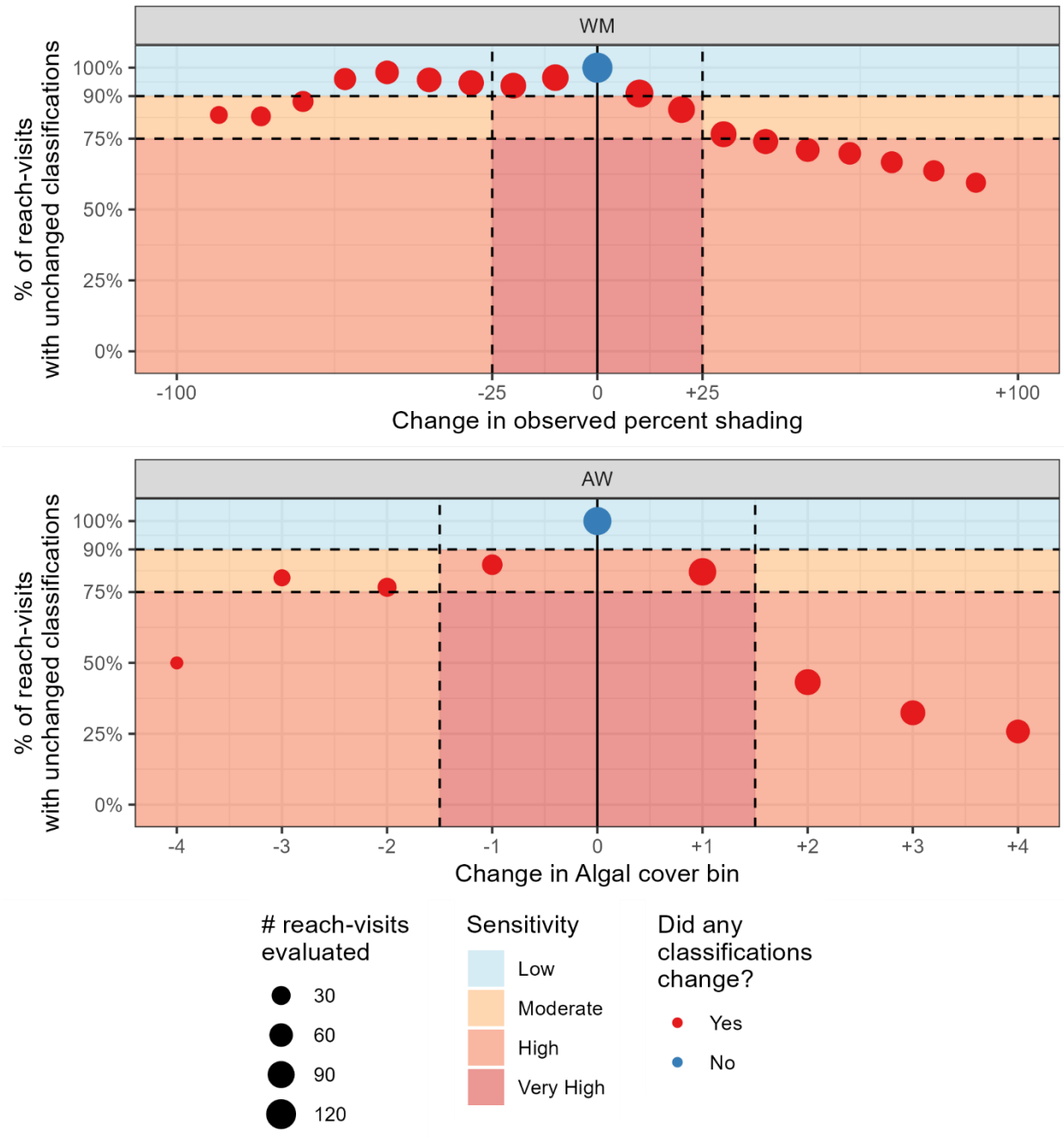


Figure 20c. Sensitivity of the AW and WM SDAMs to changes in indicator values: percent shading (top) and algal cover (bottom). Algal cover is visually estimated in bins: 0 (none detected), 1 (<2% cover), 2 (2 to 10% cover), 3 (10 to 40% cover), and 4 (≥40% cover).

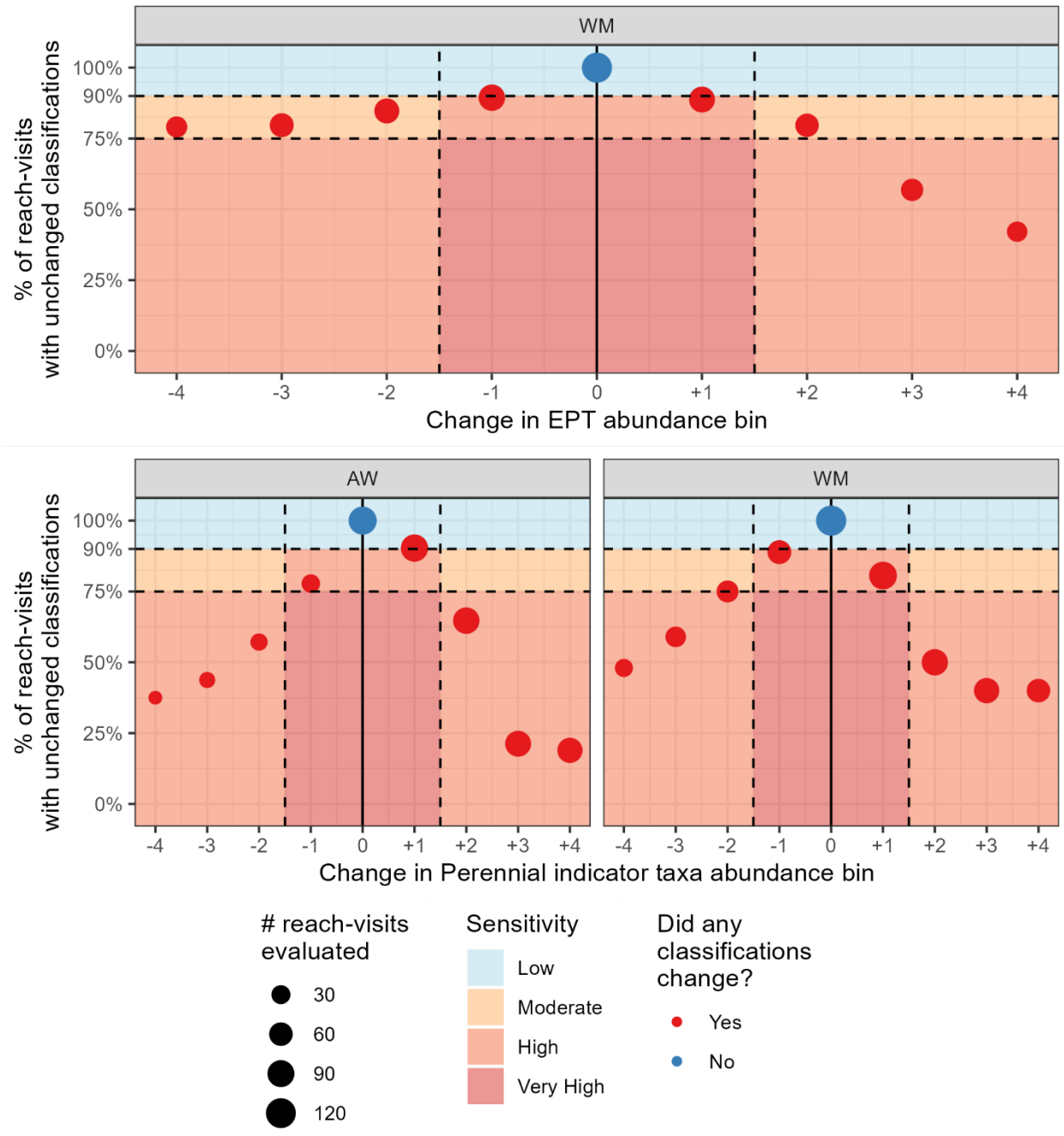


Figure 20d. Sensitivity of the AW and WM SDAM to changes in indicator values: EPT abundance (top; WM only) and perennial indicator taxa (bottom). EPT abundance is measured in bins: 0 (none detected), 1 (1- 4 individuals), 2 (5 to 9 individuals), 3 (10 to 19 individuals), and 4 (≥20 individuals). The abundance of perennial indicator taxa is also measured in bins: 0 (none detected), 1 (1 to 4 individuals), 2 (5 to 9 individuals), 3 (10 to 19 individuals), and 4 (20 or more individuals).

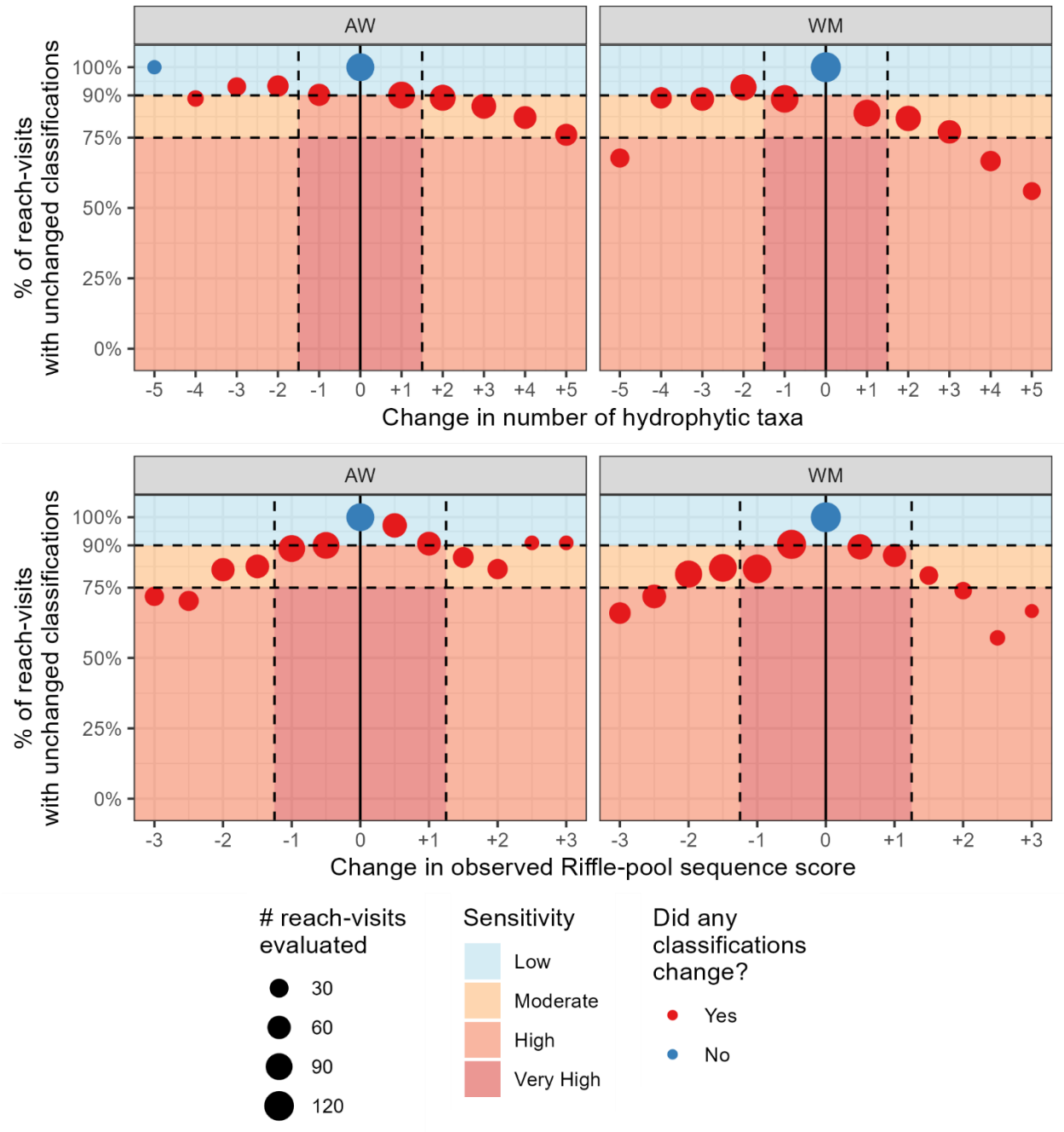


Figure 20e. Sensitivity of the AW and WM SDAMs to changes in indicator values: number of hydrophytic plant taxa (top) or riffle-pool sequence score (bottom).

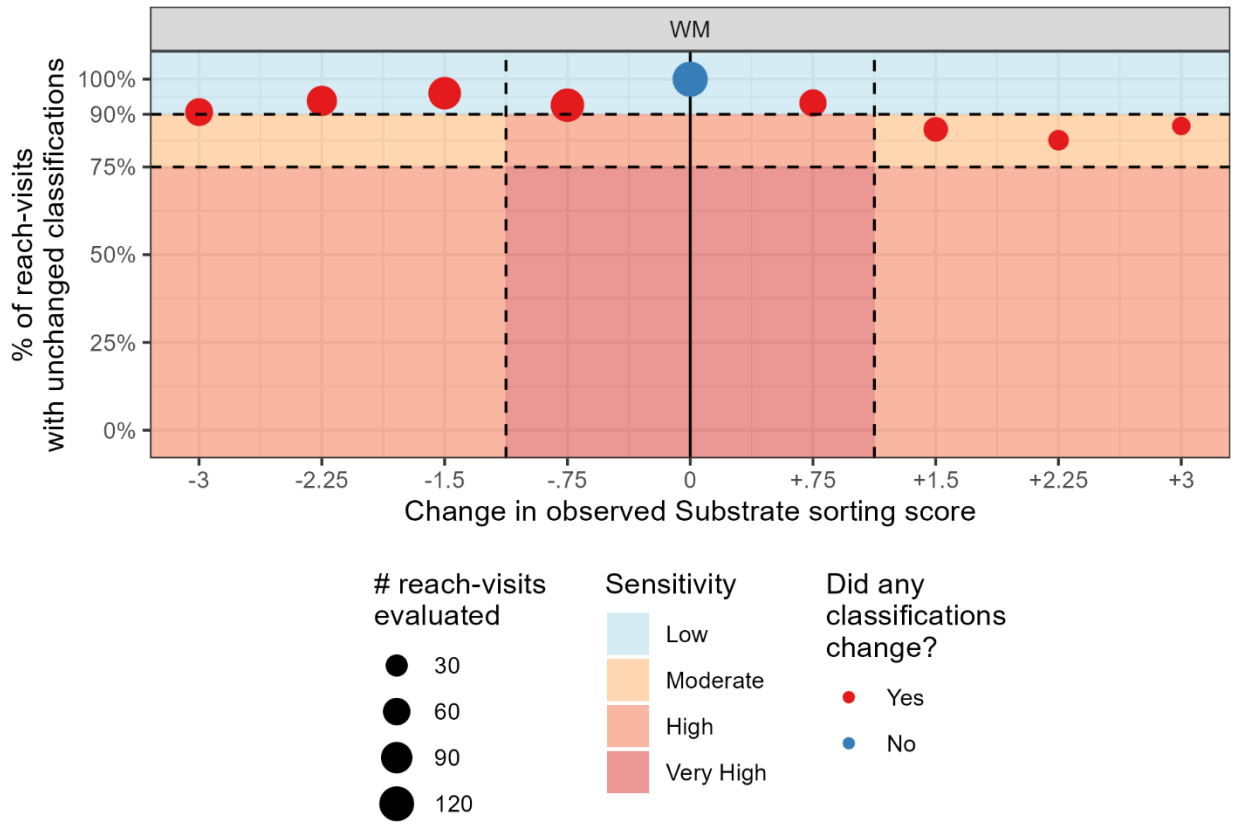


Figure 20f. Sensitivity of the WM SDAM to changes in indicator values: Substrate sorting score.

3.7.3 Sensitivity to time of sampling

Within the West, seasons were defined based on precipitation or temperature using reach-specific long-term monthly climatic statistics shown in Appendix B: Indicators and metrics evaluated for SDAM development. For each reach, we used climate metrics to assign months to wet or dry seasons, and to cold or warm seasons, based on whether they were above or below the median monthly statistic for that reach. Reach visits were then assigned to these seasons based on the month of the data collection event.

Reaches in the AW and WM have consistent seasonal patterns of temperature, with July or August typically being the warmest months; for one reach in the Montana and four reaches in Texas, April marked the start of the warm season, whereas for 390 other reaches, the warm season began in May. In contrast, precipitation patterns varied considerably among reaches within both regions (Figure 21). Some reaches, such as those in coastal California, receive nearly all precipitation during winter months, with pronounced drought during the summer. Other reaches, such as those in the Arizona-New Mexico mountains, receive the highest rainfall in the summer, with a smaller peak in mid-winter. Thus, seasonal definitions based on precipitation

varied considerably within the study area, compared to the more consistent temperature-based definition.

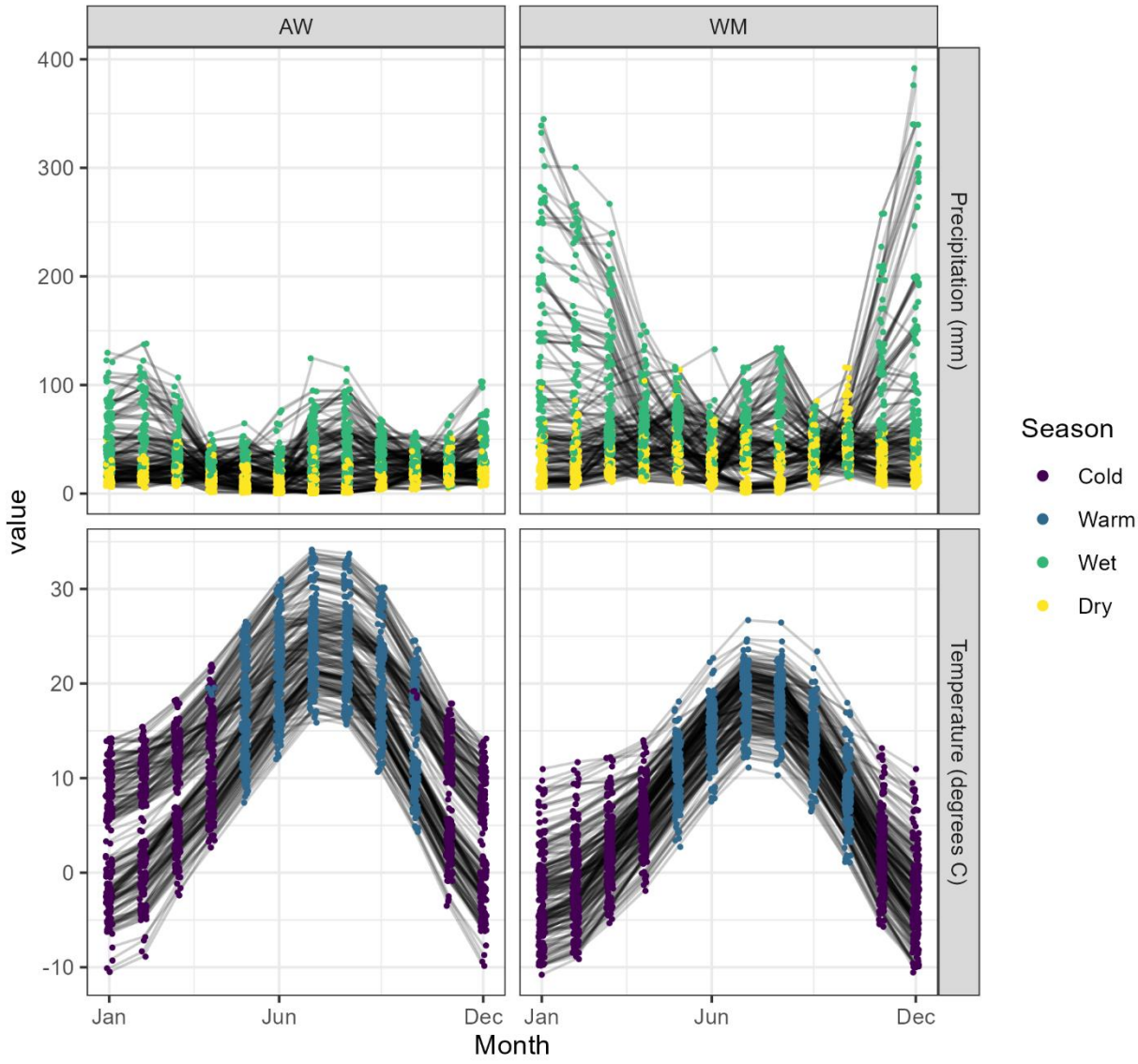


Figure 21. Long-term monthly averages of precipitation and temperature for reaches in the AW and WM.

Within the AW, reach visits were evenly distributed across seasons, whether they were defined by precipitation or temperature (Figure 21, Table 32). However, data were much less evenly distributed within the WM, as few sampling events occurred during the cold season (specifically, 22% of training data and 15% of testing data). This imbalance likely reflects the inaccessibility of many parts of the WM during winter months. In comparison, reach visits were more evenly distributed across seasons defined by precipitation. More reach visits in the AW and WM were in the Dry season than in the Wet season (Table 32). Over half of the reach visits in both regions occurred during Normal precipitation conditions and the remainder was roughly even between Drier than Normal precipitation and Wetter than Normal precipitation in the AW but in the WM

there were more than twice as many reach visits during Drier than Normal than during Wetter than Normal conditions. The inferential power for making accuracy comparisons between seasons and APC classes is limited by the large differences in number of reach visits, especially for the testing sets.

Table 32. Number of reach visits in each season as defined by temperature or precipitation, WebWIMP (WW), and Antecedent Precipitation Condition (APC) for the Arid West (AW) and Western Mountains (WM) training and testing datasets.

Region	Time of sampling comparisons	Classes	Dataset	
			Train	Test
AW	Temperature-based	Cold	136	50
		Warm	189	51
	Precipitation-based	Dry	147	45
		Wet	178	56
	WebWIMP season (WW)	Dry	240	71
		Wet	85	30
	Antecedent Precipitation Condition (APC)	Normal	208	57
		Drier than normal	61	23
		Wetter than normal	56	21
WM	Temperature-based	Cold	79	19
		Warm	276	104
	Precipitation-based	Dry	178	64
		Wet	177	59
	WebWIMP season (WW)	Dry	271	103
		Wet	84	20
	Antecedent Precipitation Condition (APC)	Normal	199	75
		Drier than normal	102	34
		Wetter than normal	54	14

Looking at the temperature and precipitation derived seasons overall, PvIvE accuracy tended to be greater during the warm season than the cold season in both the AW and WM (Figure 22); although this difference did not occur in the AW testing dataset. In contrast, EvALI and PvLTP accuracy statistics were more consistent across these seasons.

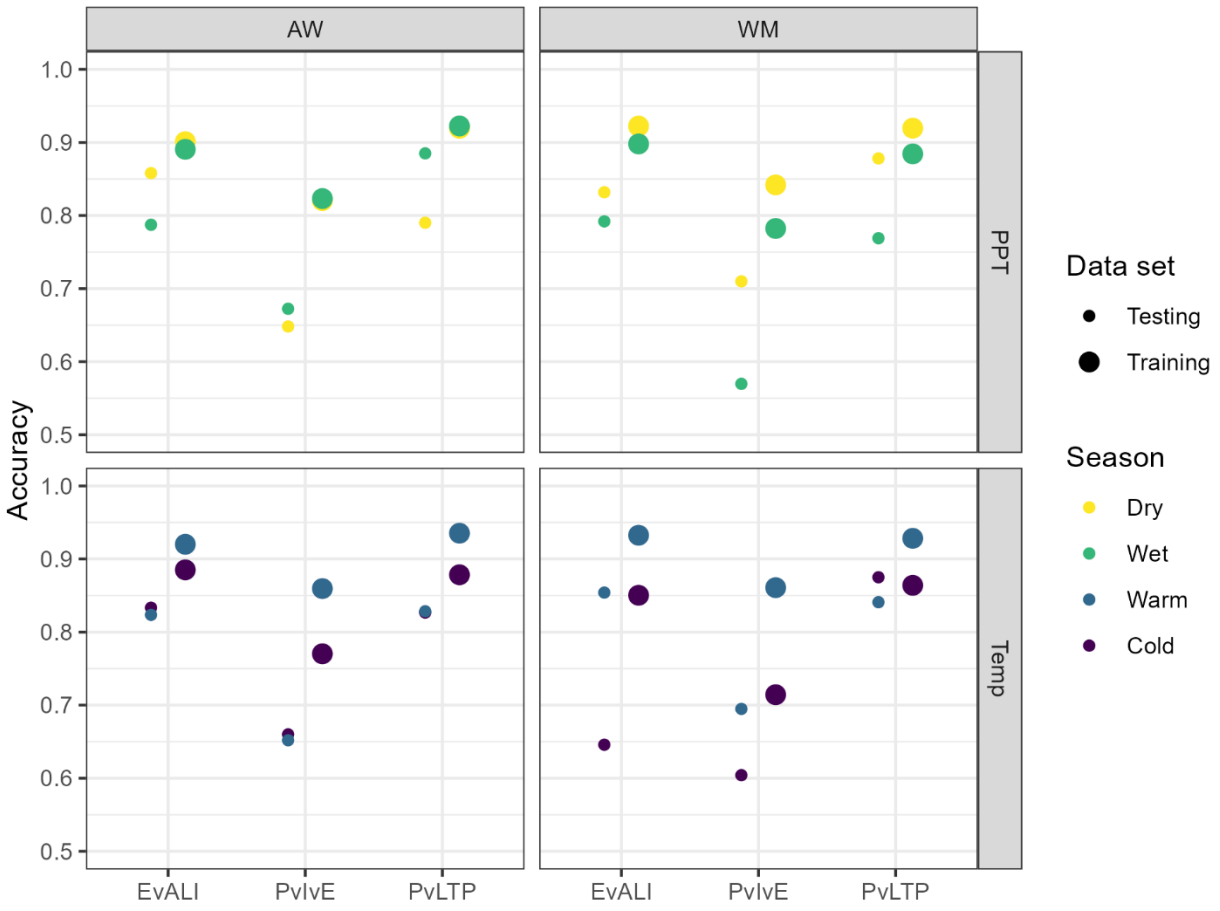


Figure 22. Accuracy of the AW and WM SDAMs in cold, warm, wet, and dry seasons derived from precipitation (PPT) and temperature (Temp) data.

Consistent with accuracy across the entire AW and WM datasets, EvALI and PvLTP accuracy were higher than PvIvE accuracy within seasons and APC conditions. In the seasonal analysis, the training accuracy was higher than testing accuracy, except for the PvLTP accuracy in AW and WM, and EvALI in AW being comparable between training and testing datasets (Figure 26). The training data set had more similar EvALI and PvIvE accuracy between Dry and Wet seasons in the AW than in the WM. Classification accuracy of the testing datasets for the AW and WM was higher during Wet season than Dry season except for EvALI accuracy in the WM, which had higher accuracy during the Dry season, with the inverse pattern in the training datasets (Figure 23). The classification accuracy of reach visits during Normal precipitation conditions was consistently higher than those of reach visits during Drier and Wetter than normal precipitation in the WM training dataset but not for the AW training dataset or the AW or WM testing datasets.

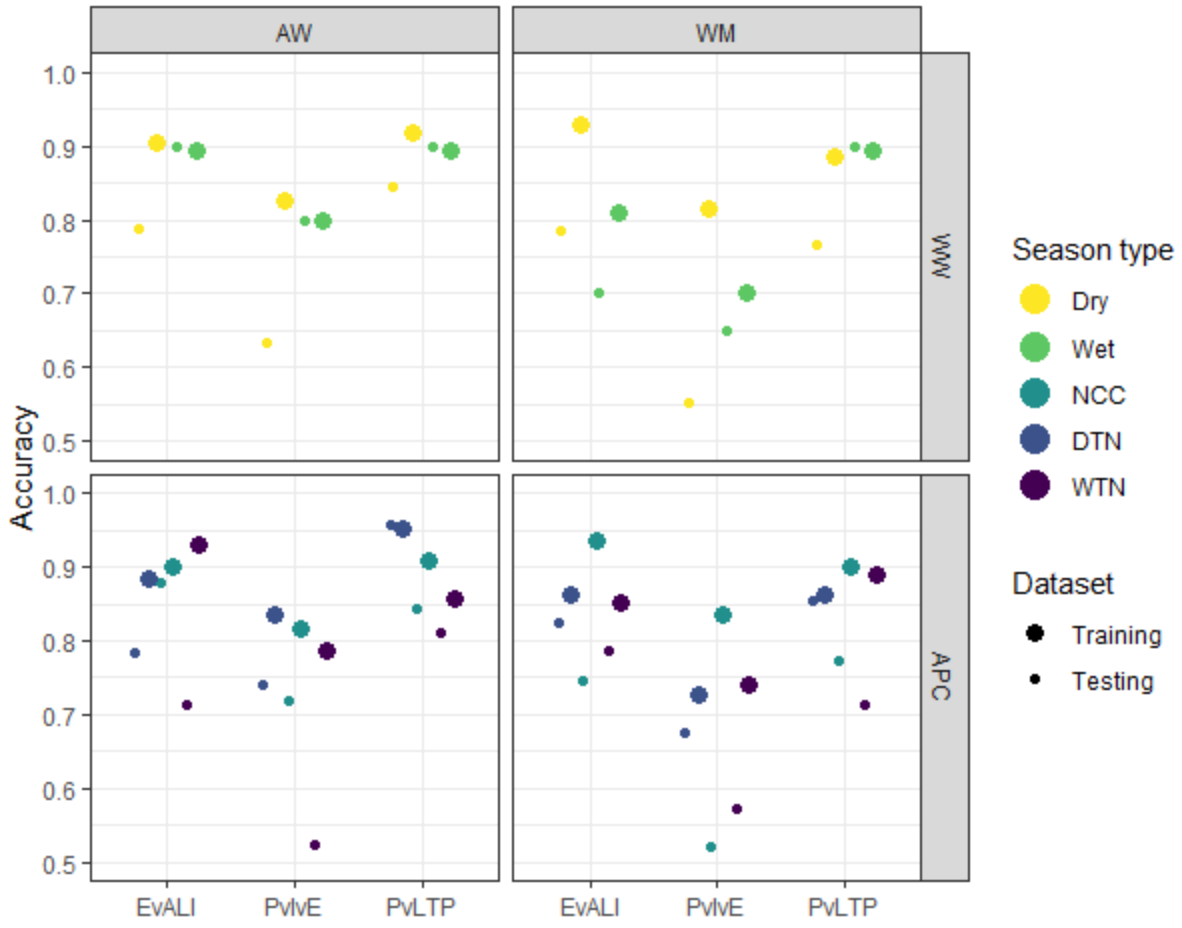


Figure 23. Accuracy of the Arid West (AW) and Western Mountains (WM) SDAMs in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) estimations for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN).

The directionality of classification errors of the AW and WM SDAMs did not vary systematically with WW or APC (Figure 24). Within the test dataset and specific to the EvALI and PvIvE comparisons, the AW tended to predict shorter streamflow duration classes during the dry season and longer streamflow duration classes during the Wet season. Classification errors of either SDAM were not more consistently balanced between predicting shorter or longer streamflow duration classes during Normal precipitation conditions than Drier or Wetter than normal conditions. There was also no greater frequency of the AW or WM SDAMs to predict shorter or longer streamflow duration classes among reach visits that occurred during Drier than normal or Wetter than normal conditions than among those during Normal precipitation (Figure 24).

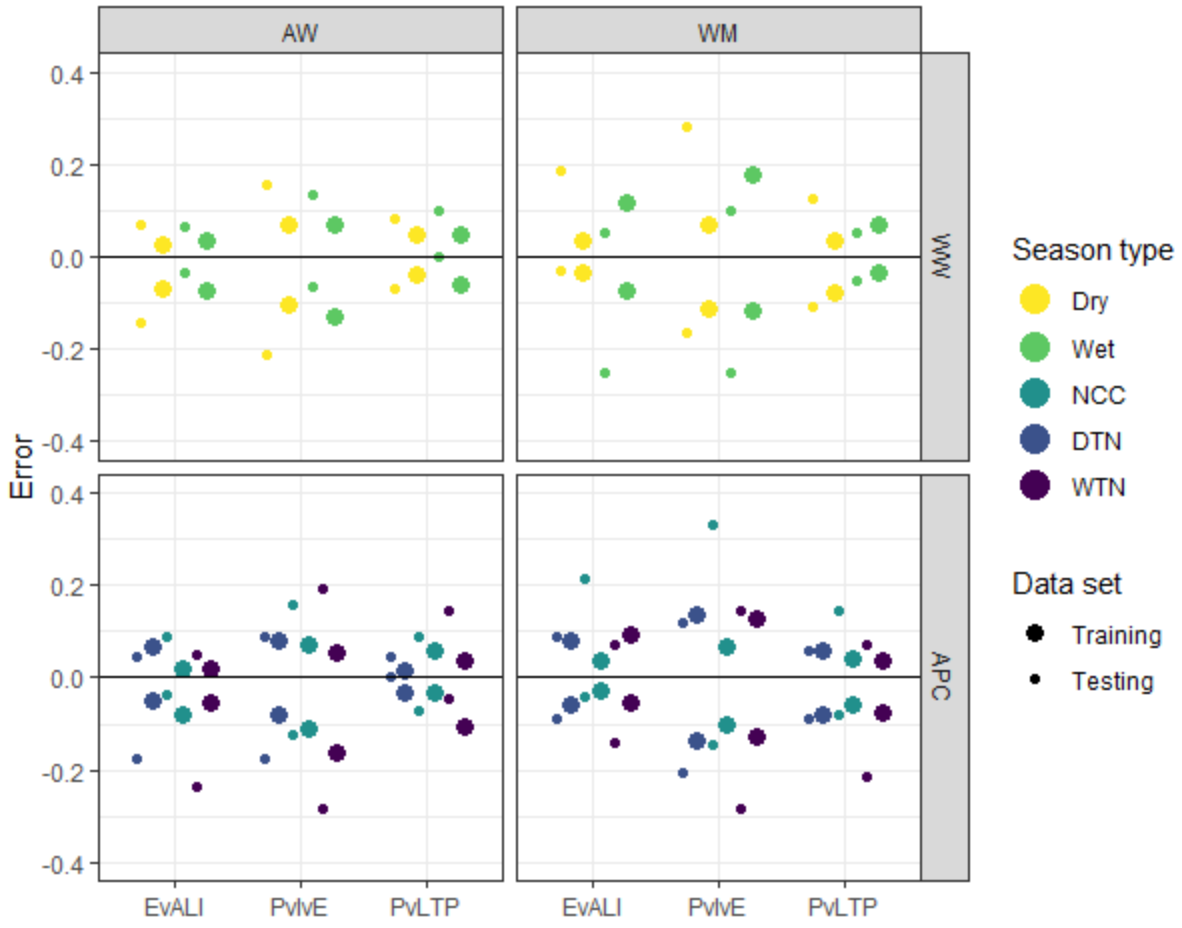


Figure 24. Error by direction of the Arid West (AW) and Western Mountains (WM) SDAMs in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN). Negative values represent the proportion of error predicting shorter flow duration classifications and positive values represent the proportion of error predicting longer flow duration classifications.

3.7.4 Comparison to other SDAMs

3.7.4.1 Comparison to other SDAMs used within the West or adjacent regions

We compared the performance of the final AW and WM SDAMs to other SDAMs used in adjacent areas or within the region using the testing datasets. The two methods were compared to the final GP SDAM (Section 4. Development of the GP SDAM) and the PNW SDAM (Nadeau 2015), both used in adjacent areas, and Phase 1 of the NM (NMED 2011), used within the defined study region. Single indicators were applied for methods that require them (i.e., the PNW SDAM, and NM). By almost every measure, the final AW and WM SDAMs had higher accuracy than other methods for the testing dataset in the West (Table 33, Figure 25), an aggregated region of the AW and WM as delineated for this study (Figure 7).

Table 33a. Accuracy of the AW SDAM compared to the WM, GP, and PNW SDAMs and the NM hydrology protocol in the AW region. Calculations were based on the testing dataset for reaches within the AW.

Measure	AW SDAM	WM SDAM	GP SDAM	PNW SDAM	NM
EvALI	0.82	0.71	0.70	0.79	0.75
PvIvE	0.64	0.55	0.60	0.60	0.55
PvLTP	0.83	0.83	0.87	0.79	0.80

Table 33b. Accuracy of the WM SDAM compared to the AW, GP, and PNW SDAMS and the NM hydrology protocol in the WM region. Calculations were based on the testing dataset for reaches within the WM.

Measure	WM SDAM	AW SDAM	GP SDAM	PNW SDAM	NM
EvALI	0.84	0.81	0.78	0.79	0.79
PvIvE	0.69	0.61	0.62	0.56	0.54
PvLTP	0.84	0.78	0.82	0.77	0.75

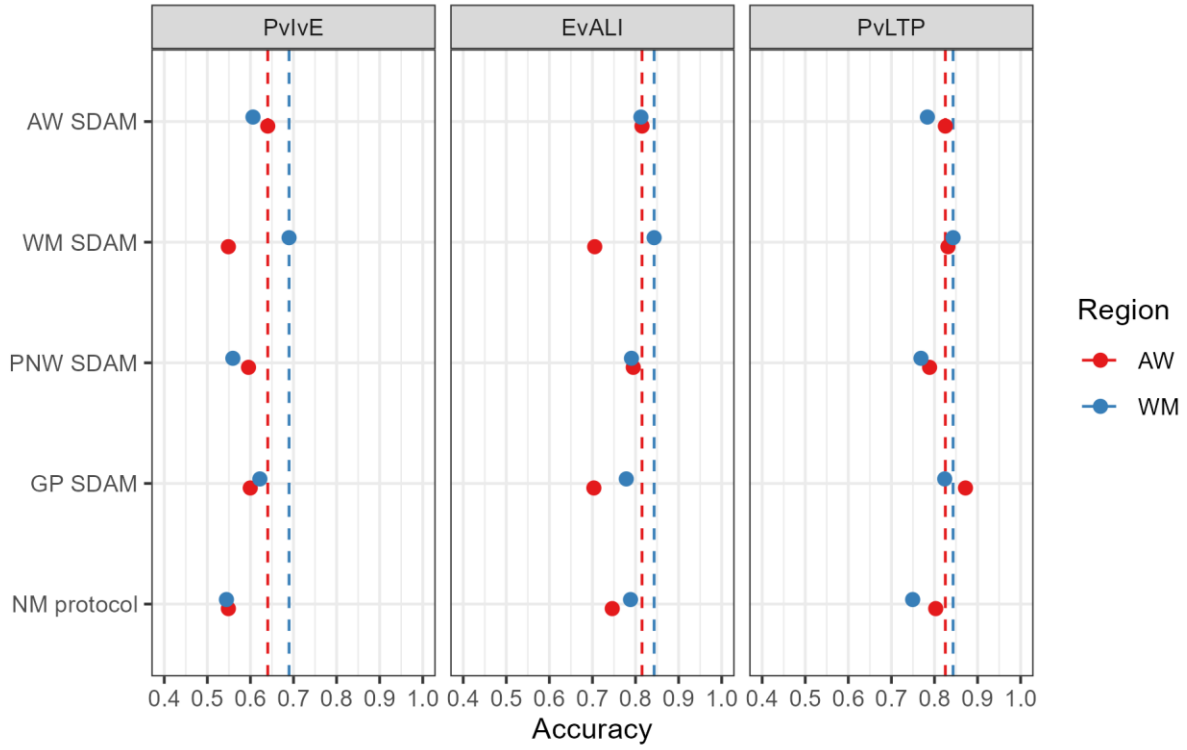


Figure 25. Comparison of classification accuracy of the AW and WM SDAMs to other methods of stream duration classification. Only results from the testing dataset are shown.

3.7.4.2 Comparison of the AW and WM SDAMs to the NM hydrology protocol within New Mexico

We compared the accuracy of the final AW and WM SDAMS to the NM hydrology protocol for reaches within the appropriate region of New Mexico. This comparison is intended to help practitioners understand the relative performance of the three methods available for use in that state. For these analyses, results for both training and testing datasets are presented (Table 34).

For the training datasets, performance of the AW SDAM was slightly better than the NM hydrology protocol, and the WM SDAM was very similar to the NM hydrology protocol; however, the testing datasets were more comparable across the three SDAMs (Figure 26). Note that for this comparison, we only evaluated the results of Phase 1 of the NM hydrology protocol, which in some cases requires a second phase of investigation to make a determination (NMED 2011).

Table 34. Number of reaches within New Mexico available for comparison of the AW and WM SDAMs to the NM hydrology protocol

Region	Set	E	I	P	Total
AW	Training	17	13	9	39
AW	Testing	0	1	4	5
WM	Training	10	18	21	49
WM	Testing	11	8	3	22

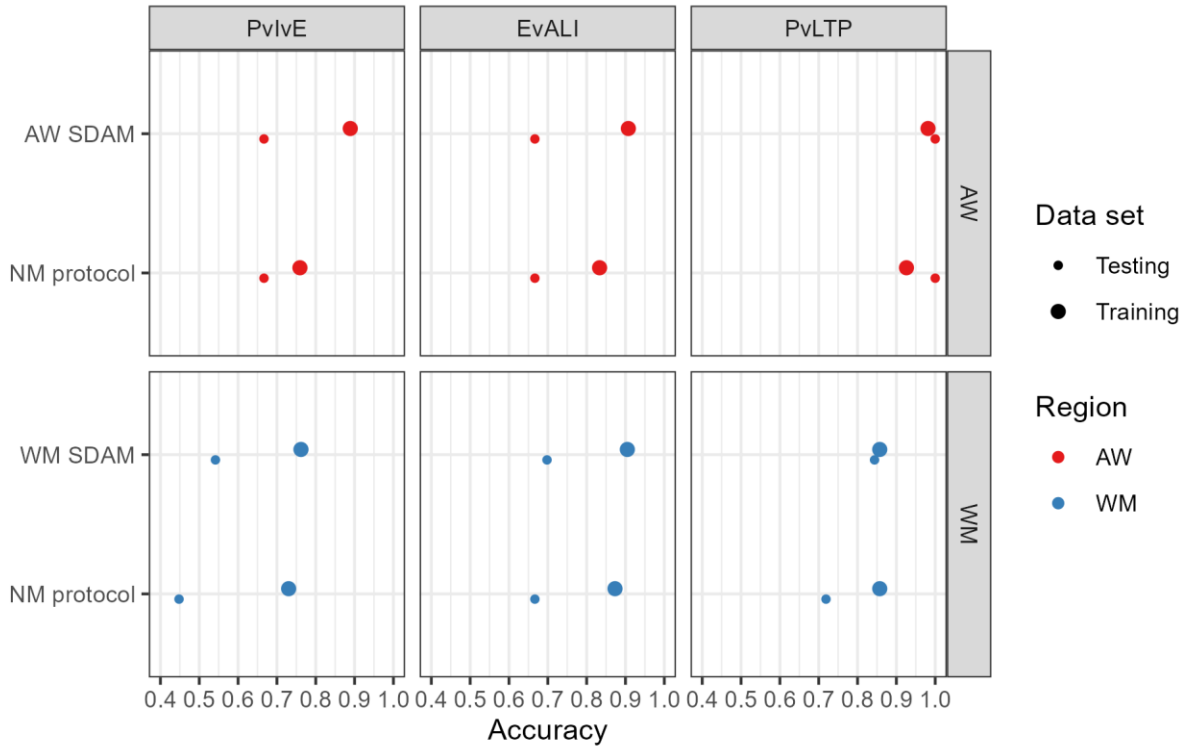


Figure 26. Accuracy of the AW SDAM, WM SDAM, and the NM hydrology protocol for reaches within the state of New Mexico.

3.7.4.3 Comparison to beta SDAMs

The two beta methods were developed from much smaller datasets in terms of number of reaches (AW: 89 vs 190; WM: 149 vs 200) and reach visits (AW: 101 vs 467; WM: 235 vs 637) than the final SDAMs, Thus, the final SDAMs are trained with a data set that has much better representation of the environmental conditions within the two regions compared to the beta SDAMs.

By some measures, the final AW and WM SDAMs had similar accuracy measurements of testing dataset to those reported for the beta methods. For example, the beta AW SDAM had EvALI accuracy of 0.88 compared to 0.82 for the final AW SDAM, and PvIvE accuracy of 0.50 vs 0.64 (beta method accuracy data from Mazor et al. 2023, but accuracy calculated as shown in Table 13). For the beta WM, the EvALI accuracy was similar in the beta and final methods (0.88 and 0.84, respectively; data from Mazor et al. 2022 but calculated as shown in Table 13), but much lower PvIvE accuracy (0.50 vs. 0.69, respectively). PvLTP accuracy was not reported for either beta method. However, performance statistics for the beta methods were calculated from different testing datasets compared to the final SDAMs, which could affect the comparison. There were too few reaches that were withheld from training datasets for both final and beta methods to provide meaningful comparisons with independent data (i.e., 12 reaches in the WM, 28 in the AW).

4. Development of the GP SDAM

The following section describes the specific data analysis steps used to develop and evaluate the final Great Plains (GP) SDAM.

4.1 Study area

The GP spans the central U.S. from Canada to Mexico and encompasses all or portions of 15 states (Figure 27). It includes areas largely dominated by native prairie-type vegetation (tall, short, and mixed grass) that generally receive less than 40 inches of precipitation each year. However, significant forested areas are also found in the northeast part of the Northern GP region, where average yearly rainfall totals are closer to the upper end of the range (30 to 40 inches). Ephemeral and intermittent reaches may be found at any position within a watershed but are more common in smaller headwaters, where flow accumulation is insufficient to sustain longer-duration flows. Ephemeral and intermittent reaches are also generally more common in semi-arid parts of these regions, where mean annual precipitation totals are lowest (10-20 inches), and evapotranspiration is relatively high.

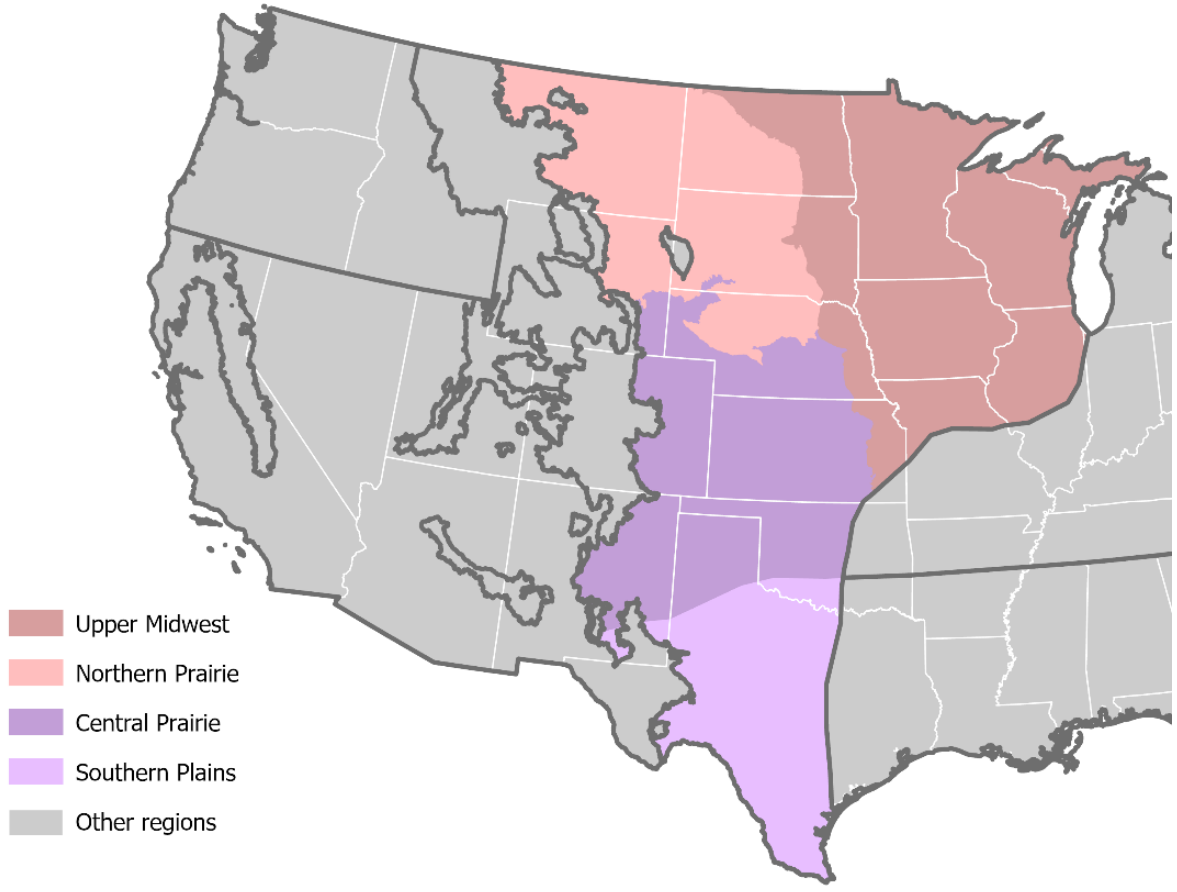


Figure 27. Map of the Great Plains region and its subregions.

There are several large and/or growing metropolitan areas within or partially within the GP, including Austin, Chicago, Dallas, Denver, Kansas City, Minneapolis, Milwaukee, and San Antonio. Thus, there are places within the GP regions where the need for an SDAM in permitting and management programs is particularly high. In addition, development associated with oil and natural gas, as well as agricultural uses that may require more and/or modified water sources due to climate change, is prevalent across the GP (Vengosh et al. 2014, Perkin et al. 2017). Within a portion of the Southern GP region, there is one protocol currently in use, applicable to New Mexico (New Mexico Environment Department [NMED] 2011).

4.2 Data collection

Data were collected at 287 reaches (1141 reach visits) in the Great Plains across four subregions: the Upper Midwest, the Northern Prairie, the Central Prairie, and the Southern Plains (Figure 28). Of these 287 reaches, 19 reaches were excluded from method development (13 in the Northern Great Plains and 6 in the Southern Great Plains) because a streamflow duration classification could not confidently be determined. The remaining 268 reaches were used for method development (Table 35). Based on logger data, gages, or other sources of hydrologic information, 27% of the reaches were classified as ephemeral, 38% as intermittent, and 35% as perennial. All three classes were represented in all four subregions, although ephemeral reaches were less common in the Upper Midwest, where only 5 reaches were identified. Of the 268 reaches included in the final Great Plains analysis, each reach was visited between one and six times, yielding a total of 1069 samples.

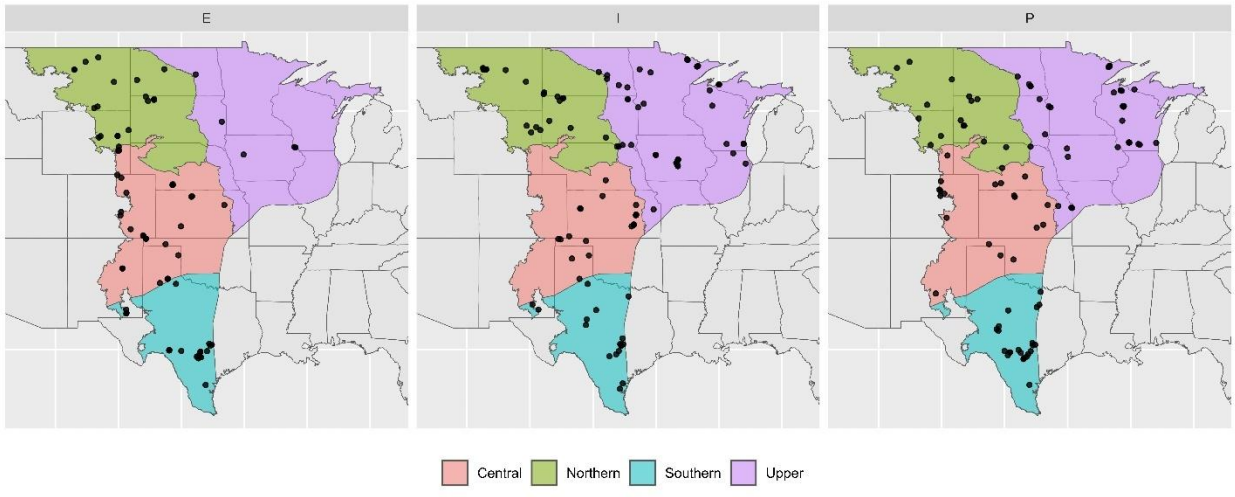


Figure 28. The study reaches used to create the final GP SDAM. The four subregions are also shown.

Table 35. Distribution of reaches used to develop the GP SDAM. Instrumented reaches were visited two to six times and had Stream Temperature, Intermittency, and Conductivity loggers installed and single-visit reaches were visited once (rarely, twice) and did not have loggers installed. Gaged reaches were located at reaches that had either nearby active or inactive USGS stream gages. Preferred reaches had multiple pre-existing data sources (e.g., water presence loggers and local expert observations) to support direct streamflow-duration classification. Acceptable reaches were classified based on local expert observations or short term logger studies) and were used to fill geographical gaps for study reaches.

Class	Subregions	Single visit, gaged	Single visit, preferred	Instrumented, gaged or other hydrologic data	Instrumented, preferred (based on multiple sources)	Instrumented, acceptable	Total
Ephemeral	***Total***	***14***	***14***	***6***	***8***	***30***	***72***
Ephemeral	Northern Prairie	3	6	0	1	9	19
Ephemeral	Upper Midwest	0	2	0	1	2	5
Ephemeral	Central Prairie	7	4	1	3	13	28
Ephemeral	Southern Plains	4	2	5	3	6	20
Intermittent	***Total***	***13***	***26***	***11***	***16***	***37***	***103***
Intermittent	Northern Prairie	4	3	4	3	7	21
Intermittent	Upper Midwest	1	13	1	1	26	42
Intermittent	Central Prairie	2	7	2	8	4	23
Intermittent	Southern Plains	6	3	4	4	0	17
Perennial	***Total***	***32***	***4***	***24***	***15***	***18***	***93***
Perennial	Northern Prairie	9	0	5	0	1	15
Perennial	Upper Midwest	8	1	5	7	14	35
Perennial	Central Prairie	8	1	7	4	0	20
Perennial	Southern Plains	7	2	7	4	3	23

4.3 Metric screening

Candidate indicator data were used to create 95 candidate metrics for consideration in the final Great Plains method. Of these candidates, 45 were biological, 7 were geomorphological, 2 were indirect hydrologic measurements, and 41 were geospatial (Appendix B).

As an initial data exploration step, the relationships between streamflow duration class and the 95 candidate indicators were ordinated using nonmetric multidimensional scaling using Gowers' distance (Gower 1971). Convex hulls were drawn around each flow class to help visualize their distributions in ordination space. In the Great Plains, the ordination spaces for intermittent and

perennial reaches were widely overlapping (Figure 29). The hull for ephemeral reaches was smaller, though largely contained within the area of the intermittent/ephemeral space. Candidate metrics contained within all three hulls may be less apt to clearly differentiate between streamflow duration classes.

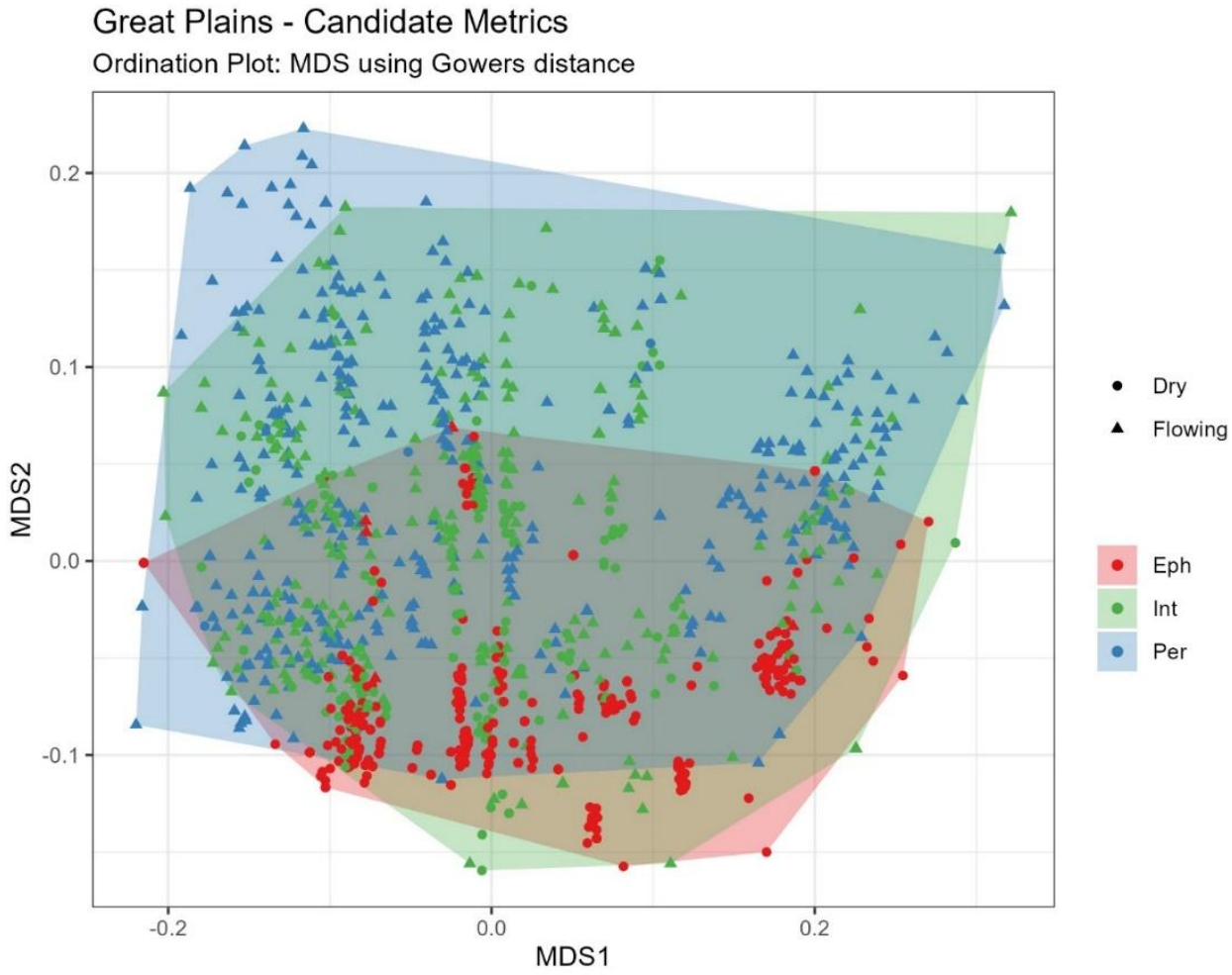


Figure 29. Nonmetric multidimensional scaling (MDS) plot of metric values at perennial (Per), intermittent (Int), and ephemeral (Eph) reaches in the Great Plains.

Next, candidate metrics were evaluated using criteria for inclusion in the SDAM methods (Table 11, Appendix B). Metric screening using these statistical processes only eliminated 5 out of the possible 95 candidate metrics (Table 36). Four metrics (ephemeral_ISAsubregion_abundance, ephemeral_ISAsubregion_taxa, Liverwort_cover, and snake_score) failed because they failed the percent dominance screen (>97% of samples were the same valuea). One metric (i.e., the ChannelDimensions_score_NM) failed to meet any responsiveness criteria. These five metrics were eliminated from further consideration.

Table 36. Screening criteria for metrics in the Great Plains. Metrics are described in Appendix B. The following fields provide the screening criteria – PctDom: Percent of reach samples with the most common value (typically zero). Min: minimum value. Max = maximum value. Range: Maximum possible value minus minimum possible value for the candidate metric. PvIvE: F-statistic from a comparison of mean values at perennial, intermittent, and ephemeral reaches. EvALI: Absolute t-statistic from a comparison of mean values at ephemeral and at least intermittent reaches. PvNP: Absolute t-statistic from a comparison of mean values at perennial and non-perennial reaches. PvIWet: Absolute t-statistic from a comparison of mean values at flowing intermittent and perennial reaches. EvIdry: Absolute t-statistic from a comparison of mean values at non-flowing intermittent and ephemeral reaches. rf_MDA: Variable importance from a random forest model, measured as mean decrease in accuracy. Screened: Indicates if the metric passed or failed screening criteria.

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Algae	AlgalCover_Live_NoUpstream	0.51	0	4	214.8	26.6	15.2	4.7	5.3	0.016	Pass
Biological	Algae	AlgalCover_LiveOrDead_NoUpstream	0.49	0	4	217.5	26.3	15.2	4.7	5.8	0.025	Pass
Biological	Algae	ironox_bfscore_NM	0.89	0	1.5	17.9	7.9	0.6	3.3	1.9	0.000	Pass
Biological	Aquatic invertebrates	BMI_presence	0.65	0	1	537.0	31.5	23.9	4.1	6.0	0.030	Pass
Biological	Aquatic invertebrates	ephemeral_ISA_abundance	0.93	0	21	4.0	0.9	3.6	3.1	0.8	0.000	Pass
Biological	Aquatic invertebrates	ephemeral_ISA_taxa	0.93	0	1	11.2	3.0	2.7	3.6	2.0	0.000	Pass
Biological	Aquatic invertebrates	ephemeral_ISAsubregion_abundance	0.97	0	21	3.7	1.9	2.5	1.3	1.8	0.000	**Fail**
Biological	Aquatic invertebrates	ephemeral_ISAsubregion_taxa	0.97	0	1	0.2	0.5	0.5	0.6	1.2	0.000	**Fail**
Biological	Aquatic invertebrates	Ephemeroptera_abundance	0.64	0	39	127.5	14.9	12.4	8.6	1.5	0.006	Pass
Biological	Aquatic invertebrates	Ephemeroptera_taxa	0.64	0	6	178.5	17.6	15.1	9.2	1.4	0.006	Pass
Biological	Aquatic invertebrates	ephinteph_ISA_abundance	0.69	0	23	35.8	5.5	4.8	6.7	3.2	0.005	Pass
Biological	Aquatic invertebrates	ephinteph_ISA_taxa	0.69	0	4	58.4	12.1	1.0	6.1	4.8	0.002	Pass
Biological	Aquatic invertebrates	ephinteph_ISAsubregion_abundance	0.65	0	30	28.9	6.7	2.4	4.8	3.8	0.003	Pass
Biological	Aquatic invertebrates	ephinteph_ISAsubregion_taxa	0.65	0	4	56.4	12.6	0.5	4.9	5.0	0.002	Pass
Biological	Aquatic invertebrates	EPT_abundance	0.55	0	58	189.2	17.5	15.0	10.8	1.8	0.013	Pass
Biological	Aquatic invertebrates	EPT_taxa	0.55	0	10	262.6	20.7	18.1	11.7	1.7	0.013	Pass
Biological	Aquatic invertebrates	GOLD_abundance	0.49	0	32	56.1	13.0	6.5	0.1	3.0	0.005	Pass
Biological	Aquatic invertebrates	Noninsect_abundance	0.55	0	51	66.4	15.4	7.7	1.5	3.3	0.003	Pass
Biological	Aquatic invertebrates	OCH_abundance	0.56	0	30	36.7	11.4	3.6	0.7	4.2	0.006	Pass
Biological	Aquatic invertebrates	OCH_taxa	0.56	0	6	89.5	17.8	8.2	1.3	4.9	0.007	Pass
Biological	Aquatic invertebrates	perennial_ISA_abundance	0.86	0	20	57.1	9.6	8.2	6.6	1.1	0.003	Pass
Biological	Aquatic invertebrates	perennial_ISA_taxa	0.86	0	3	104.9	12.5	11.2	8.9	1.4	0.003	Pass
Biological	Aquatic invertebrates	perennial_ISAsubregion_abundance	0.81	0	20	93.7	11.6	10.5	8.4	1.4	0.006	Pass
Biological	Aquatic invertebrates	perennial_ISAsubregion_taxa	0.81	0	2	167.0	14.4	14.2	11.2	1.4	0.007	Pass
Biological	Aquatic invertebrates	perennial_NC_abundance	0.52	0	44	189.5	18.4	15.2	9.9	1.9	0.015	Pass
Biological	Aquatic invertebrates	perennial_NC_live_abundance	0.52	0	44	185.7	18.5	15.0	9.8	2.0	0.014	Pass
Biological	Aquatic invertebrates	perennial_NC_taxa	0.51	0	9	261.1	21.0	18.4	10.8	1.9	0.015	Pass
Biological	Aquatic invertebrates	perintper_ISAsubregion_abundance	0.92	0	20	25.2	6.7	5.4	4.1	0.5	0.000	Pass
Biological	Aquatic invertebrates	perintper_ISAsubregion_taxa	0.92	0	3	43.2	8.5	7.1	5.5	0.5	0.000	Pass
Biological	Aquatic invertebrates	Plecoptera_abundance	0.94	0	12	12.4	6.2	3.9	1.7	0.0	0.000	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Aquatic invertebrates	Plecoptera_taxa	0.94	0	2	28.1	8.2	5.7	3.7	1.0	0.000	Pass
Biological	Aquatic invertebrates	Richness	0.36	0	18	293.6	26.6	18.5	7.2	4.6	0.031	Pass
Biological	Aquatic invertebrates	TotalAbundance	0.36	0	102	185.7	22.2	14.4	6.0	4.2	0.031	Pass
Biological	Aquatic invertebrates	Trichoptera_abundance	0.74	0	24	111.7	13.2	11.4	8.9	1.2	0.006	Pass
Biological	Aquatic invertebrates	Trichoptera_taxa	0.74	0	5	154.2	16.1	14.1	9.3	1.0	0.005	Pass
Biological	Vegetation	DifferencesInVegetation_score	0.28	0	3	172.0	18.1	13.9	4.8	5.6	0.012	Pass
Biological	Vegetation	hydrophytes_present	0.23	0	8	209.5	22.8	13.1	3.0	8.8	0.022	Pass
Biological	Vegetation	Liverwort_cover	0.97	0	3	9.4	2.9	3.5	3.1	0.4	0.000	**Fail**
Biological	Vegetation	Moss_cover	0.92	0	3	20.6	8.6	4.9	2.0	1.5	0.000	Pass
Biological	Vegetation	PctShading	0.35	0	100	9.7	3.5	1.0	2.9	2.3	0.006	Pass
Biological	Vegetation	UplandRootedPlants_score	0.59	0	3	407.5	23.0	21.4	4.9	6.4	0.031	Pass
Biological	Vertebrates	Amphibian_presence	0.85	0	1	28.4	11.2	3.1	1.3	3.3	0.000	Pass
Biological	Vertebrates	AmphSnake_PA	0.83	0	1	31.7	11.6	4.0	0.5	3.5	0.001	Pass
Biological	Vertebrates	Fish_PA	0.67	0	1	199.1	21.0	16.6	7.5	2.2	0.010	Pass
Biological	Vertebrates	snake_score	0.98	0	1	9.1	3.9	3.4	3.0	1.4	0.000	**Fail**
Geomorph	Geomorph	BankWidthMean	0.01	0.233333	68.33333	55.5	6.4	10.0	6.6	0.7	0.022	Pass
Geomorph	Geomorph	ChannelDimensions_score_NM	0.53	0	3	0.8	1.0	1.2	1.1	1.1	0.001	**Fail**
Geomorph	Geomorph	RifflePoolSeq_score	0.29	0	3	55.6	9.1	9.0	2.8	0.6	0.002	Pass
Geomorph	Geomorph	SedimentOnPlantsDebris_score	0.27	0	1.5	53.6	10.2	7.2	0.5	1.8	0.002	Pass
Geomorph	Geomorph	Sinuosity_score	0.53	0	3	28.1	7.9	3.4	0.1	4.7	0.001	Pass
Geomorph	Geomorph	Slope	0.42	0	25	10.9	2.9	5.4	2.7	0.5	0.002	Pass
Geomorph	Geomorph	SubstrateSorting_score	0.35	0	3	102.4	9.5	14.2	7.1	1.2	0.006	Pass
Geospatial	Precipitation	ppt	0.01	282.56	1037.8	39.8	9.0	5.9	1.2	3.4	0.013	Pass
Geospatial	Precipitation	ppt.11121	0.01	6.213333	65.82667	24.6	5.1	6.4	2.9	0.1	0.013	Pass
Geospatial	Precipitation	ppt.234	0.01	11.00333	79.42333	26.2	6.8	5.7	2.0	2.3	0.011	Pass
Geospatial	Precipitation	ppt.567	0.01	39.12667	128.9333	50.8	11.0	3.3	1.6	5.9	0.017	Pass
Geospatial	Precipitation	ppt.8910	0.01	23.01	102.5267	43.9	9.7	6.5	1.7	3.0	0.022	Pass
Geospatial	Precipitation	ppt.m01	0.03	4.18	55.62	21.4	3.4	6.4	3.7	1.6	0.023	Pass
Geospatial	Precipitation	ppt.m02	0.02	4.9	70.74	12.2	2.5	4.8	2.8	1.4	0.010	Pass
Geospatial	Precipitation	ppt.m03	0.02	8.74	89.28	15.2	4.4	5.0	2.4	1.1	0.011	Pass
Geospatial	Precipitation	ppt.m04	0.01	10.79	98.43	45.9	11.0	5.1	0.5	5.2	0.020	Pass
Geospatial	Precipitation	ppt.m05	0.01	26.77	137.45	37.9	8.9	3.1	1.3	5.0	0.016	Pass
Geospatial	Precipitation	ppt.m06	0.01	35.84	148.25	39.0	9.1	3.0	1.6	4.9	0.013	Pass
Geospatial	Precipitation	ppt.m07	0.02	29.33	123.92	46.9	11.3	2.8	1.4	6.7	0.027	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Geospatial	Precipitation	ppt.m08	0.01	16.79	124.75	36.0	10.1	3.1	0.7	5.5	0.017	Pass
Geospatial	Precipitation	ppt.m09	0.01	21.95	118.74	46.9	9.7	6.9	1.8	2.6	0.019	Pass
Geospatial	Precipitation	ppt.m10	0.02	19.04	109.8	29.1	5.3	7.0	3.5	0.0	0.012	Pass
Geospatial	Precipitation	ppt.m11	0.02	7.91	84.4	31.8	6.8	6.7	2.6	1.5	0.012	Pass
Geospatial	Precipitation	ppt.m12	0.02	5.65	67.46	18.7	4.3	5.7	2.6	0.4	0.014	Pass
Geospatial	Snow persistence	MeanSnowPersistence_01	0.01	0	52.76868	47.2	11.2	4.6	0.2	6.4	0.018	Pass
Geospatial	Snow persistence	MeanSnowPersistence_05	0.01	0.09638	51.06417	46.4	11.0	4.6	0.1	6.3	0.019	Pass
Geospatial	Snow persistence	MeanSnowPersistence_10	0.01	0.073745	51.58821	45.5	10.9	4.7	0.2	6.2	0.019	Pass
Geospatial	Temperature	temp.11121	0.01	-9.65333	15.75333	32.4	7.8	1.4	1.9	7.5	0.013	Pass
Geospatial	Temperature	temp.234	0.01	-4.26333	19.23333	31.5	7.9	2.0	1.3	7.2	0.014	Pass
Geospatial	Temperature	temp.567	0.01	14.92333	28.17333	22.3	6.6	2.8	0.1	5.4	0.010	Pass
Geospatial	Temperature	temp.8910	0.01	11.6	27.01333	23.0	6.8	2.5	0.3	5.8	0.012	Pass
Geospatial	Temperature	temp.m01	0.01	-14.56	13.73	39.2	8.9	1.7	1.9	8.0	0.021	Pass
Geospatial	Temperature	temp.m02	0.02	-11.69	15.84	36.8	8.6	1.9	1.7	7.7	0.021	Pass
Geospatial	Temperature	temp.m03	0.02	-4.76	19.13	31.9	8.0	2.0	1.3	7.2	0.012	Pass
Geospatial	Temperature	temp.m04	0.02	3.58	22.83	23.6	6.7	1.9	0.8	6.3	0.010	Pass
Geospatial	Temperature	temp.m05	0.02	10.59	26.23	19.9	6.0	1.9	0.6	5.8	0.010	Pass
Geospatial	Temperature	temp.m06	0.01	15.71	28.77	20.0	6.2	2.7	0.1	5.1	0.010	Pass
Geospatial	Temperature	temp.m07	0.02	18.31	29.52	29.3	7.9	4.2	0.9	5.1	0.012	Pass
Geospatial	Temperature	temp.m08	0.02	17.31	29.91	27.1	7.6	3.7	0.5	5.4	0.011	Pass
Geospatial	Temperature	temp.m09	0.02	12.3	27.52	22.0	6.6	2.4	0.3	5.6	0.013	Pass
Geospatial	Temperature	temp.m10	0.02	5.17	23.71	21.5	6.3	1.7	0.9	6.1	0.011	Pass
Geospatial	Temperature	temp.m11	0.01	-3.09	18.89	24.6	6.6	1.0	1.7	6.7	0.010	Pass
Geospatial	Temperature	temp.m12	0.01	-11.46	14.64	31.9	7.7	1.2	2.0	7.5	0.012	Pass
Geospatial	Temperature	tmax	0.02	9.42	28.7	41.5	9.4	2.9	1.0	7.6	0.018	Pass
Geospatial	Temperature	tmean	0.01	3.2	22.52	28.2	7.4	2.0	1.1	6.7	0.010	Pass
Geospatial	Temperature	tmin	0.02	-3.06	16.9	16.3	5.4	1.1	1.1	5.5	0.011	Pass
Geospatial	Watershed	DRNAREA_mi2	0.03	0.01	2991.6	13.5	6.4	4.1	3.6	2.4	0.020	Pass
Geospatial	Watershed	Elev_m	0.02	9	1956	34.7	7.5	2.4	2.3	4.7	0.013	Pass
Hydrological	Indirect	HydricSoils_score	0.73	0	3	179.4	16.3	9.1	0.3	9.2	0.004	Pass
Hydrological	Indirect	WoodyJams_number	0.87	0	100	3.4	1.8	2.2	3.7	1.3	0.000	Pass

4.4 Data stratification and preparation

Reaches in the Great Plains final dataset were divided into training (80%) and testing (20%) subsets balanced by streamflow duration class and subregion (Northern, Central,

Upper, Southern). Samples from the 80% of reaches were used to develop (or "train") the model and are referred to hereafter as the training dataset. Samples in the "test" dataset were withheld from method development, only being used to inform the final model selection and refinement, by evaluating the model on novel reaches.

Of the 268 reaches included in the final Great Plains analysis, each reach was visited between one and six times, yielding a total of 1069 samples. Figure 30 shows the distribution of repeat reach visits, with the majority of reaches being sampled six times.

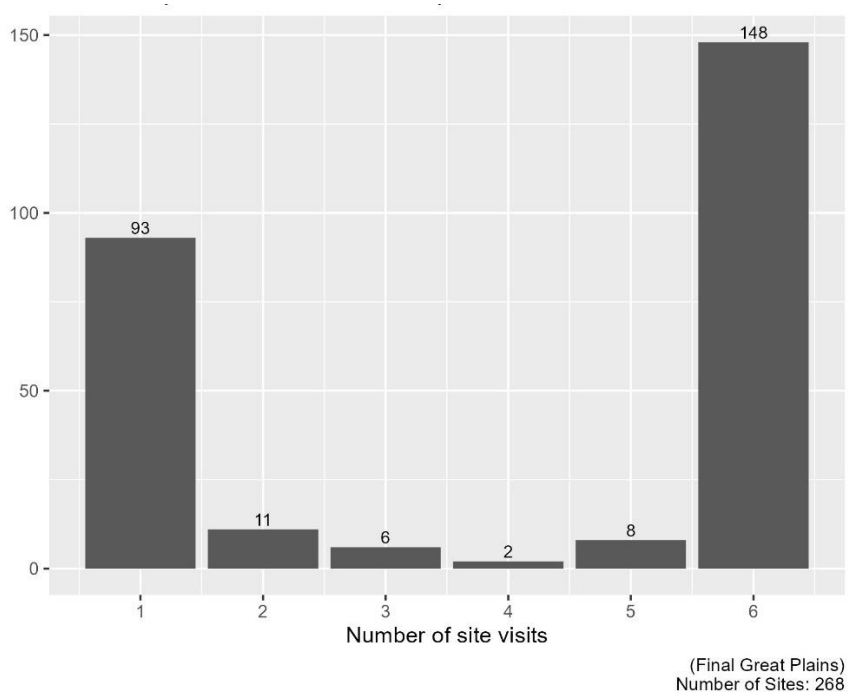


Figure 30. Distribution of number of visits across the 268 study reaches.

To minimize bias, oversampling was performed on the training dataset. The goal of rebalancing the dataset in this way was to get each reach as close as possible to six samples apiece, so that all reaches would be as equally represented as possible during model development. Oversampling is a common preprocessing step that serves to give under-represented classes more visibility in the data (Mohammed et al. 2020). Oversampling was performed on the training dataset only (no manipulations were conducted on the test dataset) and included the following steps:

- If a reach was sampled one time, each sample was repeated six times.
- If a reach was sampled twice, each sample was repeated three times.
- If a reach was sampled three times, each sample was repeated twice.
- If a reach was sampled four, five or six times, the samples were left as-is.

The result of the oversampling process was a more balanced training dataset with each study reach having as close to six samples as possible (Figure 31). The rebalanced (oversampled)

training dataset containing 1490 samples was used in the next step of preliminary method development.

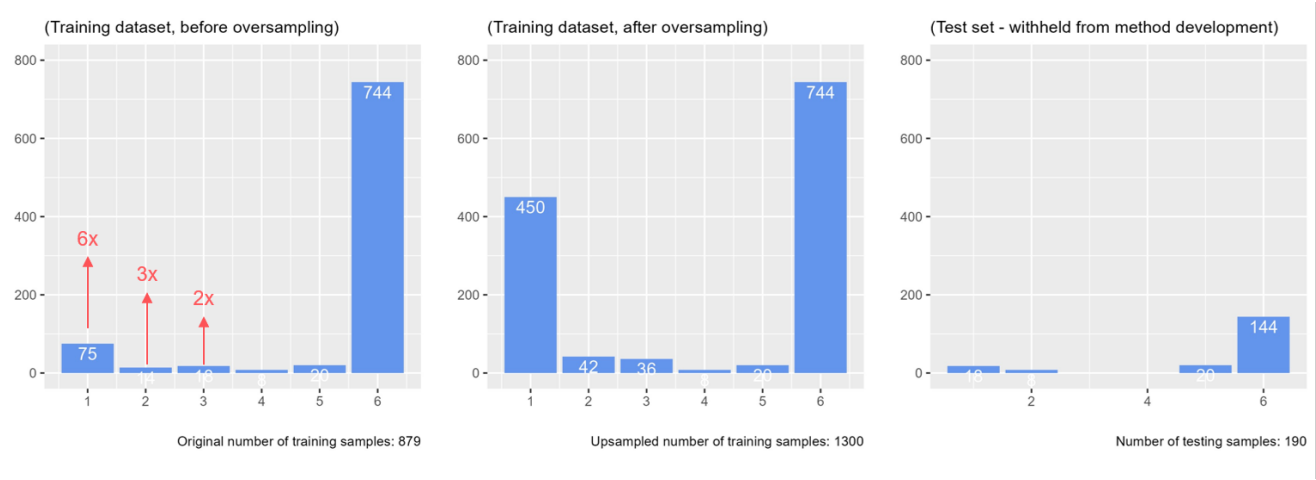


Figure 31. Distribution of number of reach visits before and after oversampling was performed. No rebalancing was performed on the test set.

4.5 Model development

Multiple approaches were attempted for developing preliminary models for the Great Plains region. Models that either entirely excluded or included only a limited number of geospatial metrics were considered (see Mazor et al. (2021b) for a discussion of the implications of including geospatial metrics in SDAMs). Additionally, model versions that were stratified by subregion (Southern Great Plains versus the Northern/Central/Upper Great Plains) compared to a single, unstratified Great Plains model were considered. The choice to separate between the Northern/Central/Upper Great Plains versus the Southern Great Plains was based off of performance during the beta method as well as the original ordinary high water mark (OHWM) publication that used a similar split (Wohl et al. 2016). See the beta Great Plains data supplement for more information on the stratification of the Great Plains into subregions (Eddy et al. 2022). The following preliminary models were constructed [brackets show abbreviations used in subsequent figures]:

- Unstratified: entire Great Plains. No geospatial metrics included. [(NoGIS) Unstrat]
- Unstratified: entire Great Plains. 2 geospatial metrics included [(2GIS)Unstrat]
- Stratified: Southern GP. No geospatial metrics included [(NoGIS) S]
- Stratified: Southern GP. 2 geospatial metrics included [(2GIS) S]
- Stratified: North, Upper, and Central GP. No geospatial metrics included [(NoGIS) UNC]
- Stratified: North, Upper, and Central GP. 2 geospatial metrics included [(2GIS) UNC]

The 6 preliminary models were compared to determine the degree of improved performance by the inclusion of GIS metrics and strata-specific models. Model design characteristics and optimal

number of metrics selected by RFE are shown in Table 37, and the selected metrics for each model are shown in Figure 32.

Table 37. Design characteristics of the preliminary models. GIS: included geospatial metrics. # samples: number of samples used in model training or testing. RFE OOB error rate: out-of-bag (OOB) error rate of the best model produced by recursive feature elimination.

Model set	Stratum	# samples (training)	# samples (testing)	# metrics eligible	# metrics chosen	RFE OOB error rate	Cohen's Kappa
Unstratified models							
Unstratified NoGIS	Entire Great Plains	1300	190	44	12	0.11	0.84
Unstratified 2GIS	Entire Great Plains	1300	190	84	10	0.07	0.90
Models stratified by region							
Stratified NoGIS	Upper, Central, Northern	1006	157	42	11	0.13	0.83
Stratified 2GIS	Upper, Central, Northern	1006	157	84	10	0.05	0.91
Stratified NoGIS	Southern	294	33	42	20	0.11	0.83
Stratified 2GIS	Southern	294	33	84	10	0.06	0.91

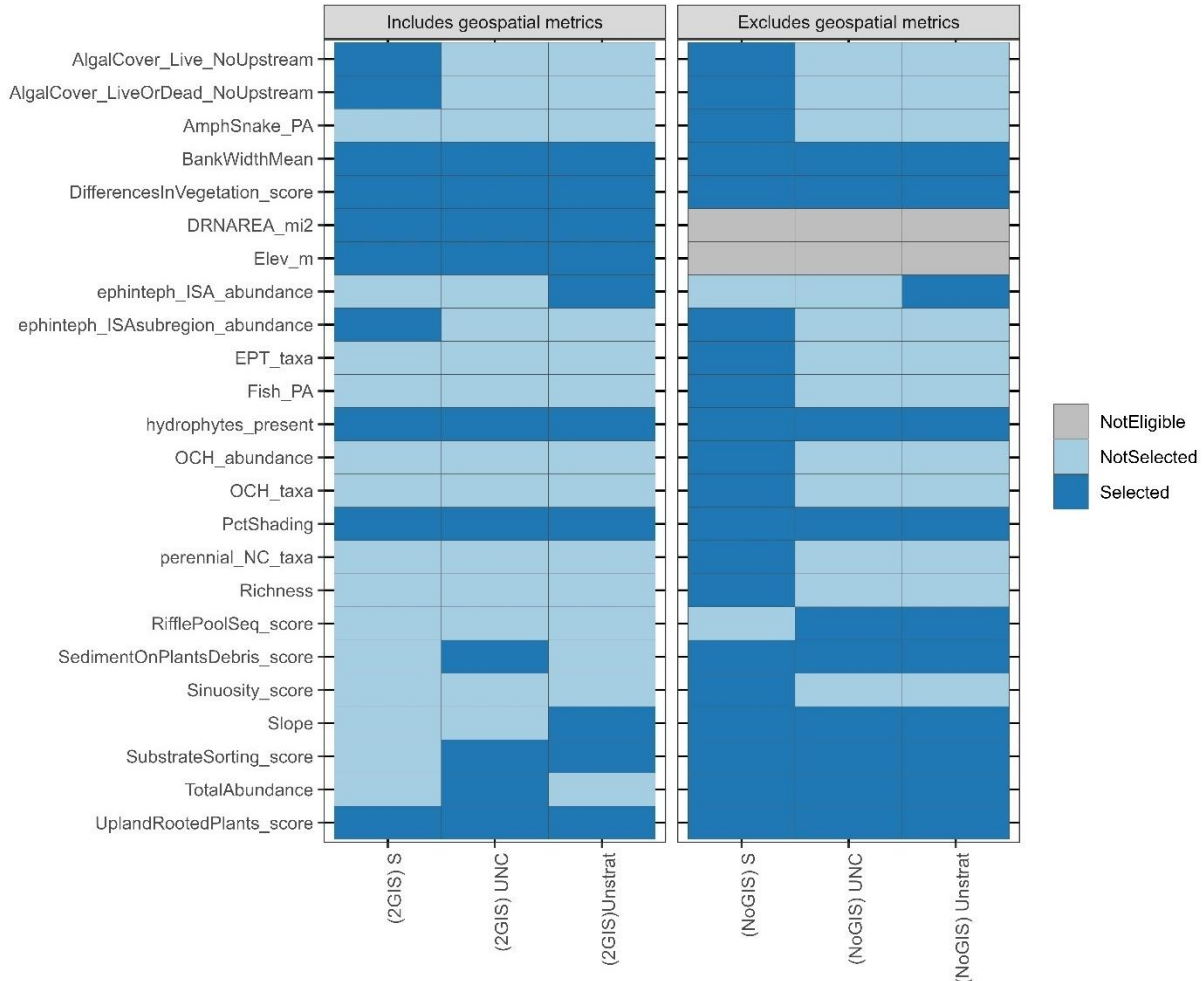


Figure 32. Metrics (left) selected by RFE for each model version (bottom). Grey tiles indicate that a screened metric was ineligible for selection in that model set (e.g., Elev_m was ineligible for models that did not allow GIS metrics).

Biological metrics, particularly those based on aquatic invertebrates, were among the most widely selected metrics across model sets (Figure 32). Among non-biological metrics, mean bankfull width was the only frequently selected geomorphological metric.

Accuracy of a model's ability to correctly distinguish among ephemeral, intermittent, and perennial streamflow classes was assessed on both the training and testing datasets independently. Training and testing measures were compared against each other to see if models validated poorly (training dataset accuracy substantially higher than testing dataset accuracy), which would suggest that models may be overfit for the training reaches and not generally predictive for streamflow duration classification. The performance of unstratified models was evaluated for individual strata by examining results for reaches within the four strata separately.

In general, the accuracies in the training dataset were highest for models including GIS metrics. However, performance of the models containing geospatial data sharply decreased when evaluated against the testing dataset, indicating that the GIS models were overfitting to the training dataset (Figure 33). The unstratified version of the models appeared to perform just as well or better on the testing sets than their unstratified counterparts. Because the stratified models did not show significant improvement (accuracy of training or testing datasets) over a single model encompassing the entire Great Plains that included a strata metric, separate models for each subregion were deemed unnecessary. Thus, the decision, which was affirmed by the RSC, was to use the unstratified model without GIS data. Furthermore, the strength of the unstratified (no GIS) model increases when looking at the ability of the model to accurately distinguish between ephemeral and at least intermittent (EvALI) compared to distinguishing between all three classes (PvIvE; Figure 33).

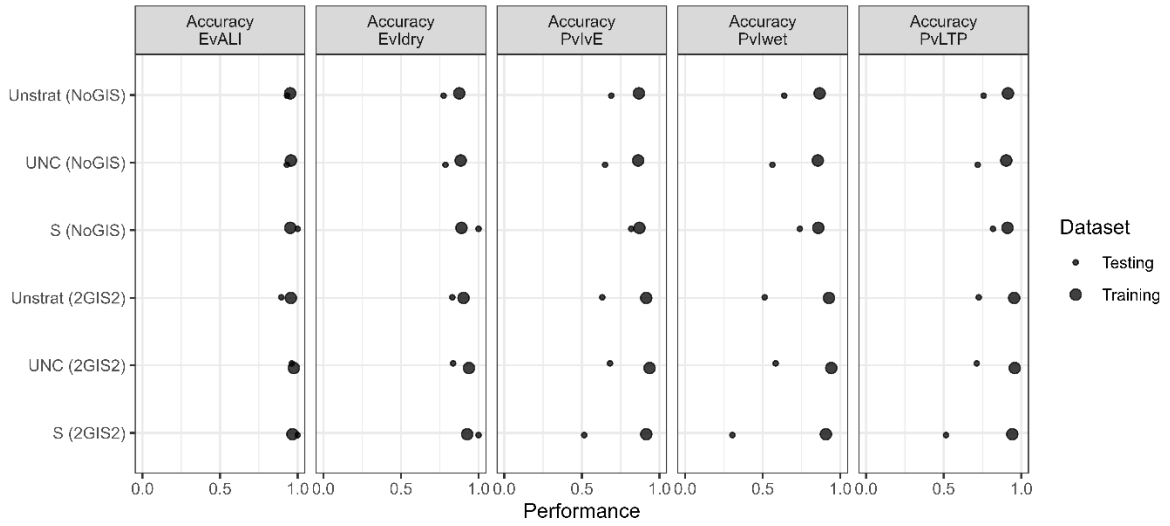


Figure 33. Accuracy of the preliminary GP models by strata

Based on the comparison of performance of the preliminary models, the unstratified NoGIS model was selected for further refinement. The unstratified (no GIS) model contained twelve metrics that were selected via RFE. The metrics are shown in Figure 34 by their order of importance. Here, importance to the random forest model is considered through mean decrease in accuracy.

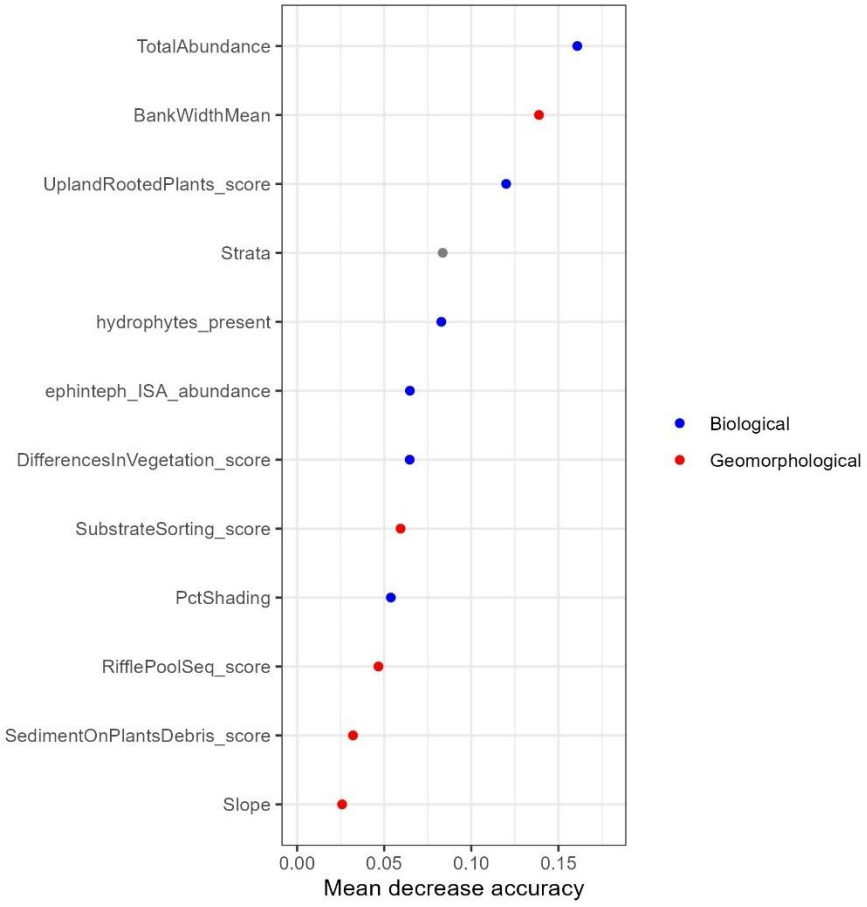


Figure 34. Metrics included in the unstratified (no GIS) model, by their order of importance. Mean Decrease in Accuracy is the relative loss in predictive performance when the particular variable is omitted from the model.

BankWidthMean was among the most important variables in the Great Plains model, an observation made in the GP beta and other SDAMs as well.

To evaluate the overall performance of the unstratified (no GIS) model, confusion matrices were created for both training and testing datasets (Table 38). No perennial-ephemeral misclassifications occurred in either testing or training datasets. The unstratified (no GIS) model had similar misclassification predictions of intermittent reach samples as ephemeral or perennial reaches in the testing and training datasets.

Table 38. Accuracy of the preliminary GP model (unstratified, no GIS) model. Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors."

GP training data

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	314	18	0
Intermittent	28	412	32
Perennial	0	52	402
At least intermittent	1	12	13
Less than perennial	9	6	1
Need more information	0	0	0

GP testing data

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	40	7	0
Intermittent	3	41	21
Perennial	0	13	50
At least intermittent	0	7	5
Less than perennial	2	1	0
Need more information	0	0	0

4.6 Model refinement

Upon selection of the unstratified (no GIS) model, the next step was to refine the selected model. The purpose of refining the preliminary model was to make the SDAM easier to implement in the field while improving (or at least not sacrificing) performance, by removing or simplifying the most difficult to measure metrics. Simplifying metrics was intended to improve efficient SDAM application and facilitate method use and transparency. Improving the efficiency of the SDAM application also contributes to ensuring that the SDAM can be applied during a single reach visit.

First, difficult-to-measure or potentially redundant metrics were attempted to be dropped or simplified (either by removing them or replacing them with an alternative metric). Some metrics were eliminated because they were closely related to another metric in the selected model (i.e., they described similar stream characteristics, such as rooted upland plants and hydrophyte presence). Metrics that were more time-consuming to measure were dropped if a simpler alternative was available and continuous metrics were converted to binary or ordinal metrics based on visual interpretation of their distributions. (Binary and ordinal metrics are typically more rapid to measure and easier to standardize than continuous metrics.) Accuracy

and repeatability measures were re-evaluated to ensure that overall model performance was not substantially diminished by the modifications.

The suite of metrics of the selected model was iteratively refined while monitoring model accuracy and repeatability. In each iteration, one or more metrics were either eliminated, binned, or otherwise simplified. The impact of each iterative refinement on performance was assessed, and the highest performing refined model was selected. Performance was assessed in terms of accuracy measures: PvIvE (i.e., proportion of reach samples classified corrected as perennial, intermittent, or ephemeral), EvALI (i.e., proportion of reach samples classified correctly as ephemeral or at least intermittent), and EvIdry (i.e. proportion of dry reach samples classified correctly as ephemeral or intermittent reaches, perennial and flowing reaches excluded).

Key steps of the model refinement phase are shown in Table 39, Table 40, and Figure 35. Performance according to three accuracy metrics on the testing dataset are shown compared to the original unstratified NoGIS model version (V0). Callout boxes indicate how the metrics were changed compared to model V0. For example, a refinement made between V0 and V1 was dropping Percent Shading and Strata, which indicated a negligible decrease in PvIvE accuracy and an improvement in EvALI measures. Dropping Percent Shading was attempted in order to reduce the amount of specialized equipment (i.e. a densiometer) required by field crews.

Other refinements explored binning metrics (especially hydrophyte and aquatic invertebrate metrics) to reduce the amount of effort required for data collection by field crews. New binned metrics introduced in the refinement phase included: hydrophytes_2 (0-1, 2+ hydrophytes), TotalAbund_0_10 (0, 1-9, or 10+ individuals), BankWidth_10 (<10, 10+ m), UplandRooted_PA (present/absent), ephISAabund_PA (present/absent).

After consultation with the RSC, the final GP model selected was the Version 5 refinement of the unstratified (no GIS) model.

Original (V0) NoGIS Unstrat

- BankWidthMean
- DifferencesInVegetation_score
- RifflePoolSeq_score
- SedimentOnPlantsDebris_score
- SubstrateSorting_score
- ephinteph_ISA_abundance
- hydrophytes_present
- TotalAbundance
- UplandRootedPlants_score
- PctShading
- Slope
- Strata

Refined (V5)

- BankWidthMean
- DifferencesInVegetation_score
- RifflePoolSeq_score
- SedimentOnPlantsDebris_score
- SubstrateSorting_score
- hydrophytes_2
- TotalAbund_0_10
- UplandRooted_PA

The Version 5 refinement differs from the unstratified (no GIS) model as follows:

- Number of hydrophytic species recorded ranged from 0 to eight species in the original dataset. This was simplified in the refined model into two discrete groups: fewer than two species and two or more species). This metric binning was based on visual interpretation of the metric distributions across streamflow duration classes and through trial-and-error testing.
- TotalAbundance, originally a count ranging from 0 to 102 in the original model., was binned into three discrete groups (equal to 0, 1-9, 10+). The choice of these groups was based on visual interpretation of the metric distributions across ephemeral, intermittent, and perennial classes, and through trial-and-error testing.
- UplandRootedPlants_score, originally codified as a score from 0-3 was simplified into a presence or absence metric.
- ephinteph_ISA_abundance, Strata, Percent Shading, and Slope were dropped from the model.

Table 39: Steps in refining the GP model. Step 5, marked with an asterisk (*) was the final selected refined model. Grey boxes identify indicators included in a refined model, B = binned modification, and S = simplified modification.

Step	Preceding step	DifferencesInVegetation_score	PctShading	BankWidthMean	UplandRootedPlants_score	Slope	Strata	hydrophytes_present	RiffflePoolSeq_score	ephinteph_ISA_abundance	TotalAbundance	SubstrateSorting_score	SedimentOnPlantsDebris_score	Modification
0	NA													Preliminary model
1	0				S									Dropped Strata and PctShading; Simplified UplandRootedPlants_score to presence / absence
2	1				S			B		S				Binned Hydrophytes to fewer than 2 and 2 or more species; Simplified ephinteph_ISA_abundance to presence / absence
3	2				S			B		S	B			Binned TotalAbundance at 0 and 10
4	3				S			B		S	B			Dropped slope
5*	4				S			B			B			Dropped ephinteph_ISA_abundance

Step	Preceding step	DifferencesInVegetation_score	PctShading	BankWidthMean	UplandRootedPlants_score	Slope	Strata	hydrophytes_present	RiffflePoolSeq_score	ephinteph_ISA_abundance	TotalAbundance	SubstrateSorting_score	SedimentOnPlantsDebris_score	Modification
6	5				S			B		S				Added back in Simplified ephinteph_ISA_abundance as presence / absence; Dropped TotalAbundance
7	5			B	S			B			B			Binned BankWidthMean at 10

Table 40. Training and testing accuracy of each step in the refinement of the GP. Step 5, marked with an asterisk (*) was the final selected refined model. Statistics are calculated for the entire GP region.

Step	Preceding step	EvALI testing	EvIdry testing	PvIvE testing
0	None	0.93	0.77	0.69
1	0	0.82	0.70	0.64
2	1	0.81	0.69	0.61
3	2	0.81	0.69	0.60
4	3	0.80	0.67	0.62
5*	4	0.82	0.71	0.65
6	5	0.80	0.68	0.64
7	5	0.82	0.71	0.67

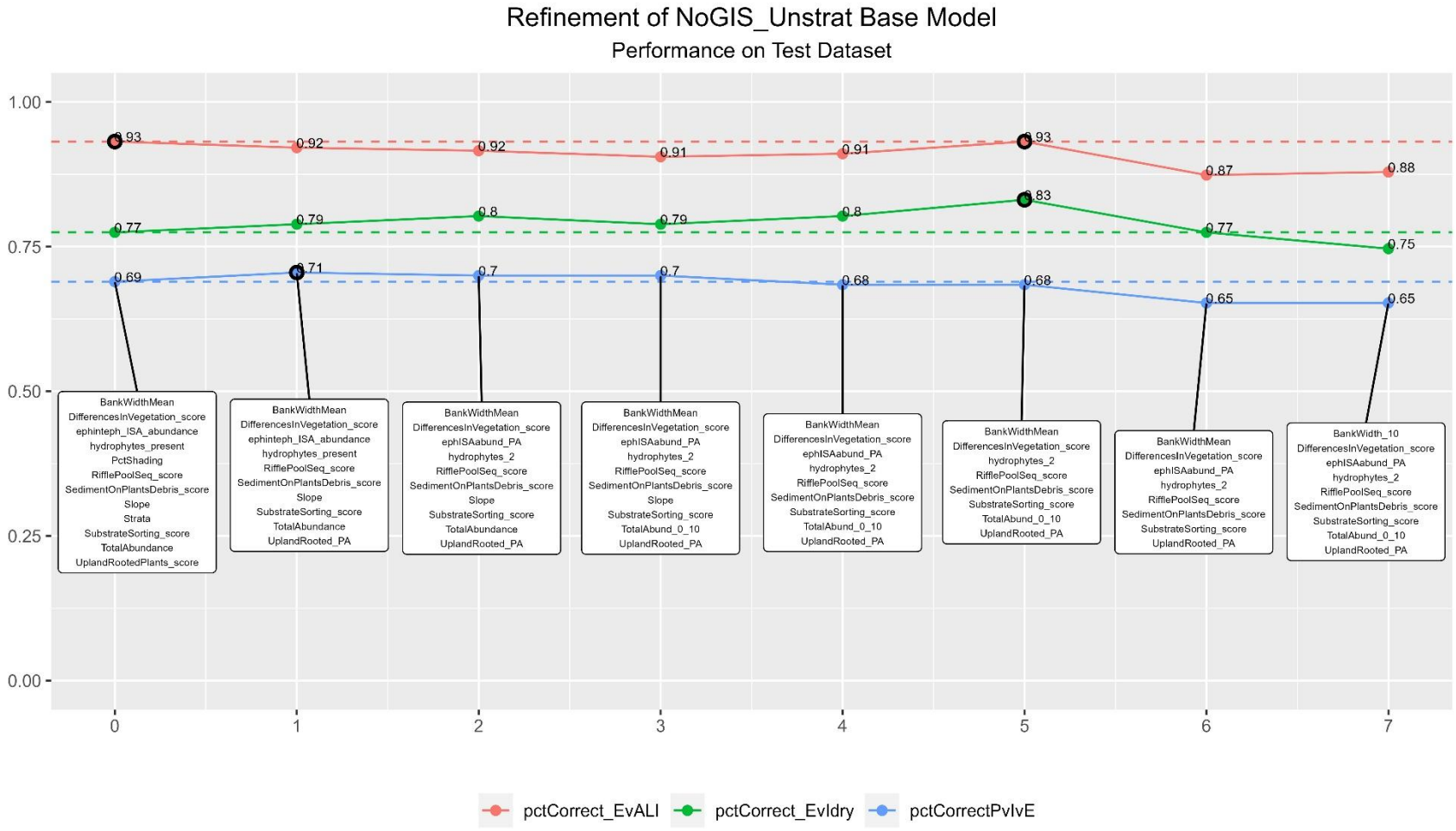


Figure 35. Impact of refinement of metric set on the model performance relative to the unstratified (no GIS) GP model using the testing dataset. Each refinement description is relative to the description at 0 (unstratified, no GIS model). Black circles indicate the highest accuracy scores. Dashed lines show performance of the original unstratified (no GIS) model.

4.6.1 Single indicators

Single indicators can supersede a model classification of *ephemeral* to make it change to *at least intermittent*. Single indicators provide technical benefits (i.e., improved EvALI accuracy) as well as non-technical benefits, such as greater acceptance of the SDAM, given public understanding of the role of streamflow duration in supporting biological organisms and rapidity of determining a flow classification. Single indicators are also used in other SDAMs (e.g., Nadeau et al. 2015, Dorney and Russell 2018, Mazor et al. 2021a); for instance, indicators can include the presence of fish, iron-oxidizing bacteria, hydric soils, and/or aquatic vertebrates (amphibians and reptiles), among others.

Within the GP, all single indicators (Table 14) had minimal impact on performance or introduced more errors than they corrected. The impacts of including each single indicator in the final Great Plains model (V5 refinement) are summarized in Figure 36. Based on these results, the RSC did not recommend including any of the evaluated single indicators in the final SDAM GP.

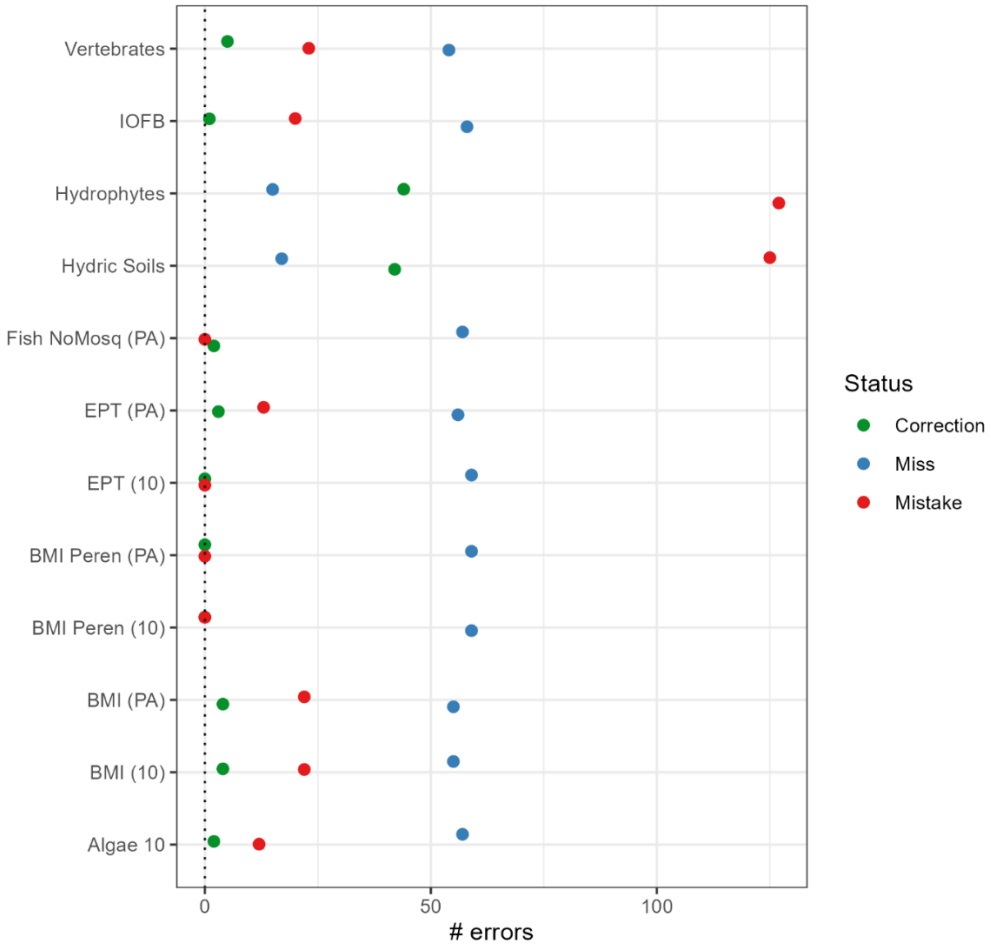


Figure 36. Impact of including single indicators in the final GP model. Green dots represent the number of corrections, which were generally fewer than the number of mistakes introduced (red dots). Missed opportunities (blue dots) indicate cases where the single indicator could have helped improve performance but did not.

4.7 Performance assessment of final models

The performance of the final Great Plains model when evaluated with the test data was an overall PvIvE accuracy of 0.68. EvALI accuracy was significantly higher, with a final testing accuracy of 0.93. PvLTP testing accuracy was also higher than PvIvE (0.75 in the testing dataset). No samples in either the training or testing datasets resulted in a need more information (Table 41). Only 11 visits across training and testing resulted in ALI or LTP outcomes (Table 41, Table 42). Ephemeral streams were assessed 13 times during flowing conditions, and perennial streams were assessed 4 times during non-flowing conditions in the GP (Table 42).

Table 41: V5 final model performance. Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors".

Region	Set	Predicted class	E	I	P
GP	Training	E	347	18	0
GP	Training	I	1	454	5
GP	Training	P	0	27	441
GP	Training	ALI	1	0	2
GP	Training	LTP	3	1	0
GP	Training	NMI	0	0	0
GP	Testing	E	43	10	1
GP	Testing	I	2	38	23
GP	Testing	P	0	20	49
GP	Testing	ALI	0	1	3
GP	Testing	LTP	0	0	0
GP	Testing	NMI	0	0	0

Table 42: Comparison of predicted and true streamflow duration classes for the Great Plains final model, with flowing and not flowing reach visits presented separately Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors."

Region	Set	Predicted class	E (flowing)	E (not flowing)	I (flowing)	I (not flowing)	P (flowing)	P (not flowing)
GP	Training	E	12	335	1	17	0	0
GP	Training	I	0	1	347	107	4	1
GP	Training	P	0	0	25	2	438	3
GP	Training	ALI	1	0	0	0	2	0
GP	Training	LTP	0	3	1	0	0	0
GP	Training	NMI	0	0	0	0	0	0
GP	Testing	E	0	43	1	9	1	0
GP	Testing	I	0	2	22	16	23	0
GP	Testing	P	0	0	19	1	49	0
GP	Testing	ALI	0	0	1	0	3	0
GP	Testing	LTP	0	0	0	0	0	0
GP	Testing	NMI	0	0	0	0	0	0

There was one perennial reach in the testing dataset where the model misclassified a reach visit as ephemeral. This reach in Wisconsin was visited 6 times, indicator data from all but one visit were classified by the final Great Plains model as intermittent. The one visit (in September 2021) when the reach was classified as ephemeral was distinct in that surface flow was discontinuous and no aquatic invertebrates were observed. Although some potential single indicators were observed at this reach visit (specifically, frog vocalizations, iron-oxidizing fungi or bacteria, and hydrophytes), they were not selected for use in the final GP SDAM because they reduced the overall accuracy of the method when other reaches were considered (Figure 36). No fish were observed. Performance of the final GP model (v5) by subregion is displayed in Figure 37.

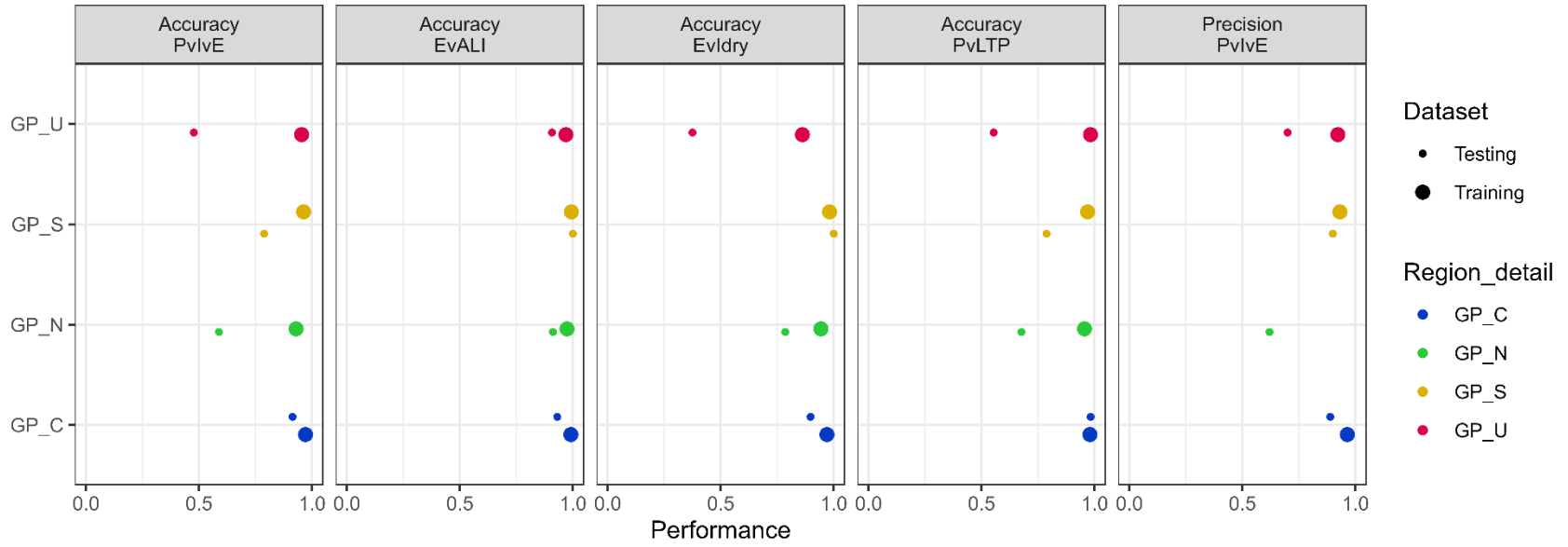


Figure 37. Performance (classification accuracy) of the final Great Plains model (v5) across four subregions of the Great Plains (GP_C = Central Prairies, GP_N = Northern Prairies, GP_S = Southern Plains, and GP_U = Upper Midwest). Large dots indicate performance on the training dataset; small dots represent model performance on the testing dataset.

4.7.1 Accuracy at disturbed reaches

Disturbances were noted at 422 (39%) reach visits in the GP. Urban or developed land use represented 49% of these reach visits, and 51% were associated with agricultural land use. Accuracy was comparable between disturbed and undisturbed reach visits (Table 43).

Table 43. Comparison of accuracy at disturbed and undisturbed reach visits. Training performance does not include oversampled data.

Set	Measure	# disturbed reaches	# disturbed visits	Disturbed accuracy	# undisturbed reaches	# undisturbed visits	Undisturbed accuracy	Difference in accuracy (Disturbed minus undisturbed)
Testing	EvALI	27	64	0.92	39	126	0.94	-0.02
Testing	PvLTP	27	64	0.86	39	126	0.68	0.18
Testing	PvIvE	27	64	0.80	39	126	0.63	0.17
Testing	EvIdry	13	24	0.79	18	47	0.94	-0.15
Testing	PvIwet	18	40	1.00	28	79	0.92	-0.08
Training	EvALI	147	357	0.98	151	521	0.97	0.01
Training	PvLTP	147	357	0.97	151	521	0.95	0.02
Training	PvIvE	147	357	0.95	151	521	0.93	0.02
Training	EvIdry	67	125	0.85	76	212	0.93	0.02
Training	PvIwet	99	226	0.99	104	304	0.99	0.00

4.7.2 Sensitivity to variation in indicator values

Thresholds for small versus large changes were determined for each indicator included in the final GP model (Table 44).

Table 44. Threshold for large and small sensitivity.

Indicator	Threshold for large decreases	Threshold for large increases
BankWidthMean	<50% of original value	>150% of original value
DifferencesInVegetation_score	-1.25 from original value	+1.25 from original value
RifflePoolSeq_score	-1.25 from original value	+1.25 from original value
SubstrateSorting_score	-1.25 from original value	+1.25 from original value
SedimentOnPlantsDebris_score	-1.25 from original value	+1.25 from original value
TotalAbund_0_10	-1.5 bins from original value	+1.5 bins from original value
hydrophytes _2	-1.5 bins from original value	+1.5 bins from original value
UplandRooted_PA	-1.5 bins from original value	+1.5 bins from original value

The GP SDAM has high sensitivity to 3 indicators, high sensitivity to one indicator, moderate sensitivity to 3 indicators and low sensitivity to one indicator (Table 45, Figure 38).

Table 45. Sensitivity of indicators in the final GP SDAM.

Indicator	Large decrease	Small decrease	Small increase	Large increase	Sensitivity
BankWidthMean	High	High	Low	Moderate	High
DifferencesInVegetation_score	Moderate	Low	Low	Moderate	Moderate
RifflePoolSeq_score	Low	Low	Low	Low	Low
SubstrateSorting_score	Moderate	Low	Low	Moderate	Moderate
SedimentOnPlantsDebris_score	Moderate	Low	Low	Moderate	Moderate
TotalAbund_0_10	High	Very High	Very High	High	Very High
hydrophytes _2	N/A	Very High	Very High	N/A	Very High
UplandRooted_PA	N/A	Very High	Very High	N/A	Very High

The following figures show the relative sensitivity of each indicator. In each figure, the solid vertical lines indicate no change in the indicator value. Dashed vertical lines delineate the threshold between large and small changes in indicator value. Horizontal dashed lines indicate thresholds between small, moderate, and large impacts on the percent of reach visits with unchanged classifications.

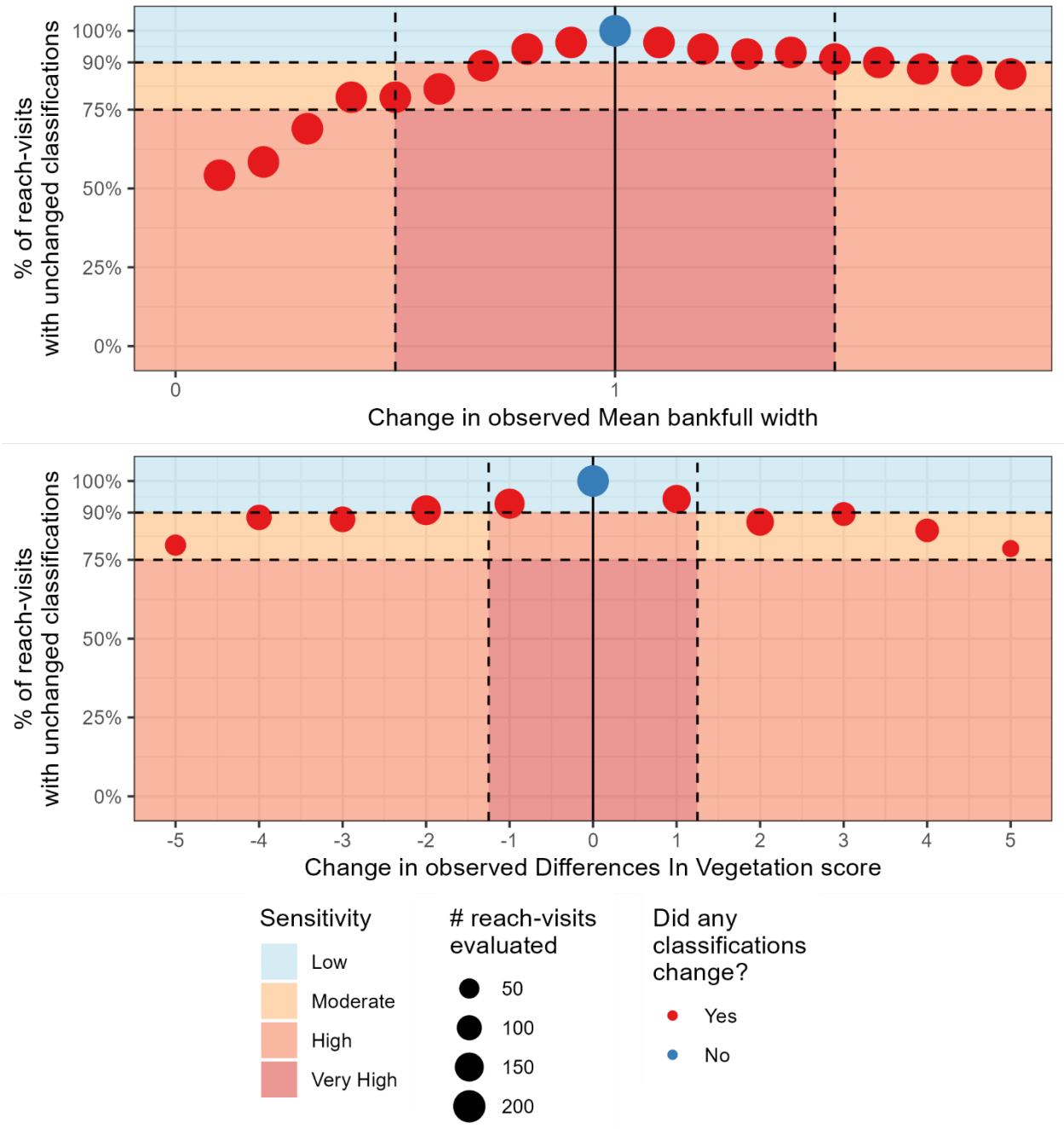


Figure 38a. Sensitivity of the GP SDAM to changes in indicator values for mean bankfull channel width (top) or difference in vegetation score (bottom). The x-axis represents the factor by which the original indicator value was multiplied (top) or change in score (bottom), and the y-axis represents the percent of reach visits with unchanged classifications with the changed indicator value (all other indicator values held constant).

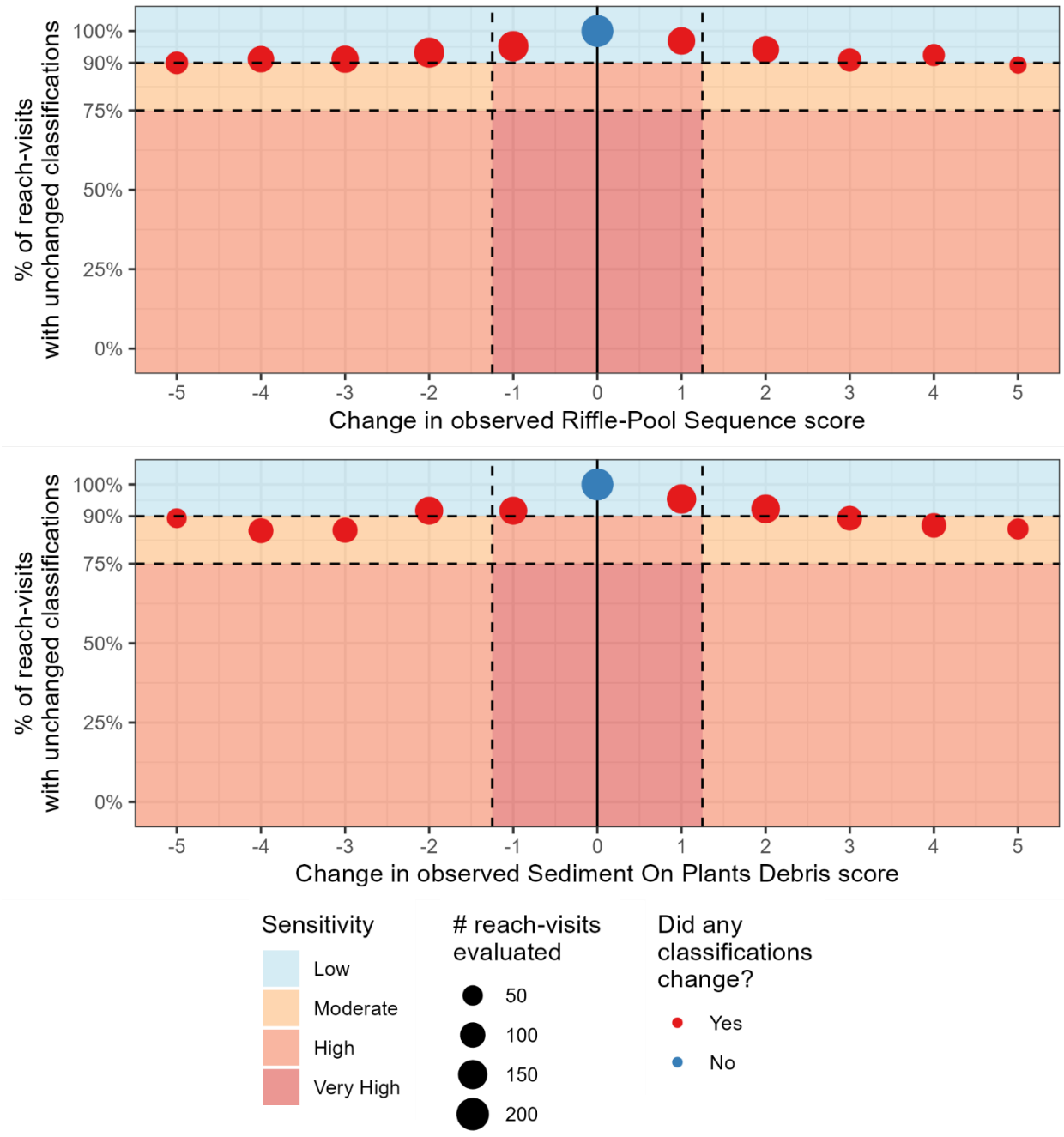


Figure 38b. Sensitivity of the GP SDAM to changes in indicator values for the riffle-pool sequence score (top) or sediment on plants and debris score (bottom).

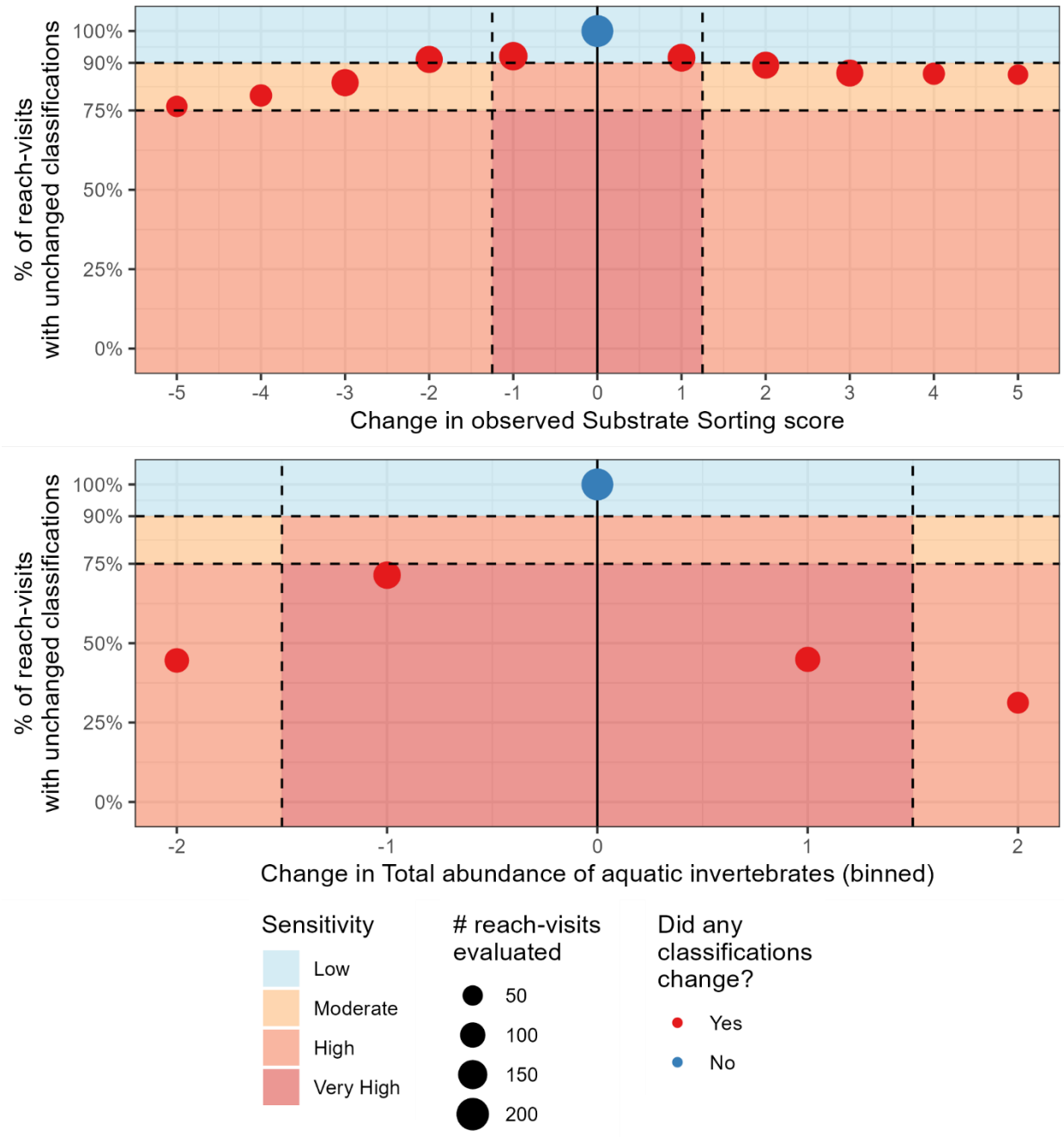


Figure 38c. Sensitivity of the GP SDAM to changes in indicator values for the the substrate sorting score (top) or the total abundance of aquatic invertebrates (bottom).

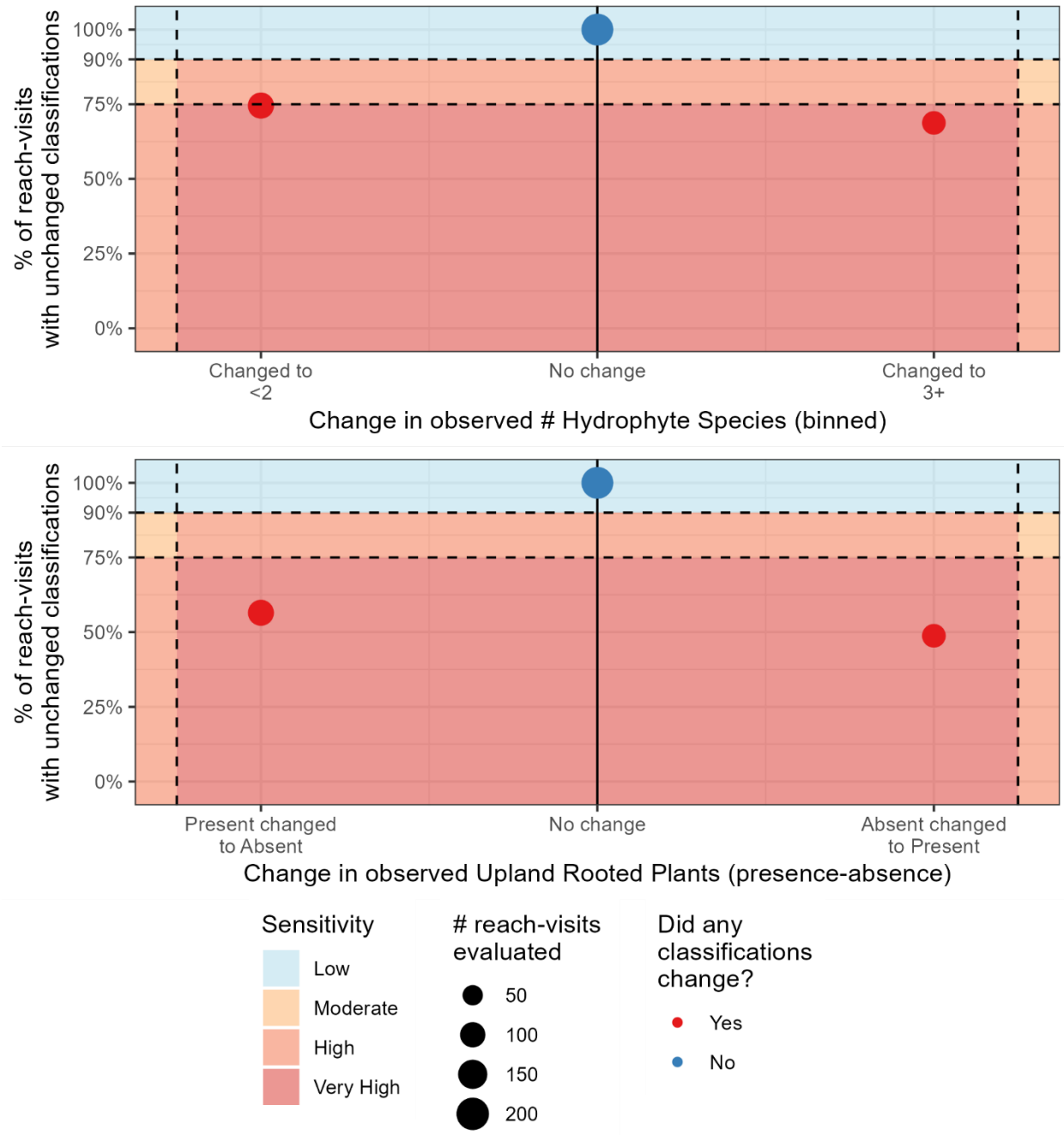


Figure 38d. Sensitivity of the GP SDAM to changes in indicator values for the number of hydropthyes (top) or upland rooted plants (bottom). Both of these indicators are scored in a binary manner.

4.7.3 Sensitivity to time of sampling

Seasonal patterns were relatively consistent across subregions of the GP based on reach-specific long-term monthly climatic statistics shown in Appendix B: Indicators and metrics evaluated for SDAM development (Figure 39). Across all subregions, the warmest months typically occurred in the summer (May to October). Within the Central Prairie, Upper Midwest, and Northern Prairie subregions, precipitation was highest in late Spring (typically May or June), whereas the Southern Great Plains showed a bimodal pattern, with highest precipitation in both Spring and

Fall. Based on these patterns, a consistent seasonal classification was applied to all reaches in the GP (Summer = May-Oct, Winter = Nov-Apr).

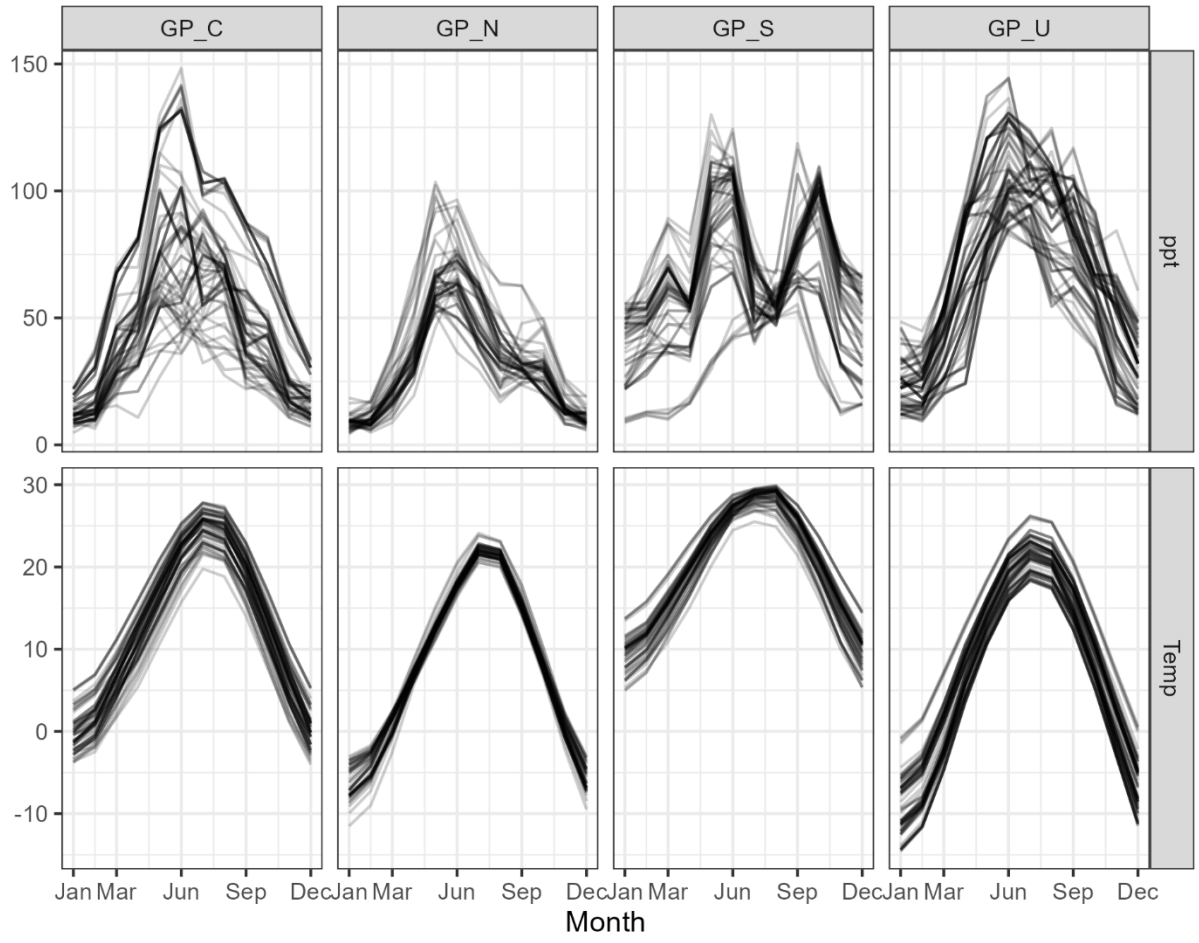


Figure 39. Long-term monthly averages of precipitation (ppt, mm) and temperature (Temp, °C) for Great Plains reaches by subregion (GP_C = Central Prairie, GP_N = Northern Prairie, GP_S = Southern Great Plains, and GP_U = Upper Midwest) to support approximate "summer" and "winter" delineations (Summer = May-Oct, Winter = Nov-Apr).

There were more reach visits during the Summer than Winter (Table 46). The number of GP training reach visits was similar between dry (55%) and wet (45%) seasons, but 55%, 31%, and 14% of reach visits occurred during Normal, Drier than Normal, and Wetter than Normal precipitation conditions, respectively (Table 46). The GP testing dataset was divided into 58% dry and 42% wet season reach visits. The distribution of the GP testing dataset was comparable to that of the training dataset, with 58% Normal, 26% Drier than Normal, and 15% Wetter than Normal. The inferential power for making accuracy comparisons, particularly between APC classes, is limited by the large differences in number of reach visits, especially for the testing sets.

Table 46. Number of reach visits in each season-based temperature and precipitation, WebWIMP (WW), and Antecedent Precipitation Condition (APC) for the Great Plains training and testing datasets.

Time of sampling comparisons	Classes	Data set	
		Train	Test
Weather-based seasons (WBS)	Summer (May – Oct)	871	111
	Winter (Nov – Apr)	429	79
WebWIMP season (WW)	Dry	714	110
	Wet	586	80
Antecedent Precipitation Condition (APC)	Normal	710	112
	Drier than normal	408	50
	Wetter than normal	182	28

Northern Prairie and Southern Great Plains subregions of the GP experienced the greatest variation of PvIvE accuracy between summer vs winter. Less variation was observed for EvALI accuracy. Classification accuracy was comparable between training and testing datasets for the EvALI comparison but was higher for the training than testing dataset for the PvIvE and PvLTP comparisons (Figure 40 and Figure 41). Based on the testing datasets, the GP SDAM performed slightly better during Winter reach visits compared to the Summer reach visits (Figure 40). The EvALI accuracy (training and testing datasets) of the GP SDAM was comparable between WW and APC but was higher for PvIvE and PvLTP in the testing dataset during the Dry than Wet season and lowest for reach visits during Wetter than normal precipitation (Figure 41). Thus, accuracy measures are considered relatively robust against seasonality changes.

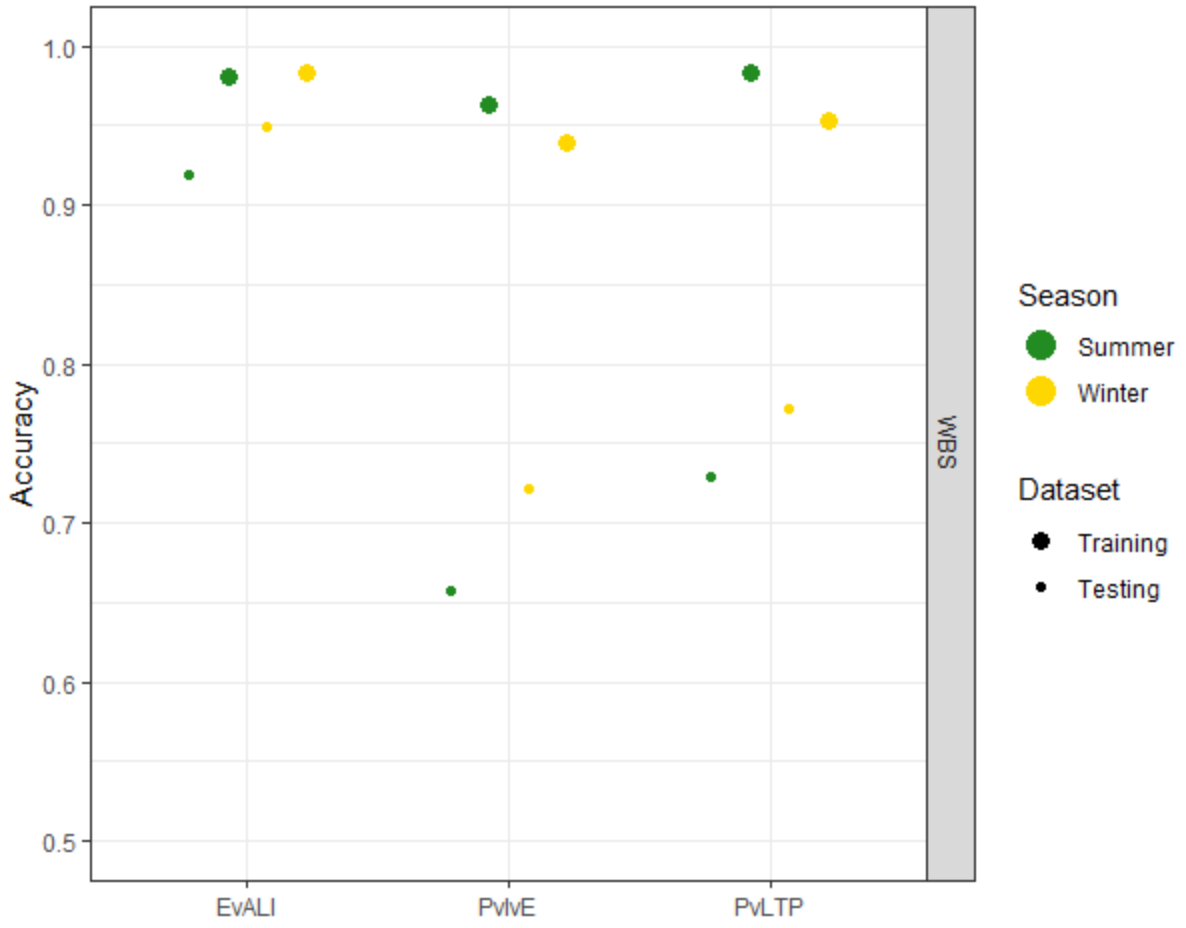


Figure 40. Accuracy of the final GP SDAM in weather-based seasons (WBS; Summer = May to October and Winter = November to April).

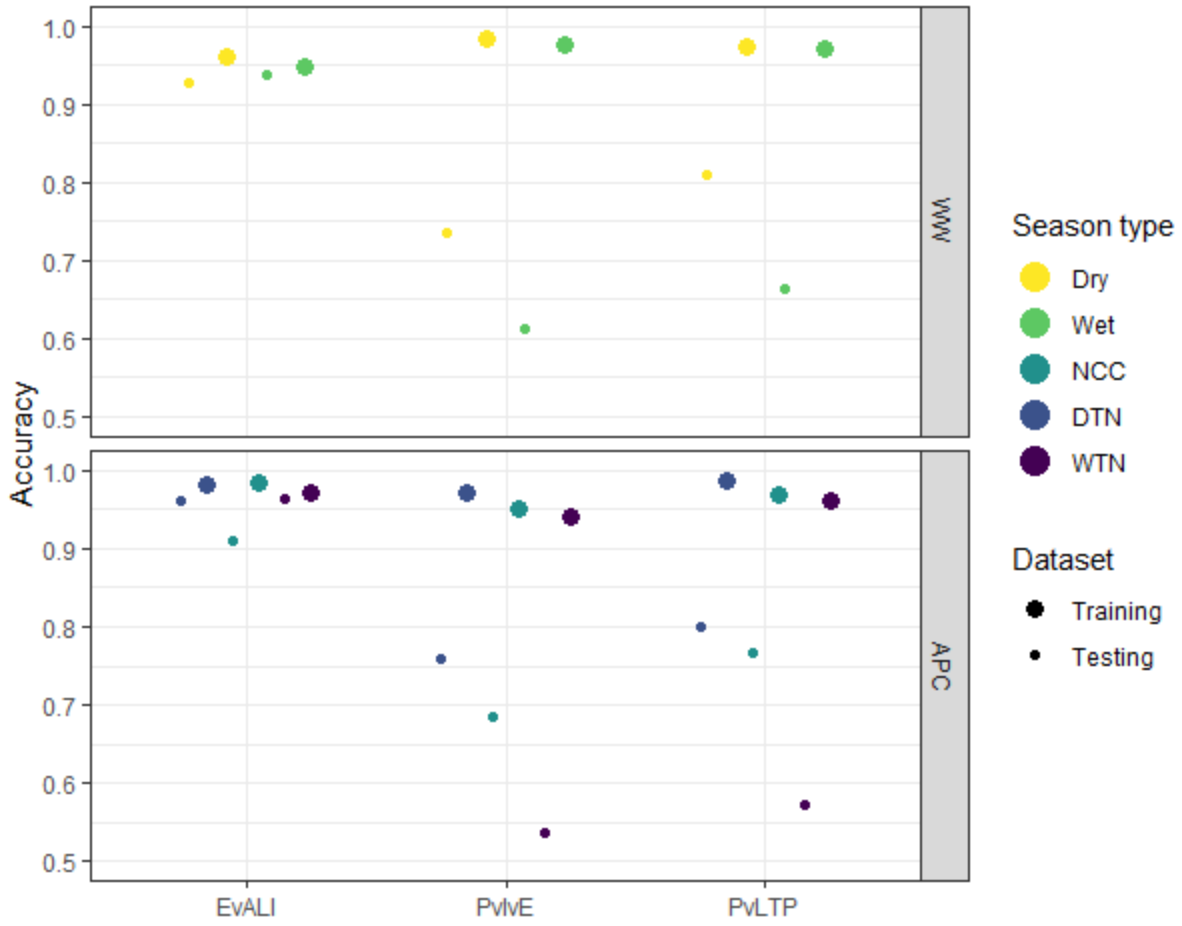


Figure 41. Accuracy of the Great Plains (GP) SDAM in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) estimations for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN).

Because the GP SDAM had high classification accuracy for the training dataset, there were very little error for characterizing directionality by WW or APC. For the testing dataset, there were more misclassifications predicting short streamflow duration classes during the Wet season than during the Dry season (Figure 41). The PvIvE and PvLTP misclassifications by the GP SDAM tended to be more balanced in their directionality during Normal precipitation than during Drier or Wetter than Normal precipitation, but EvALI misclassifications were comparably balanced among APC. Overall, there was a greater tendency for the GP SDAM to predict shorter flow duration classes than to predict longer flow duration classes, regardless of APC (Figure 42).

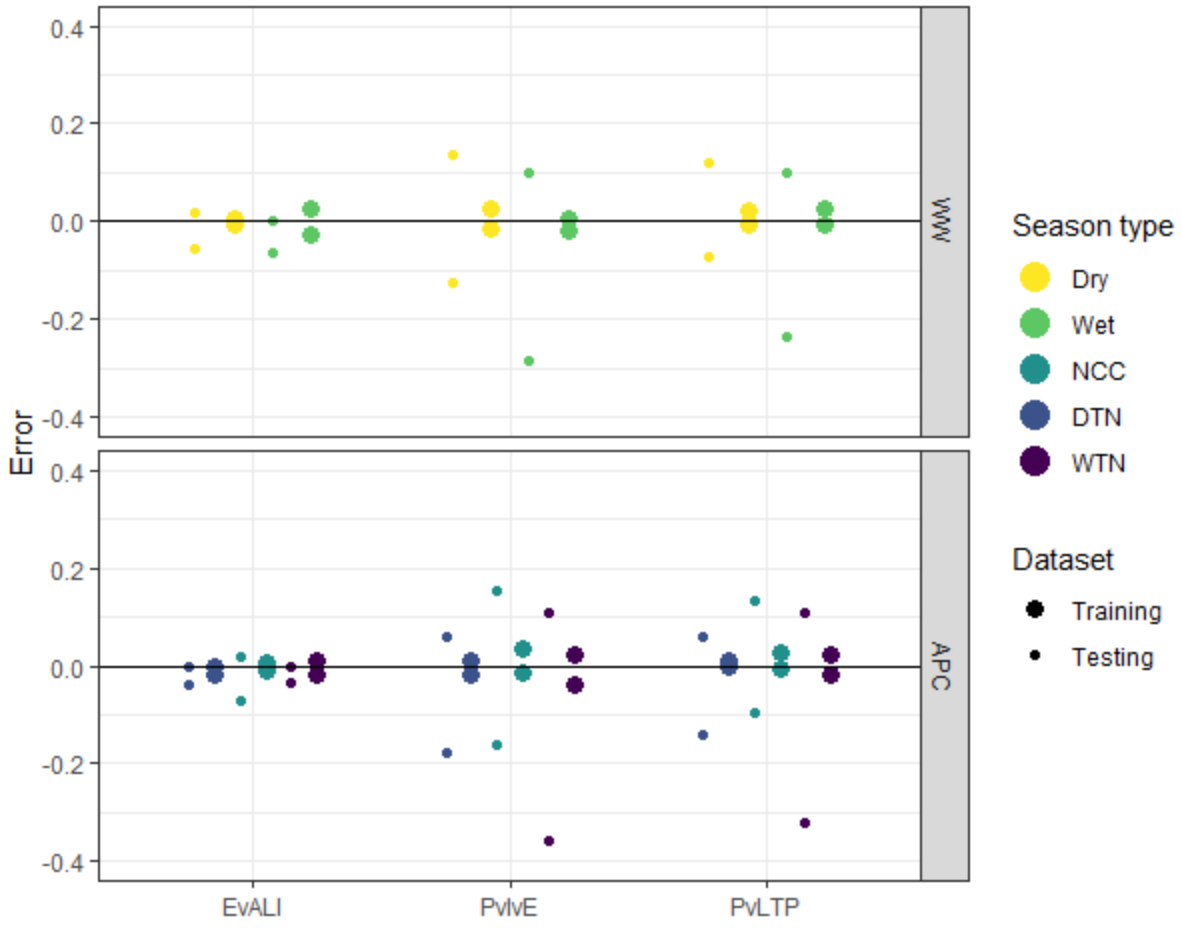


Figure 42. Error by direction of the GP SDAM in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) estimations for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN). Negative values represent the proportion of error predicting shorter flow duration classifications and positive values represent the proportion of error predicting longer flow duration classifications.

4.7.4 Comparison to other SDAMs

4.7.4.1 Comparison to other SDAMs used within the Great Plains or adjacent regions

In Figure 43, performance of the final GP SDAM is compared to adjacent SDAMs using the same final GP testing dataset. The final GP method was compared to Phase 1 of the NM hydrology protocol (NMED 2011) as well as the final AW and WM SDAMs. The final GP SDAM performed better than adjacent SDAMs on the testing data than other methods (Figure 43, Table 47).

Because the SDAMs for the NE and SE require indicators that were not measured the same way at reaches within the GP, those methods were not evaluated within the GP region.

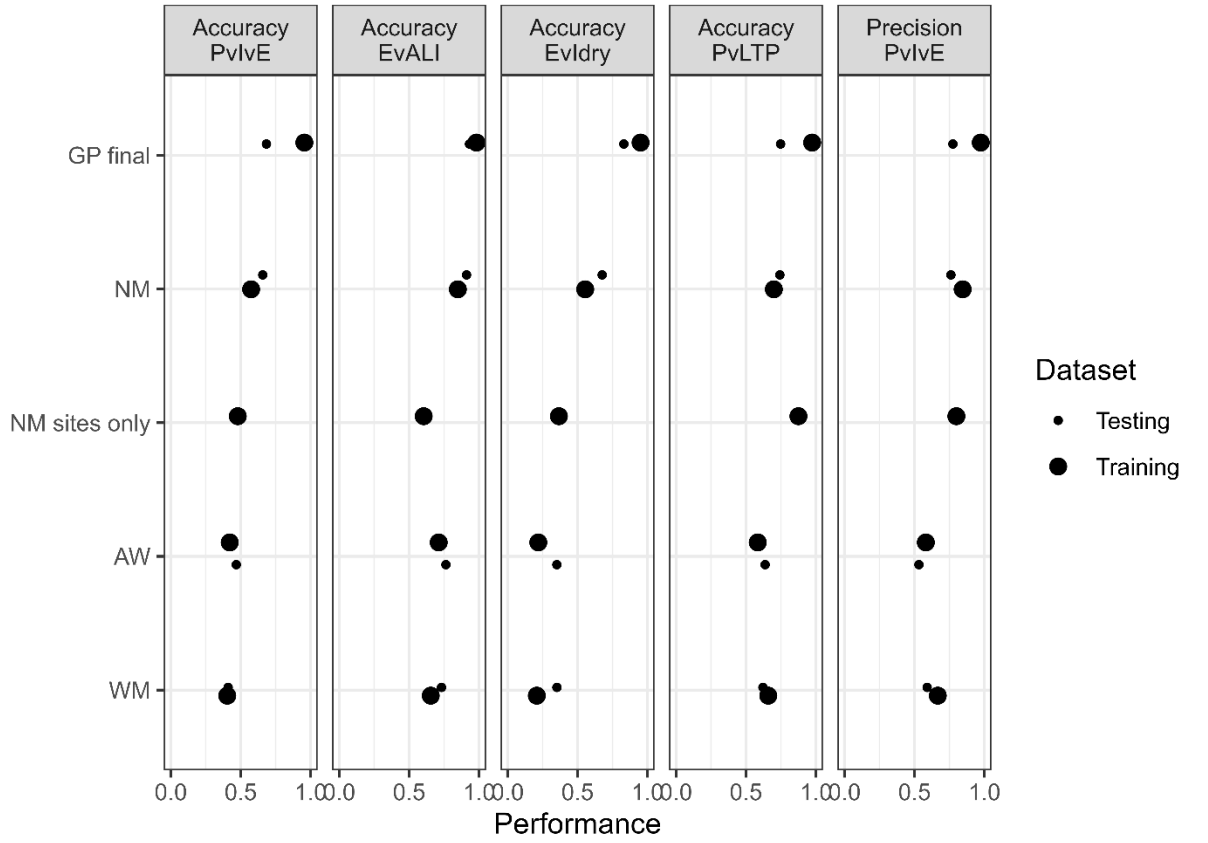


Figure 43. Comparison of performance of the final GP SDAM to other SDAMs.

Table 47. Performance of the final GP SDAM as a whole, and by subregion, as compared to other SDAMs.

SDAM	PvIvE % Accuracy (Test)	EvALI % Accuracy (Test)	EvALI dry % Accuracy (Test)	PvIvE % Precision (Test)	EvALI % Precision (Test)
GP final	68.4%	93.2%	83.1%	77.5%	91.6%
Upper	47.7% (31/65)	90.8% (59/65)	37.5% (3/8)	70.0%	90.5%
Northern	58.8% (20/34)	91.2% (31/34)	78.6% (11/14)	62.0%	88.0%
Southern	78.8% (26/33)	100.0% (33/33)	100.0% (10/10)	90.0%	100.0%
Central	91.4% (53/58)	93.1% (54/58)	89.7% (35/39)	88.9%	91.1%
GP beta	63.2%	83.7%	76.1%	65.2%	82.0%
NM	65.8%	91.1%	67.6%	76.1%	91.1%
AW	55.3%	86.8%	78.9%	65.7%	84.3%
WM	53.7%	83.7%	73.2%	67.5%	81.1%

4.7.4.2 Comparison of the GP SDAM and NM hydrology protocol within New Mexico

Because data were collected at only 8 reaches (none of which were in the testing set) within the GP portion of the state of New Mexico, we did not directly compare the performance of the GP SDAM and NM hydrology protocol within New Mexico.

4.7.4.3 Comparison to beta GP SDAM

The beta GP SDAM had PvIvE accuracy of 0.62 and an EvALI accuracy of 0.87 (Eddy et al. 2022), compared to PvIvE accuracy of 0.68 and EvALI accuracy of 0.93 for the final GP SDAM. PvLTP accuracy was not reported for the beta GP method. Performance statistics for the beta methods were calculated from different testing datasets compared to the final SDAMs, which could affect the comparison. The beta PvIvE accuracies reported (in the National Data Supplement) have been adjusted from the original values reported in 2022 beta data supplement report (Eddy et al 2022 and November 2023 addendum) to be calculated consistent with descriptions in section 2.8.4.1 Accuracy.[1] In addition, the beta GP method was developed from a smaller dataset both in number of reaches (17 fewer) and samples (528 fewer, prior to oversampling which in the beta was targeting 4 and for the final was targeting 6 samples per reach).

[1] In development of the beta methods ALI was counted as correct for PvIvE accuracy (treating ALI the same as a I result), however for the final methods ALI was not counted as correct for PvIvE accuracy.

5. Development of the NE and SE SDAMs

The specific data analysis steps described in this document follow the approach used to develop and evaluate the final SDAM for the Northeast and Southeast regions. U.S. Caribbean data were not included in final method development.

5.1 Study area

The NE and SE SDAM regions (based on Wohl et al. 2016) include states along the Atlantic and Gulf Coasts, extending west to the southeast corner of Kansas and north through Michigan to Maine. The NE includes all or part of 24 states and is considered those areas dominated by forest-type vegetation where snowmelt contributes at least some flow to streams and rivers during the year. Average yearly precipitation ranges widely across the region, but most areas receive between 40 and 50 inches of precipitation per year, on average. The SE includes all or part of 11 states and is considered those areas characterized by forest-type vegetation that are generally dominated by diverse types of rainfall runoff rather than snowmelt, including tropical storms and hurricanes. Average yearly precipitation also ranges widely across the SE region, but most areas receive between 50 and 60 inches of precipitation per year, on average. Ephemeral and intermittent reaches may be found at any position within a watershed but are more common in smaller headwaters, where flow accumulation is insufficient to sustain longer-duration flows (Fritz et al. 2008). Ephemeral and intermittent reaches may also be more common along the western boundary of the SE region and more southern parts of the NE, where average yearly precipitation totals are lowest (40 inches or less), and evapotranspiration is relatively high (Hammond et al. 2021).

The NE and SE regions as defined above include many metropolitan areas, including the New York City area (largest by population), Houston, Philadelphia, Baltimore-Washington D.C., Miami, and Atlanta, as well as some of the nation's fastest growing cities, such as Orlando, Raleigh-Durham, and Charlotte. Thus, there are places within the NE and SE where the need for an SDAM in permitting and management programs is, or continues to be, particularly high. In addition, development associated with oil and natural gas, as well as agricultural uses that may require more and/or modified water sources due to climate change, occur across the NE and SE (Freeman and Marcinek 2006, Kanno and Vokoun 2010, Rugel et al. 2012, Vengosh et al. 2014). The state of North Carolina developed its own method (NC protocol) for statewide use in streams of all sizes and streamflow durations (NCDWQ 2010, Dorney and Russell 2018), which has served as a model for other methods developed in the NE and SE region, such as those developed for Tennessee (TDEC 2020) and parts of Virginia (James City County, VA 2009). Both the North Carolina and Tennessee methods were developed to support Clean Water Act section 401 programs and state level rules (e.g., riparian buffers in NC), whereas the James City County method was developed to comply with requirements under the Virginia Chesapeake Bay Preservation Act. Ohio also has an method (OH protocol) for headwater streams (drainage areas

<1.0 sq. mile; OH EPA 2020), which was developed to differentiate among classes of small headwater streams for appropriate characterization for aquatic life use designation.

5.2 Data collection

Data were collected at 209 reaches in the NE and 157 reaches in the SE. Three of these reaches in the NE and 4 in the SE were excluded from method development because a streamflow classification could not confidently be determined. The remaining 206 reaches in the NE and the 153 reaches in the SE were used for method development (Figure 44, Table 48). Each reach was visited between one and five times, yielding a total of 706 samples for the NE and 556 samples for the SE. Based on logger data, gages, or other sources of hydrologic information, 20% of the reaches in the NE were classified as ephemeral, 47% as intermittent, and 33% as perennial. Within the SE, 22% were classified as ephemeral, 47% as intermittent, and 31% as perennial. All states within the regions were sampled, with the exceptions of New Jersey and Delaware.

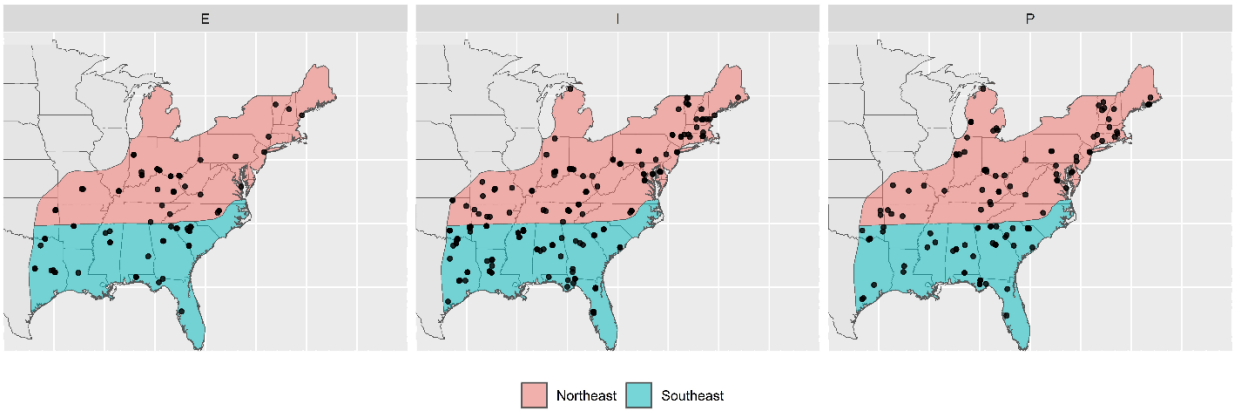


Figure 44. Location of reaches sampled to develop the NE and SE SDAMs.

Table 48. Distribution of reaches used to develop the NE and SE SDAM. Instrumented reaches were visited two to five times and had Stream Temperature, Intermittency, and Conductivity loggers installed and single-visit reaches were visited once (rarely, twice) and did not have loggers installed. Gaged reaches were located at reaches that had either nearby active or inactive USGS stream gages. Preferred reaches had multiple pre-existing data sources (e.g., water presence loggers and local expert observations) to support direct flow classification. Acceptable reaches were classified based only on local expert observations and were used to fill geographical gaps for study reaches.

Region	Class	Single visit, gaged	Single visit, preferred	Instrumented, gaged or other hydrologic data	Instrumented, preferred (based on multiple sources)	Instrumented, acceptable (based on BPJ or logger)	Total
NE	Ephemeral	3	15	0	7	16	41
NE	Intermittent	10	26	2	11	47	96
NE	Perennial	14	15	6	11	23	69
SE	Ephemeral	0	16	0	4	14	34
SE	Intermittent	5	17	1	14	35	72
SE	Perennial	8	7	1	8	23	47

5.3 Metric screening

5.3.1 Data corrections

In the process of reviewing the reaches and samples from the testing and training dataset, a few data corrections were performed on the dataset. These corrections are summarized below:

- A total of 7 reaches with an unknown streamflow duration were removed from the dataset (amounting in 22 visits total).
- Reach NHNE8905_V was removed due to missing values during its only reach visit
- Missing values were replaced for the following metrics, that did not vary significantly between the other site visits: *fp_entrenchmentratio_mean* (7 NAs)*, Slope* (2 NAs)*, BankWidthMean* (1 NA)*, PctShading* (1 NA)*, oh_substrate_dominance_ndom* (1 NA)*, oh_substrate_dominance_score* (1 NA)*, oh_substrate_significance_score* (1 NA), *SedimentOnPlantsDebris_score* (1 NA)

This results in 358 reaches (205 in NE, 153 in SE) and 1262 reach visits used in model development.

5.3.2 Metric Screening

Candidate indicator data were used to create 113 candidate metrics for inclusion in developing the final East method. Of these candidates, 51 were biological, 16 were geomorphological, 4 were indirect hydrologic measurements, and 42 were geospatial (Appendix B).

Ordination plots were created using these candidate metrics to gain an initial impression of the metrics to distinguish between flow class. Convex hulls were drawn around each flow class to help visualize their distributions in ordination space. The ordination of all candidate metrics for NE and SE samples showed extensive overlap of intermittent, ephemeral, and perennial reaches, indicating the challenge of separating samples by flow class (Figure 45). This was the case for both the Northeastern and Southeastern regions.

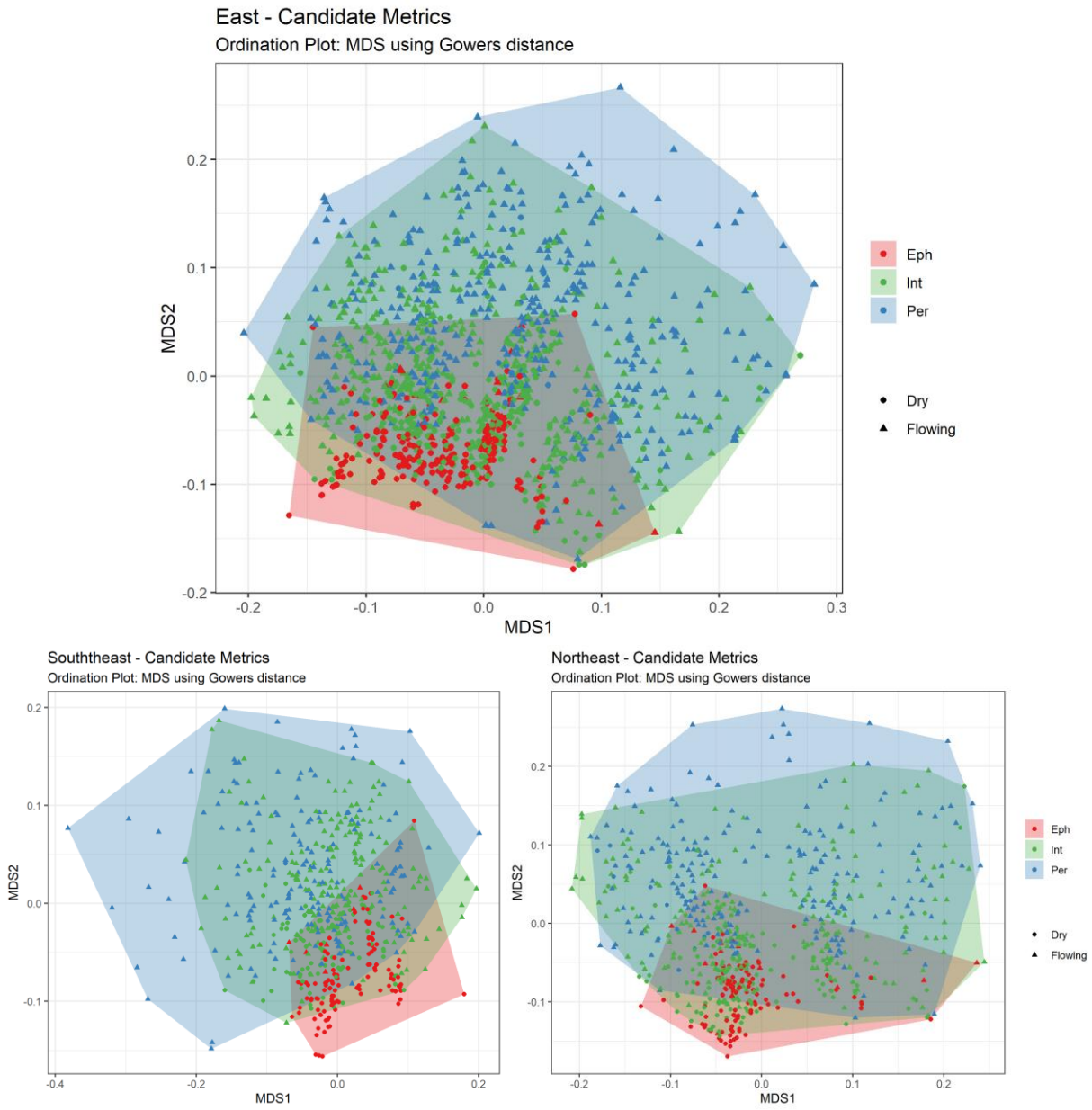


Figure 45. Ordination plot for the entire NESE. Ordination plots for (A) the NE and (B) SE separately.

After the initial data exploration, candidate metrics were evaluated using criteria for inclusion in the final SDAMs for the NE and SE (Table 11). Candidate metrics had to meet at least one responsiveness criterion, in addition to the distribution criterion, to be considered in further analyses. The performance of these metrics grouped or categorized by type and indicator are shown in Table 49. Three metrics failed screening: multiyear_tadpole_abundance, ppt.m07, and ppt.m09. These metrics failed because they had a Percent Dominance (PctDom) score greater than 95% and/or because they did not meet at least one of the responsiveness criteria. These three metrics were eliminated from further consideration.

Table 49. Screening criteria for metrics in the Northeast and Southeast. Metrics are described in Appendix B: Indicators and metrics evaluated for SDAM development. The following fields provide the screening criteria – PctDom: Percent of reach samples with the most common value (typically zero). Min: minimum value. Max = maximum value. Range: Maximum possible value minus minimum possible value for the candidate metric. PvIvE: F-statistic from a comparison of mean values at perennial, intermittent, and ephemeral reaches. EvALI: Absolute t-statistic from a comparison of mean values at ephemeral and at least intermittent reaches. PvNP: Absolute t-statistic from a comparison of mean values at perennial and non-perennial reaches. PvIWet: Absolute t-statistic from a comparison of mean values at flowing intermittent and perennial reaches. EvIdry: Absolute t-statistic from a comparison of mean values at non-flowing intermittent and ephemeral reaches. rf_MDA: Variable importance from a random forest model, measured as mean decrease in accuracy. Screened: Indicates if the metric passed or failed screening criteria.

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Algae	AlgalCover_Live_ NoUpstream	0.65	0	4	63.8	18.1	7.0	0.9	5.0	0.001	Pass
Biological	Algae	AlgalCover_LiveOr Dead_NoUpstream	0.64	0	4	64.9	18.4	7.0	0.9	5.2	0.001	Pass
Biological	Algae	ironox_bfscore_NC	0.85	0	3	31.6	10.1	6.0	2.9	3.3	0.001	Pass
Biological	Aquatic invertebrates	BMI_presence	0.75	0	1	350.7	21.3	18.8	1.5	7.6	0.014	Pass
Biological	Aquatic invertebrates	BMI_score_alt4	0.50	0	3	326.8	23.8	18.2	3.1	7.6	0.013	Pass
Biological	Aquatic invertebrates	Ephemeroptera_ abundance	0.62	0	31	117.2	16.4	11.7	7.5	2.7	0.005	Pass
Biological	Aquatic invertebrates	Ephemeroptera_taxa	0.62	0	5	180.6	20.9	15.1	8.9	3.4	0.006	Pass
Biological	Aquatic invertebrates	ephinteph_ISA_ abundance	0.58	0	32	40.8	6.6	5.2	7.6	4.9	0.010	Pass
Biological	Aquatic invertebrates	ephinteph_ISA_taxa	0.58	0	5	54.6	8.0	4.8	8.8	5.3	0.006	Pass
Biological	Aquatic invertebrates	EPT_abundance	0.50	0	87	131.0	19.0	12.3	7.2	3.8	0.007	Pass
Biological	Aquatic invertebrates	EPT_taxa	0.50	0	14	192.9	23.2	15.3	8.4	4.1	0.008	Pass
Biological	Aquatic invertebrates	GOLD_abundance	0.47	0	36	44.5	13.3	6.1	1.4	4.6	0.002	Pass
Biological	Aquatic invertebrates	Noninsect_ abundance	0.47	0	46	30.7	9.7	1.7	2.6	5.1	0.003	Pass
Biological	Aquatic invertebrates	OCH_abundance	0.61	0	25	34.2	12.1	5.9	3.0	4.0	0.001	Pass
Biological	Aquatic invertebrates	OCH_taxa	0.61	0	5	80.0	16.8	9.4	4.5	5.3	0.001	Pass
Biological	Aquatic invertebrates	perennial_ISA_ abundance	0.60	0	55	151.4	16.0	13.0	9.8	2.1	0.020	Pass
Biological	Aquatic invertebrates	perennial_ISA_taxa	0.60	0	9	262.9	20.6	17.5	12.4	2.8	0.020	Pass
Biological	Aquatic invertebrates	perennial_NC_ abundance	0.47	0	72	163.2	19.0	13.5	9.0	3.4	0.014	Pass
Biological	Aquatic invertebrates	perennial_NC_live_ abundance	0.48	0	72	160.3	18.7	13.4	9.0	3.3	0.013	Pass
Biological	Aquatic invertebrates	perennial_NC_taxa	0.47	0	14	276.7	24.2	18.1	11.3	4.0	0.017	Pass
Biological	Aquatic invertebrates	perintper_ISA_ abundance	0.54	0	61	128.5	18.6	12.3	7.3	3.1	0.011	Pass
Biological	Aquatic invertebrates	perintper_ISA_taxa	0.54	0	10	220.9	23.0	16.3	9.7	3.6	0.010	Pass
Biological	Aquatic invertebrates	Plecoptera_ abundance	0.72	0	35	42.1	13.4	6.7	2.2	2.0	0.001	Pass
Biological	Aquatic invertebrates	Plecoptera_taxa	0.72	0	7	66.4	16.1	8.7	3.3	2.7	0.001	Pass
Biological	Aquatic invertebrates	Richness	0.25	0	20	278.5	27.8	17.5	7.9	7.2	0.015	Pass
Biological	Aquatic invertebrates	TotalAbundance	0.25	0	112	146.4	21.1	11.9	4.8	6.6	0.018	Pass
Biological	Aquatic invertebrates	Trichoptera_ abundance	0.71	0	24	88.3	14.3	9.8	7.1	2.9	0.003	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Biological	Aquatic invertebrates	Trichoptera_taxa	0.71	0	4	127.2	18.4	12.5	7.7	3.3	0.002	Pass
Biological	Vegetation	DifferencesIn Vegetation_score	0.36	0	3	121.0	18.4	9.8	1.7	7.2	0.007	Pass
Biological	Vegetation	FibrousRootedPlants_score	0.58	0	3	57.9	6.9	10.4	4.3	2.0	0.002	Pass
Biological	Vegetation	hydrophytes_obl_ present	0.73	0	10	33.7	14.0	4.6	0.9	4.8	0.002	Pass
Biological	Vegetation	hydrophytes_obl_ present_inchannel	0.83	0	8	18.9	10.6	3.3	0.4	3.4	0.001	Pass
Biological	Vegetation	hydrophytes_present	0.34	0	16	81.9	16.3	8.9	3.8	5.2	0.003	Pass
Biological	Vegetation	hydrophytes_present_inchannel	0.70	0	12	13.8	7.8	2.5	0.0	2.5	0.000	Pass
Biological	Vegetation	Liverwort_cover	0.91	0	3	22.0	7.7	5.2	3.6	2.7	0.000	Pass
Biological	Vegetation	Moss_cover	0.94	0	3	9.0	8.0	2.5	0.5	2.7	0.000	Failk
Biological	Vegetation	PctShading	0.43	0	100	12.0	6.0	3.3	0.9	1.6	0.002	Pass
Biological	Vegetation	UplandRootedPlants_score	0.60	0	3	188.8	13.6	17.0	5.9	4.9	0.010	Pass
Biological	Vertebrates	Amphib_abundance	0.74	0	42	14.0	6.5	3.7	1.3	2.1	0.001	Pass
Biological	Vertebrates	Amphib_Juvenile_ abundance	0.86	0	38	8.7	7.2	2.7	0.7	3.5	0.000	Pass
Biological	Vertebrates	Amphib_Juvenile_ richness	0.86	0	2	25.2	10.9	4.7	1.4	4.0	0.000	Pass
Biological	Vertebrates	Amphib_richness	0.74	0	3	37.3	12.6	5.3	0.8	4.7	0.001	Pass
Biological	Vertebrates	Amphibian_presence	0.74	0	1	39.9	11.9	5.7	1.1	4.5	0.001	Pass
Biological	Vertebrates	AmphSnake_PA	0.74	0	1	39.9	11.9	5.7	1.1	4.5	0.001	Pass
Biological	Vertebrates	Fish_PA	0.77	0	1	89.7	17.7	10.5	4.7	3.1	0.002	Pass
Biological	Vertebrates	Frog_abundance	0.87	0	42	5.4	4.0	2.2	0.6	1.0	0.000	Pass
Biological	Vertebrates	Frog_richness	0.87	0	2	14.7	8.4	2.0	1.4	3.3	0.000	Pass
Biological	Vertebrates	Multiyear_tadpole_ abundance	1.00	0	2	0.8	1.9	1.0	1.7	1.0	0.000	Pass
Biological	Vertebrates	Salamander_ abundance	0.86	0	11	18.0	8.8	4.4	2.3	3.2	0.000	Pass
Biological	Vertebrates	Salamander_Juvenile_abundance	0.89	0	11	14.2	7.7	3.8	1.8	2.4	0.000	Pass
Biological	Vertebrates	Salamander_ richness	0.85	0	3	26.0	9.1	5.5	2.5	3.5	0.000	Pass
Geomorph	Geomorph	ActiveFloodplain_ score	0.29	0	3	54.3	10.1	7.1	1.2	3.7	0.002	Pass
Geomorph	Geomorph	AlluvialDep_score	0.50	0	3	52.8	12.0	7.3	1.7	3.3	0.001	Pass
Geomorph	Geomorph	BankWidthMean	0.02	0.2	52.333	62.2	14.5	8.6	4.4	4.4	0.017	Pass
Geomorph	Geomorph	Continuity_score	0.57	0	3	178.6	13.6	13.5	4.3	7.2	0.006	Pass
Geomorph	Geomorph	Depositional_score	0.33	0	3	131.1	17.6	12.4	5.0	5.7	0.003	Pass
Geomorph	Geomorph	fp_entrenchment ratio_mean	0.38	0.9795	2.5	10.7	2.0	4.5	3.6	0.1	0.001	Pass
Geomorph	Geomorph	GradeControl_score	0.30	0	1.5	26.6	4.4	7.0	5.5	3.0	0.002	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Geomorph	Geomorph	Headcut_score	0.75	0	3	19.7	5.5	3.7	0.0	1.7	0.001	Pass
Geomorph	Geomorph	NaturalValley_score	0.38	0	1.5	15.3	1.3	4.7	4.8	1.8	0.003	Pass
Geomorph	Geomorph	oh_substrate_ dominance_score	0.25	0	32	69.4	11.4	8.7	4.5	6.6	0.004	Pass
Geomorph	Geomorph	oh_substrate_total_ substrate_types	0.23	0	8	31.9	6.5	6.2	4.1	5.3	0.002	Pass
Geomorph	Geomorph	RifflePoolSeq_score	0.33	0	3	100.1	11.7	11.2	3.6	4.9	0.002	Pass
Geomorph	Geomorph	SedimentOnPlants Debris_score	0.37	0	1.5	46.4	11.7	4.8	0.3	5.3	0.001	Pass
Geomorph	Geomorph	Sinuosity_score	0.55	0	3	35.0	8.5	4.5	0.2	4.1	0.001	Pass
Geomorph	Geomorph	Slope	0.25	0	46	48.0	7.5	8.0	2.4	2.9	0.006	Pass
Geomorph	Geomorph	SubstrateSorting_ score	0.41	0	3	195.9	15.2	16.0	7.4	8.1	0.011	Pass
Geospatial	Precipitation	ppt	0.01	772.69	1717.2	6.8	0.6	3.5	4.1	1.2	0.013	Pass
Geospatial	Precipitation	ppt.11121	0.01	48.673	152.9	5.8	0.2	3.3	3.6	0.6	0.010	Pass
Geospatial	Precipitation	ppt.234	0.01	46.176	142.87	3.8	0.4	2.6	2.7	0.4	0.013	Pass
Geospatial	Precipitation	ppt.567	0.01	66.926	154.64	18.1	2.4	5.8	5.4	1.3	0.018	Pass
Geospatial	Precipitation	ppt.8910	0.01	75.76	152.54	0.4	0.9	0.3	1.4	1.9	0.012	Pass
Geospatial	Precipitation	ppt.m01	0.01	32.35	154.45	0.5	0.7	0.9	1.5	2.1	0.009	Pass
Geospatial	Precipitation	ppt.m02	0.01	33	147.82	1.7	0.8	1.8	2.1	1.3	0.009	Pass
Geospatial	Precipitation	ppt.m03	0.01	44.95	158.47	1.5	1.5	1.3	1.3	1.9	0.013	Pass
Geospatial	Precipitation	ppt.m04	0.02	57.59	138.05	14.6	1.7	4.3	4.4	2.8	0.016	Pass
Geospatial	Precipitation	ppt.m05	0.01	59.95	180.33	25.7	1.5	7.2	5.8	1.8	0.020	Pass
Geospatial	Precipitation	ppt.m06	0.01	68.97	197.13	17.7	3.4	5.7	5.3	3.6	0.021	Pass
Geospatial	Precipitation	ppt.m07	0.01	45.67	206.16	0.5	0.5	0.7	0.2	1.7	0.011	Fail
Geospatial	Precipitation	ppt.m08	0.01	51.71	232.4	6.9	0.5	3.4	2.1	3.4	0.013	Pass
Geospatial	Precipitation	ppt.m09	0.01	68.9	178.02	1.3	1.5	0.5	1.0	0.7	0.010	Fail
Geospatial	Precipitation	ppt.m10	0.02	56.92	150.35	9.1	1.0	3.7	4.7	0.1	0.013	Pass
Geospatial	Precipitation	ppt.m11	0.01	46.37	162.32	7.7	1.1	3.2	3.9	0.6	0.014	Pass
Geospatial	Precipitation	ppt.m12	0.01	47.23	166.23	12.0	0.9	5.0	4.7	0.2	0.012	Pass
Geospatial	Snow persistence	MeanSnow Persistence_01	0.00	0.0038	44.949	14.9	5.9	3.9	2.1	2.9	0.011	Pass
Geospatial	Snow persistence	MeanSnow Persistence_05	0.00	0.0825	45.129	14.8	5.8	4.0	2.1	2.9	0.011	Pass
Geospatial	Snow persistence	MeanSnow Persistence_10	0.00	0.1570	45.399	14.7	6.0	3.9	2.0	3.1	0.010	Pass
Geospatial	Temperature	temp.11121	0.01	-5.2066	17.756	8.2	3.8	3.2	2.7	4.0	0.007	Pass
Geospatial	Temperature	temp.234	0.01	-2.4533	19.506	10.9	4.3	3.8	2.9	3.7	0.007	Pass

Type	Indicator	Metric	PctDom	Min	Max	PvIvE	EvALI	PvLTP	PvIwet	EvIdry	rf_MDA	Screen
Geospatial	Temperature	temp.567	0.02	14.366	27.243	10.9	4.1	4.0	3.1	3.5	0.007	Pass
Geospatial	Temperature	temp.8910	0.01	11.92	26.616	10.1	3.9	3.9	3.2	3.6	0.008	Pass
Geospatial	Temperature	temp.m01	0.02	-9.67	15.75	9.1	4.0	3.4	2.7	4.0	0.007	Pass
Geospatial	Temperature	temp.m02	0.01	-8.09	17.33	9.6	4.0	3.6	2.8	3.8	0.007	Pass
Geospatial	Temperature	temp.m03	0.02	-3.36	19.36	11.3	4.4	3.9	3.0	3.7	0.008	Pass
Geospatial	Temperature	temp.m04	0.02	3.94	21.83	12.1	4.6	4.1	3.0	3.7	0.007	Pass
Geospatial	Temperature	temp.m05	0.02	10.34	25.4	11.3	4.1	4.1	3.4	3.6	0.008	Pass
Geospatial	Temperature	temp.m06	0.02	15.18	27.87	10.7	4.3	3.8	3.0	3.6	0.007	Pass
Geospatial	Temperature	temp.m07	0.02	17.58	28.88	10.2	4.0	3.9	3.0	3.2	0.009	Pass
Geospatial	Temperature	temp.m08	0.02	16.7	29.34	11.0	3.9	4.2	3.4	3.1	0.009	Pass
Geospatial	Temperature	temp.m09	0.02	12.67	27.35	9.9	4.0	3.8	3.1	3.7	0.008	Pass
Geospatial	Temperature	temp.m10	0.02	6.39	24.36	9.3	3.7	3.7	3.2	3.7	0.007	Pass
Geospatial	Temperature	temp.m11	0.02	0.45	20.47	8.6	3.8	3.4	2.9	4.0	0.007	Pass
Geospatial	Temperature	temp.m12	0.02	-6.4	17.05	6.9	3.6	2.9	2.5	4.1	0.007	Pass
Geospatial	Temperature	tmax	0.01	10.6	28.26	10.0	3.9	3.8	3.0	3.4	0.009	Pass
Geospatial	Temperature	tmean	0.02	4.66	22.71	10.0	4.1	3.7	3.0	3.8	0.007	Pass
Geospatial	Temperature	tmin	0.02	-1.36	17.52	9.8	4.2	3.6	2.9	4.1	0.009	Pass
Geospatial	Watershed	DRNAREA_mi2	0.05	0.002	579	24.4	7.3	4.9	4.1	2.9	0.042	Pass
Geospatial	Watershed	Elev_m	0.02	10	887	0.2	0.2	0.5	0.5	2.4	0.008	Pass
Geospatial	Watershed	StreamOrder	0.59	1	3	133.3	22.5	12.2	6.0	7.8	0.015	Pass
Hydrological	Indirect	HydricSoils_score	0.71	0	3	64.4	10.1	2.5	3.4	6.9	0.004	Pass
Hydrological	Indirect	LeafLitter_score	0.30	0	1.5	132.0	16.0	11.8	3.8	5.0	0.003	Pass
Hydrological	Indirect	ODL_score	0.39	0	1.5	51.6	9.8	7.0	2.1	4.5	0.001	Pass
Hydrological	Indirect	WoodyJams_number	0.83	0	11	2.4	1.2	2.1	2.3	1.0	0.000	Pass

5.4 Data stratification and preparation

In the Eastern region, there were a total of 1262 samples across 358 reaches that represented ephemeral, intermittent, and perennial reaches (23 reach visits were removed because unknown flow duration or incomplete data from only site visit) with a class distribution and spatial distribution shown in Figure 46.

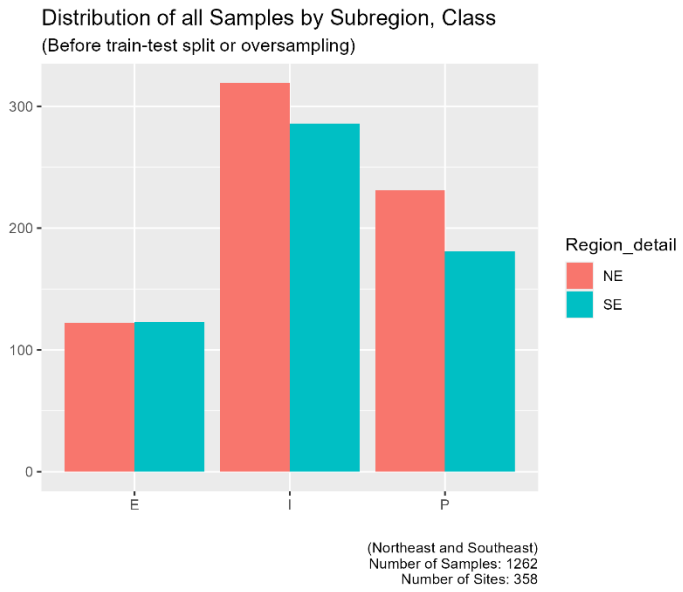


Figure 46. Distribution of samples by streamflow class for the East Region.

Of the 358 reaches included in the NE and SE datasets, each was visited between one and five times, yielding a total of 1262 samples. Figure 47 shows the distribution of repeat reach visits. To minimize bias, oversampling was performed on the training dataset. Oversampling, as described in Section 2.6.3 *Rebalancing uneven reach visits*, is a common preprocessing step that serves to give under-represented classes more visibility in the data (Mohammed et al. 2020).

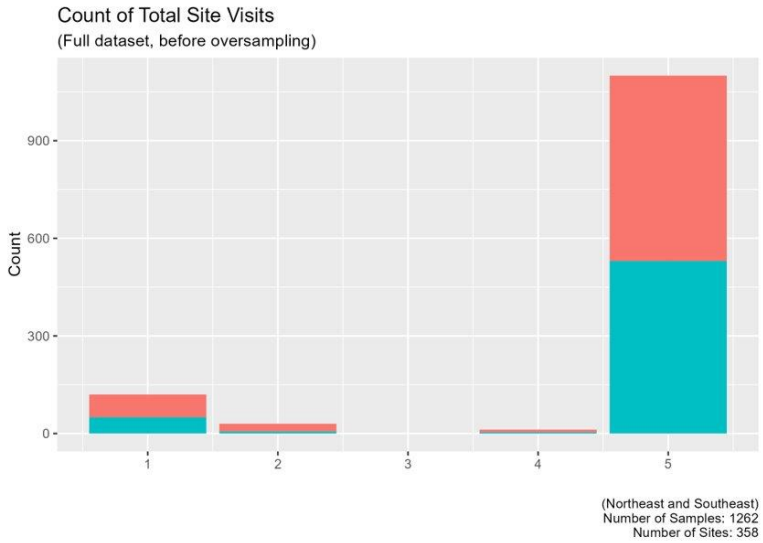


Figure 47. Total reach visits for the East Region.

The result of the oversampling process was that each study reach had as close to five sample as possible for method development and the distribution of flow classes was preserved from the original training dataset to the oversampled training dataset, which also matched well to the distribution of flow classes within the testing dataset (Figure 48). Therefore, the oversampled

training data with 1550 samples (as opposed to the original training dataset of 1022 samples) were used in subsequent model development steps.

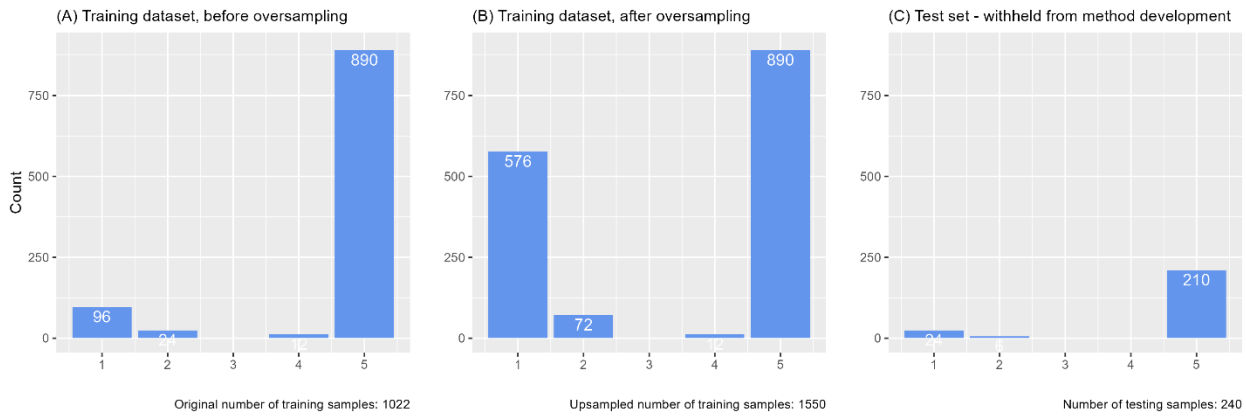


Figure 48. Distribution of reach visits in (A) training dataset before oversampling, (B) training dataset after oversampling, and the (C) testing dataset (not oversampled). Shown for each bar is the ephemeral (E), intermittent (I), and perennial (P) classification colored by class. A balanced distribution between classes is important to mitigate against bias and improve model accuracy. Note that no rebalancing was performed on the test set.

5.5 Model development

At the beginning of the analysis, it was unclear whether greater accuracy would be achieved having one model for the aggregated NE and SE (Unstratified model) or two models separated by region (Stratified models). Thus, the preliminary modeling process was applied to all three types of models In addition, it was unclear whether geospatial (GIS) metrics should be included in the model-building process or excluded (see Mazor et al. (2021b) for a discussion of the implications of including geospatial metrics in SDAMs). Therefore, a total of 6 model versions were tested to determine which model type would perform best for the East region:

- 1 unstratified model, the entire East region with No GIS metrics included
- 1 unstratified model, the entire East region with 2 GIS metrics included
- 2 stratified versions of the model (Northeast and Southeast) with No GIS metrics included
- 2 stratified versions of the model (Northeast and Southeast) with 2 GIS metrics included

For the unstratified model, the model was developed based on the entire East dataset, and then performance was evaluated for each subregion (Northeast and Southeast). For the stratified models, the model was both developed and tested on only its named region.

After running the models with unlimited GIS metrics in the beta version, it was determined that the GIS metrics dominated the candidate metrics. Hence, the model versions for consideration include GIS restrictions. In addition to the model versions that restrict all GIS metrics (“No GIS”), the “2 GIS” model versions use the top 2 GIS metrics that present themselves in the unlimited GIS model. Table 50 compares the performance of the 6 preliminary models to determine the degree of improved performance by the inclusion of GIS metrics and strata-specific models. The

selected metrics for each model are shown in Figure 49. Bank width mean, ephinteph_ISA_abundance, and percent shading were the only metrics that were selected across all 6 models.

Table 50. Design characteristics of the preliminary East models. GIS: included geospatial metrics. # samples: number of samples used in model training and testing. RFE OOB error rate: out-of-bag (OOB) error rate of the best model produced by recursive feature elimination.

Model set	Stratum	# samples (training)	# samples (testing)	# metrics eligible	# metrics chosen	RFE OOB error rate	Cohen's Kappa
Unstratified models							
Unstratified NoGIS	East	1550	240	70	18	0.15	0.76
Unstratified 2GIS	East	1550	240	72	11	0.10	0.83
Models stratified by region							
Stratified NoGIS	Northeast	888	139	70	11	0.16	0.75
Stratified NoGIS	Southeast	662	101	70	20	0.15	0.78
Stratified 2GIS	Northeast	888	139	72	10	0.08	0.86
Stratified 2GIS	Southeast	662	101	72	10	0.10	0.84

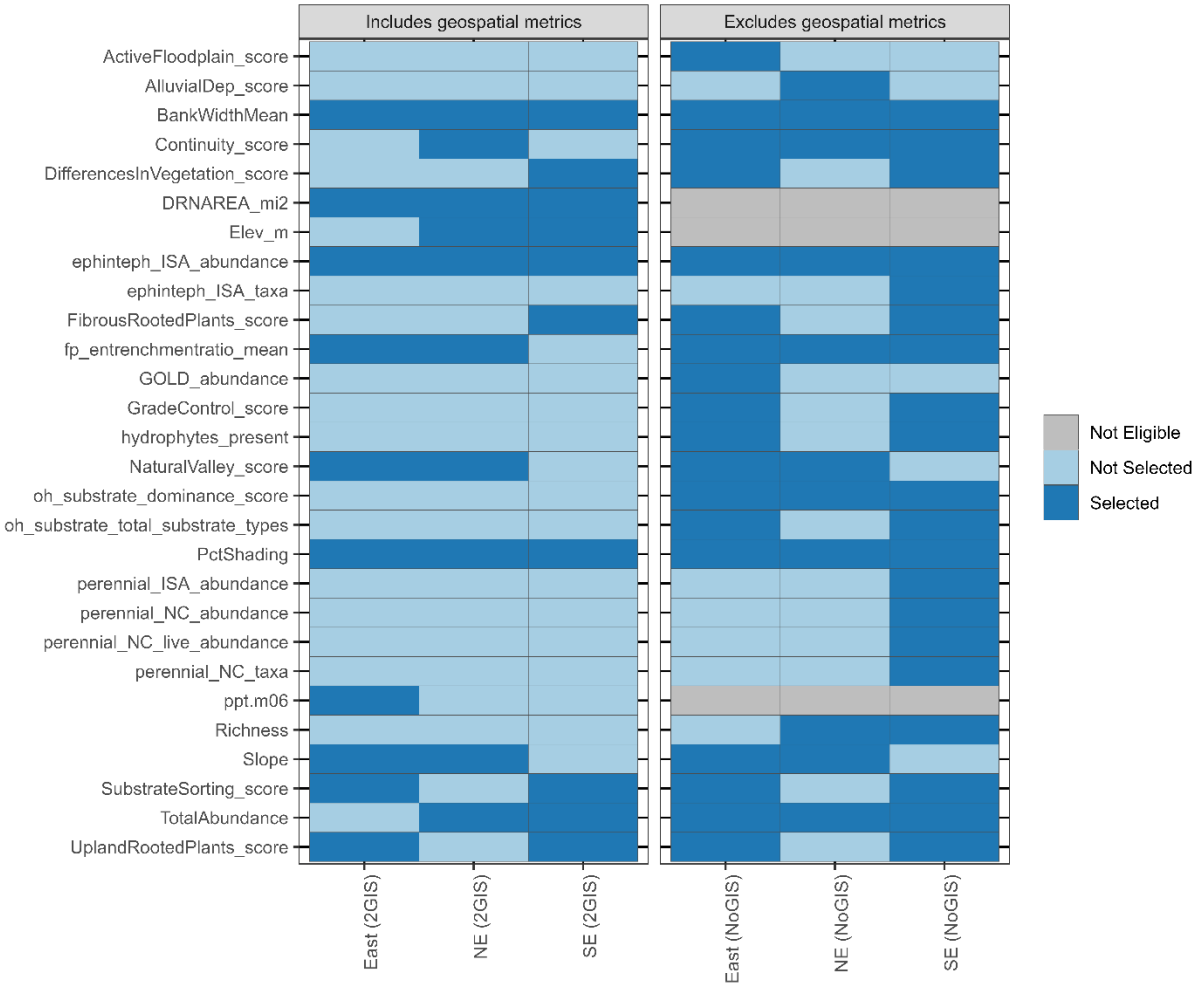


Figure 49: Comparison of Selected Metrics for each Preliminary Model.

In general, the accuracies in the training dataset were highest for models including GIS metrics. However, performance of the models containing geospatial data sharply decreased when evaluated against the testing dataset, indicating that the GIS models were overfitting to the training dataset. Through discussion with the RSC, the stratified models containing a limited number of GIS metrics were determined to be the most appropriate models. The top 2 GIS metrics were chosen by first removing highly correlated metrics from groups of geospatial metrics (e.g. precipitation, temperature, snow). Drainage area and Elev_m were the non-correlated geospatial metrics with highest performance for the initial Northeast 2GIS and Southeast 2 GIS models. The strength of each stratified (2 GIS) model increases when looking at the ability of the model to accurately distinguish between ephemeral and at least intermittent (EvALI) compared to distinguishing between all three classes (PvIvE; Figure 50).

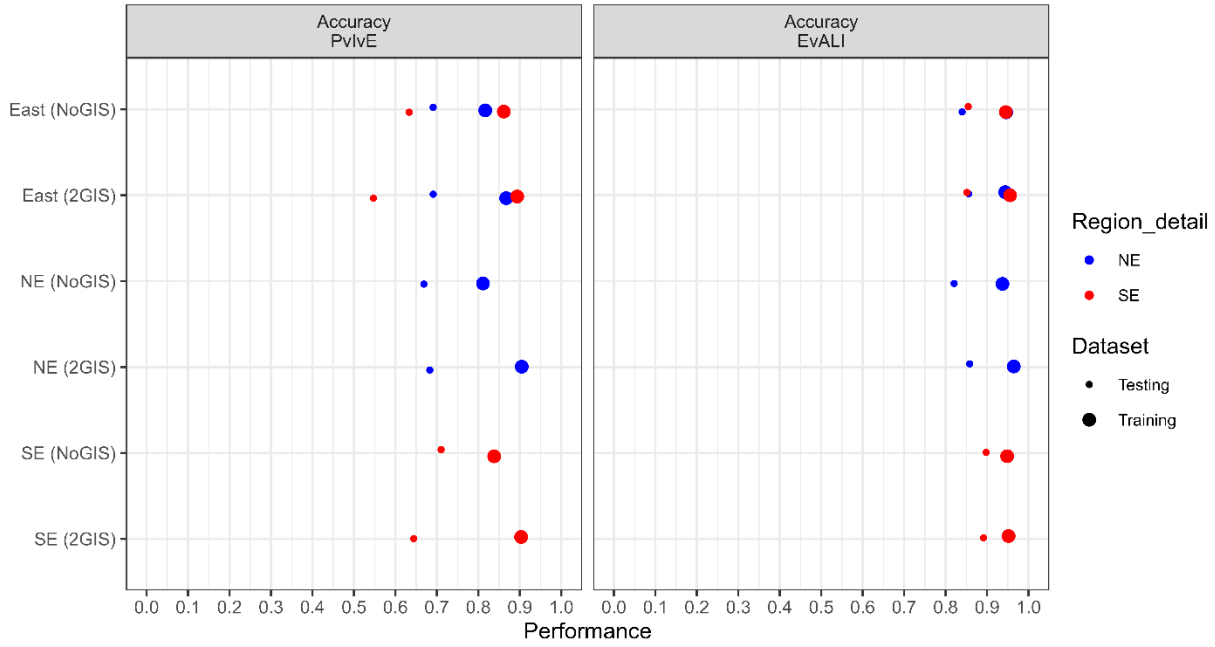


Figure 50. Preliminary model comparison.

Based on the comparison of performance of the preliminary models, the stratified 2GIS model was selected for further refinement. Both the Northeast (2GIS) and Southeast (2GIS) model contained ten metrics that were selected via RFE. The metrics are shown in Figure 51 in order of relative mean decrease in accuracy.

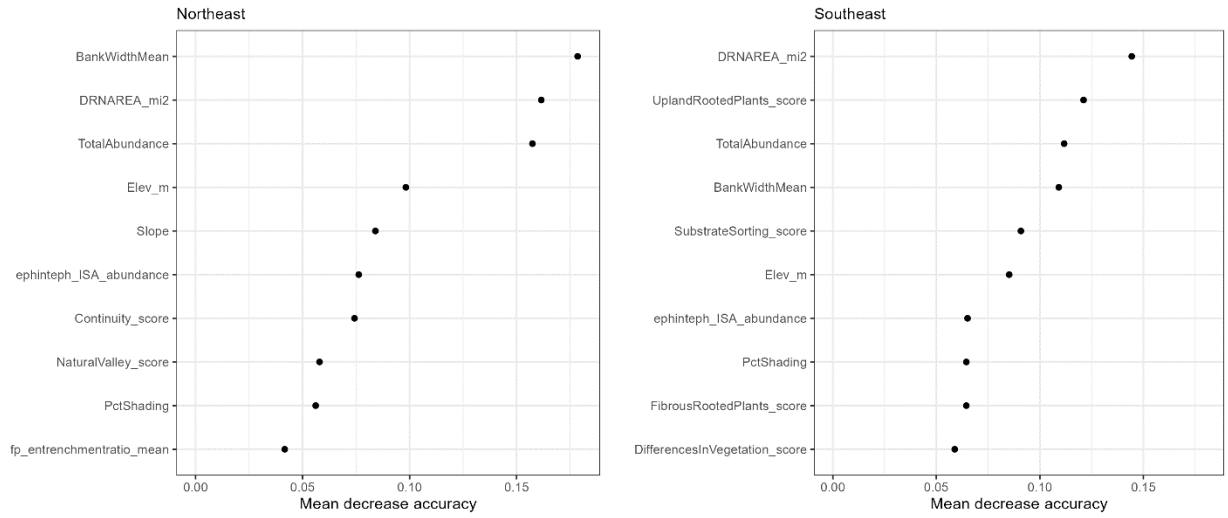


Figure 51. Metrics included in the Northeast (2 GIS) and the Southeast (2 GIS) models. Mean Decrease in Accuracy is the relative loss in predictive performance when the variable is omitted from the model.

To evaluate the overall performance of the preliminary models for the Northeast and Southeast, confusion matrices were created for both training and testing datasets (Table 51). No perennial-ephemeral misclassifications (“big errors”) occurred in either testing or training datasets.

Table 51. Accuracy of the preliminary NE and SE models. Bright green shaded cells indicate correct predictions for the PvIvE accuracy measure. Light green shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Light blue shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple shaded cells indicate "big errors" (i.e., ephemeral reaches classified as perennial, or perennial reaches classified as ephemeral) and are incorrect for all accuracy measures. Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors."

Northeast: Training

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	165	11	0
Intermittent	14	366	16
Perennial	0	21	274
At least intermittent	0	3	11
Less than perennial	2	5	0
Need more information	0	0	0

Northeast: Testing

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	22	4	0
Intermittent	8	53	14
Perennial	0	3	20
At least intermittent	0	5	1
Less than perennial	3	6	0
Need more information	0	0	0

Southeast: Training

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	137	9	0
Intermittent	15	281	21
Perennial	0	9	175
At least intermittent	0	1	6
Less than perennial	2	5	1
Need more information	0	0	0

Southeast: Testing

Predicted class	Ephemeral	Intermittent	Perennial
Ephemeral	14	0	0
Intermittent	6	34	17
Perennial	0	1	1
At least intermittent	0	6	18
Less than perennial	2	1	1
Need more information	0	0	0

5.6 Model refinement

Balancing both high accuracy and ease of field collection, the stratified Northeast and Southeast models with 2 GIS metrics were selected for further refinement. Using distribution boxplots for each of the candidate metrics selected, refinements were performed manually by systemically binning and/or dropping each metric and evaluating the effect on model performance.

The purpose of refining the preliminary model was to make the SDAM easier to implement in the field while improving (or at least not sacrificing) performance, by removing or simplifying the most difficult to measure metrics. Simplifying metrics was intended to improve efficient SDAM application and facilitate method use and transparency. Improving the efficiency of the SDAM application also contributes to ensuring that the SDAM can be applied during a single reach visit.

First, complicated or potentially redundant metrics were attempted to be dropped or simplified (either by removing them or replacing them with an alternative metric). Some metrics were eliminated because they were closely related to another metric in the selected model (i.e., they described similar stream characteristics, such as rooted upland plants and fibrous roots of upland plants presence). Metrics that were more time-consuming to measure were rejected if a simpler alternative was available and continuous metrics were converted to binary or ordinal metrics based on visual interpretation of their distributions. (Binary and ordinal metrics are typically more rapid to measure and easier to standardize than continuous metrics.) Accuracy and repeatability measures were re-evaluated to ensure that overall model performance was not substantially diminished by the modifications.

The suite of metrics of the selected model was iteratively refined while monitoring model accuracy and repeatability. In each iteration, one or more metrics were either eliminated, binned, or otherwise simplified. The impact of each iterative refinement on performance was assessed, and the highest performing refined model was selected. Performance was assessed in terms accuracy measures: PvIvE, EvALI and EvIdry.

Key steps of the model refinement phase and the final model selected are detailed in the 5.6.1 Northeast Model Refinement and 5.6.2 Southeast Model Refinement sections below. Performance according to three accuracy metrics on the testing dataset are shown compared to the original stratified 2GIS model version (V0) for both NE and SE. Callout boxes indicate how the metrics were changed compared to model V0.

5.6.1 Northeast Model Refinement

Table 52 shows the most successful refinements performed on the preliminary Northeast model, as compared to v0 (the unrefined version of the model). For example, one example of a refinement made between V0 and V1 is binning the originally continuous BankWidthMean metric into BankWidthMean_10 (<10 m or ≥ 10). The unrefined model (v0) had the following test set accuracies: PvIvE=0.68, EvALI=0.85, and EvIdry=0.73. As shown in Figure 52 and Table 53, many refinements maintained or improved their accuracy for most measures. Based on these accuracy measures and the number of errors in the confusion matrices, refinement v7 was selected as the final model. This resulted in the final selected set of indicators:

- *BankWidthMean_10_bin*
- *BMI_score_alt4*
- *DRNAREA_mi2*
- *Elev_m*
- *fp_entrenchmentratio_mean_1.5_2_bin*
- *PctShading*
- Slope_10_bin

This refined (v7) model resulted in a simpler field collection method with the following test set accuracies: PvIvE=0.73, EvALI=0.88, and EvIdry=0.73.

Table 52. Steps in refining the NE model. Step 7, marked with an asterisk (*) was the final selected refined model. Grey boxes identify indicators included in a refined model, B = binned modification, and S = simplified modification.

Step	Preceding step	DRNAREA	Continuity_score	BankWidthMean	Elev_m	Slope	Fp_entrenchmentratio_mean	NaturalValley_score	PctShading	ephinteph_ISA_abundance	TotalAbundance	Ppt8910	BMI_score_alt4	Modification
0	NA													Preliminary model
1	0			B		B								Dropped Continuity_score; Dropped Fp_entrenchmentratio_mean; Binned BankWidthMean at 10, Binned Slope at 10; Dropped Total Abundance; Added BMI_score_alt4
2	1			B		B								Dropped BMI_score_alt4; Add back Total Abundance binned at 1, 10, 33; Added ppt8910; Dropped PctShading

Step	Preceding step	DRNAREA	Continuity_score	BankWidthMean	Elev_m	Slope	Fp_entrenchmentratio_mean	NaturalValley_score	PctShading	ephinteph_ISA_abundance	TotalAbundance	Ppt8910	BMI_score_alt4	Modification
3	1			B		B								Added Continuity_score; Dropped ephinteph_ISA_abundance; Dropped NaturalValley_score
4	3			B		B								Added ephinteph_ISA_abundance; Dropped Continuity_score
5	4			B		B								Dropped ephinteph_ISA_abundance; Added Fp_entrenchmentratio_mean
6	5			B		B				B				Added ppt8910; Dropped PctShading; Added a binned ephinteph_ISA_abundance at 1 and 10
7*	5			B		B	B							Binned Fp_entrenchmentratio_mean at 1.5 and 2

Table 53. Training and testing accuracy of each step in the refinement of the NE model. Step 7, marked with an asterisk (*) was the final selected refined model. Statistics are calculated for the entire NE region.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0	None	0.96	0.85	0.90	0.73	0.91	0.68
1	0	0.95	0.87	0.87	0.69	0.89	0.68
2	1	0.97	0.88	0.93	0.73	0.95	0.67
3	2	0.94	0.88	0.86	0.76	0.87	0.73
4	3	0.96	0.91	0.89	0.77	0.94	0.76
5	4	0.95	0.88	0.87	0.69	0.89	0.71
6	5	0.96	0.91	0.90	0.74	0.94	0.65
7*	6	0.95	0.88	0.88	0.73	0.88	0.73

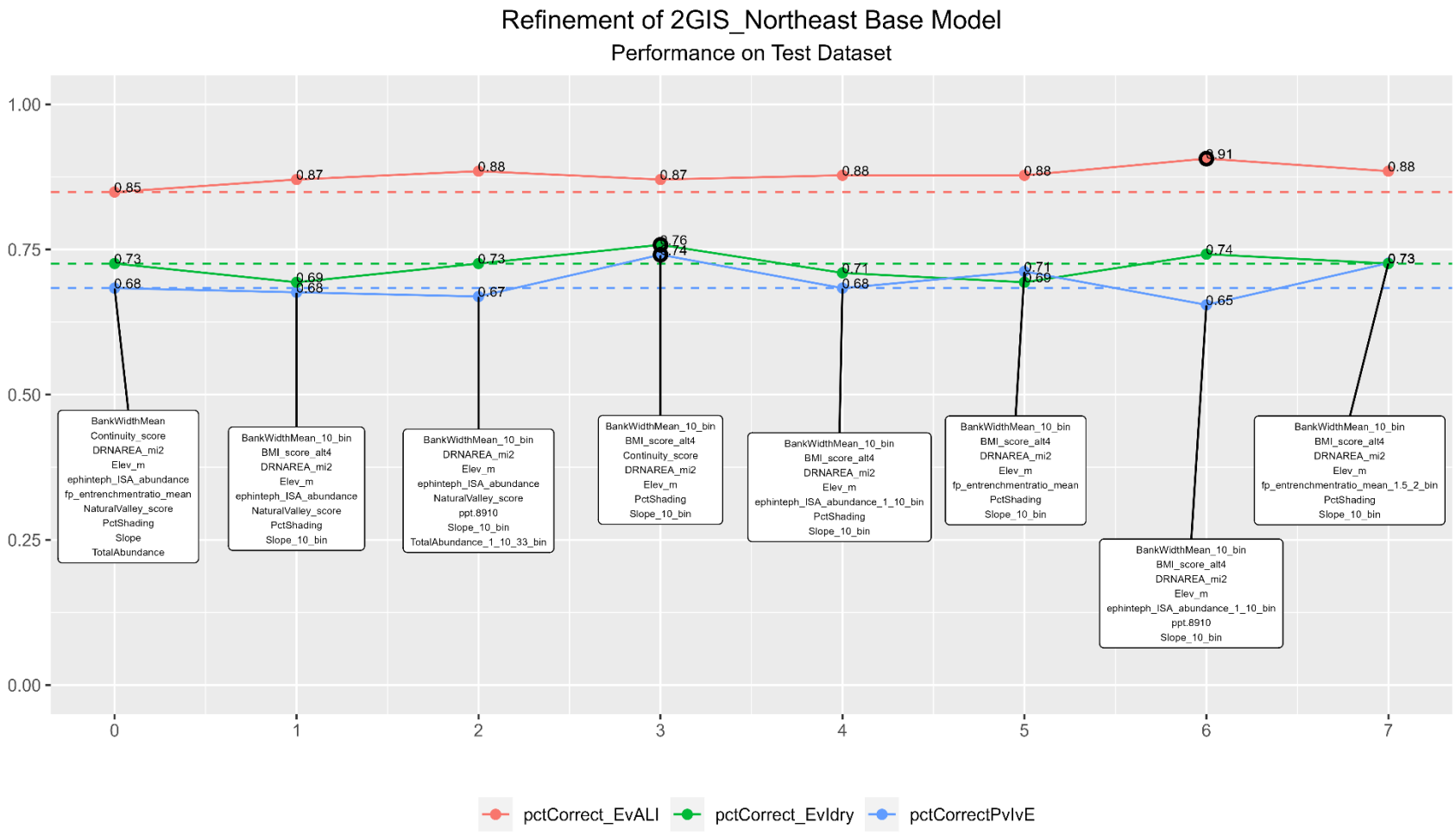


Figure 52. Northeast (2GIS) refinement versions 1 through 7. Chosen refinement is version 7.

5.6.2 Southeast Model Refinement

Table 54 and Figure 53 show some of the refinement attempts as compared to V0 (the unrefined version of the model). For example, a refinement made between V0 and V1 was simplifying the original continuous metric TotalAbundance into a binned metric called TotalAbundance_0_3_40_bin (0, 1-2, 3-39, 40+).

Testing data performance of the unrefined version (V0) of the Southeast model was PvIvE=0.65, EvALI=0.90, EvIdry=0.79 as shown in Table 55 and Figure 53. Many refinements were able to improve the test accuracy for all measures while simplifying the model. Based on these accuracy metrics and consideration of field collection difficulty, refinement version 7 was selected as the final model for the SE moving forward. This process resulted in the final selected set of indicators:

- *BankWidthMean*
- *BMI_score_alt4*
- *DRNAREA_mi2*
- *Elev_m*
- *FibrousRootedPlants_score*
- *PctShading*
- *ppt.567*
- *SubstrateSorting_score*
- *TotalAbundance_0_3_40_bin*
- *UplandRootedPlants_score*

This refined (v7) model resulted in a simpler field collection method with improvements in all accuracy measures on testing data: PvIvE=0.67, EvALI=0.91, EvIdry=0.82.

Table 54. Steps in refining the SE model. Step 7, marked with an asterisk (*) was the final selected refined model. Grey boxes identify indicators included in a refined model, B = binned modification, and S = simplified modification.

Step	Preceding step	DRNAREA	SubstrateSorting_score	BankWidthMean	Elev_m	FibrousRootedPlants_score	UplandRootedPlants_score	PctShading	ephinteph_ISA_abundance	TotalAbundance	Ppt567	DifferencesInVegetation_scor	BMI_score_alt4	Modification
0	NA													Preliminary model
1	0									B				Binned TotalAbundance at 0, 3, 40
2	0													Dropped TotalAbundance

Step	Preceding step	DRNAREA	SubstrateSorting_score	BankWidthMean	Elev_m	FibrousRootedPlants_score	UplandRootedPlants_score	PctShading	ephinteph_ISA_abundance	TotalAbundance	Ppt567	DifferencesInVegetation_scor	BMI_score_alt4	Modification
3	1			B						B				Dropped ephinteph_ISA_abundance; Binned BankWidthMean; Dropped PctShading
4	1									B				Added BMI_score_alt4; Added ppt567
5	3			B						B				Added ppt567; Dropped DifferencesInVegetation_score
6	4													Dropped PctShading
7*	4													Dropped DifferencesInVegetation_score;

Table 55. Training and testing accuracy of each step in the refinement of the SE model. Step 7, marked with an asterisk (*) was the final selected refined model. Statistics are calculated for the entire SE region.

Step	Preceding step	EvALI training	EvALI testing	EvIdry training	EvIdry testing	PvIvE training	PvIvE testing
0	None	0.95	0.90	0.88	0.79	0.90	0.65
1	0	0.95	0.91	0.89	0.82	0.90	0.64
2	1	0.95	0.87	0.87	0.74	0.90	0.65
3	2	0.95	0.91	0.89	0.76	0.88	0.61
4	3	0.96	0.92	0.89	0.82	0.91	0.67
5	4	0.96	0.91	0.92	0.79	0.92	0.64
6	5	0.96	0.92	0.89	0.82	0.92	0.65
7*	6	0.96	0.91	0.89	0.82	0.92	0.67

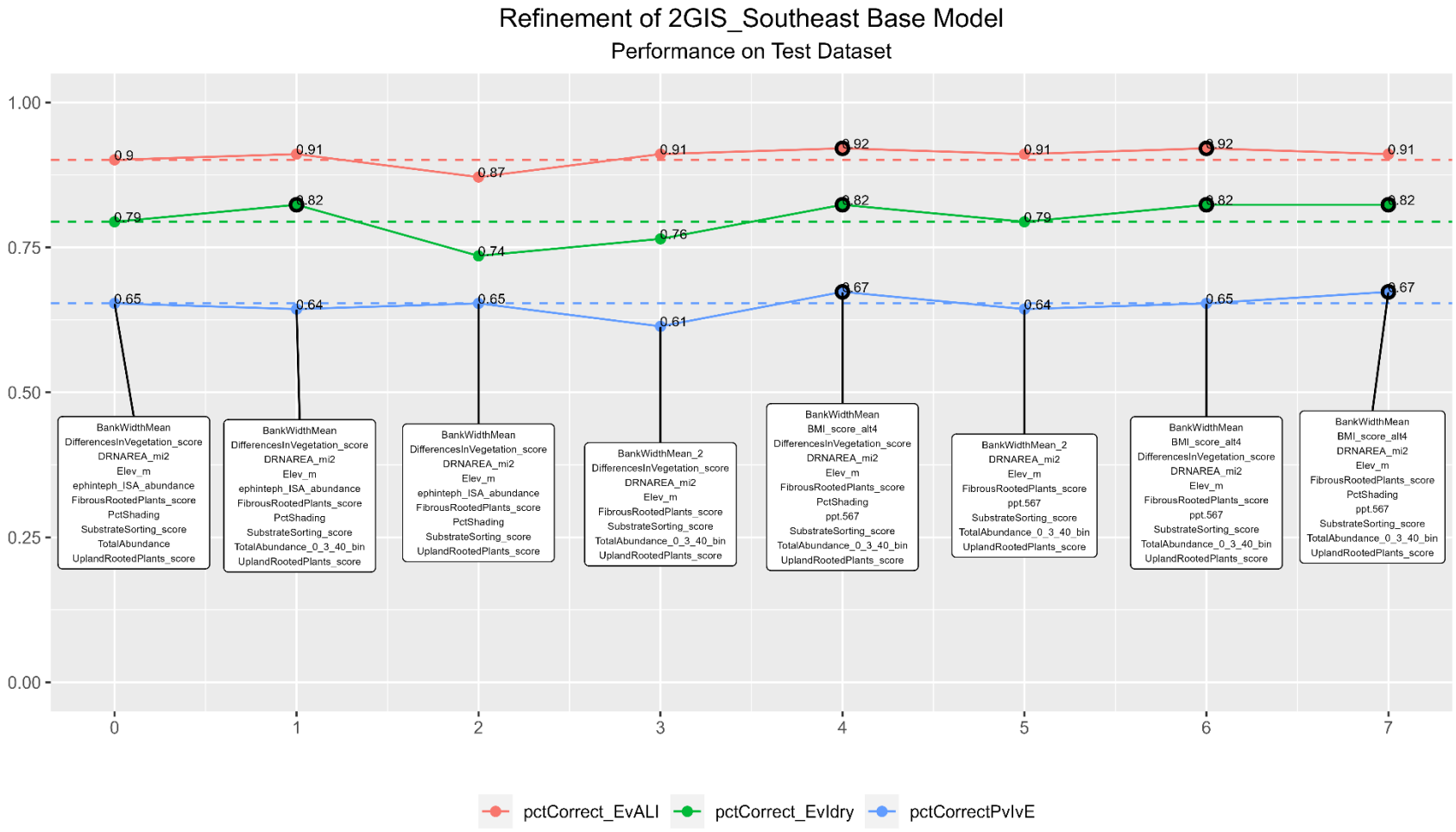


Figure 53. Southeast (2GIS) Refinement Versions 1 through 7. Chosen refinement was version 7.

5.6.3 Single indicators

Single indicators can supersede a model classification of *ephemeral* to make it change to *at least intermittent*. Single indicators provide technical benefits (i.e., improved accuracy) as well as non-technical benefits, such as greater acceptance of the SDAM, given public understanding of the role of streamflow duration in supporting biological organisms and rapidity of determining a flow

classification. Single indicators are also used in other SDAMs (e.g., Nadeau et al. 2015, Dorney and Russell 2018, Mazor et al. 2021a); for instance, indicators can include the presence of fish, iron-oxidizing bacteria, hydric soils, and/or aquatic vertebrates (amphibians and reptiles), among others. See Table 14 for list of single indicators evaluated for the final Northeast and Southeast models.

The impacts of including each single indicator in the final Northeast and Southeast models are summarized in Figure 54Figure 54 and Figure 55. All single indicators had minimal impact on performance or introduced more errors than they corrected. Therefore, the RSC did not recommend inclusion of single indicators in either the NE or SE final models.

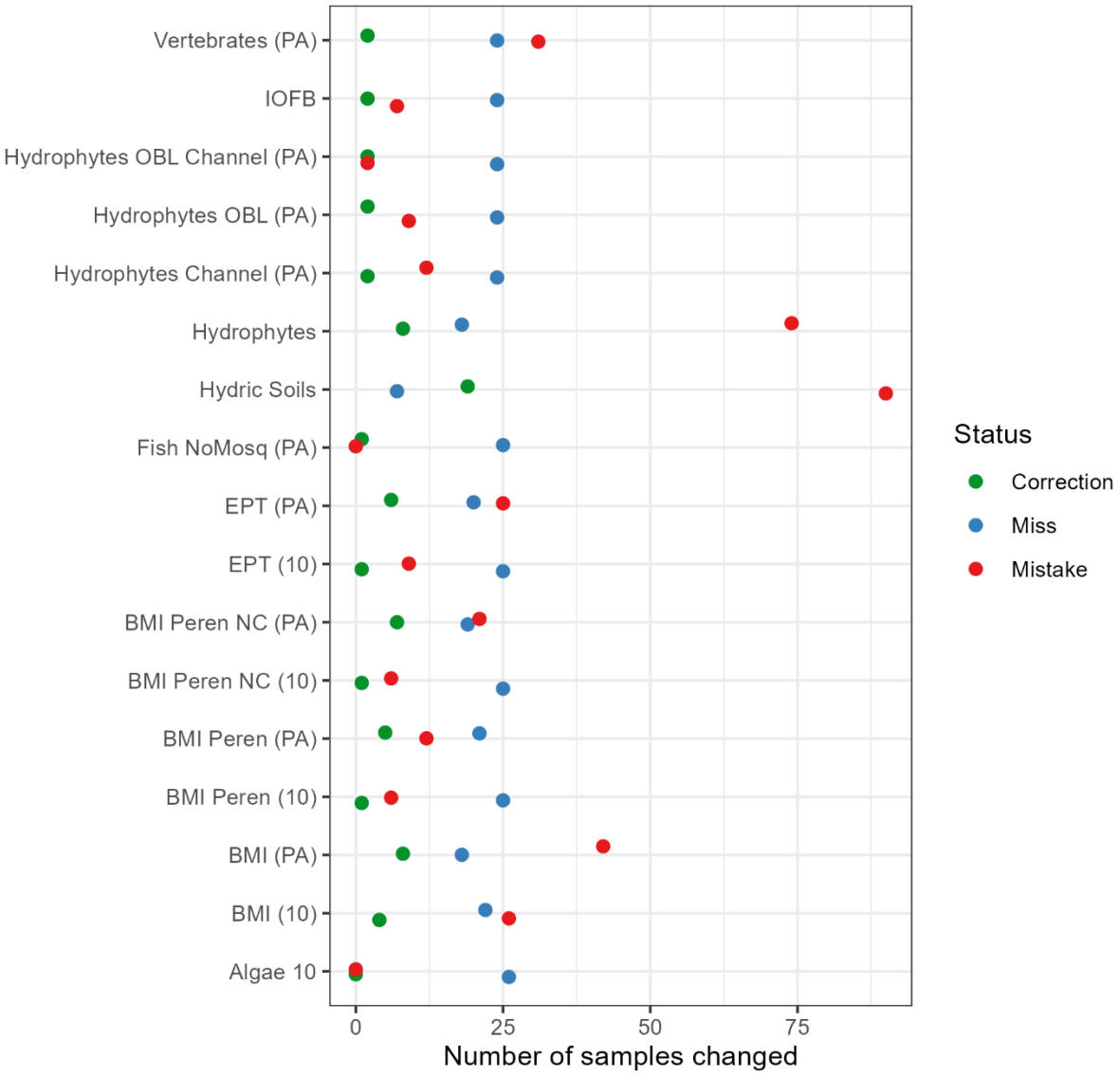


Figure 54. Impact of including single indicators in the Northeast V7 refined model. Green dots count the number of corrections, which were generally fewer than the number of mistakes introduced (red dots). Missed opportunities (blue dots) indicate cases where the single indicator could have helped improve performance but did not.

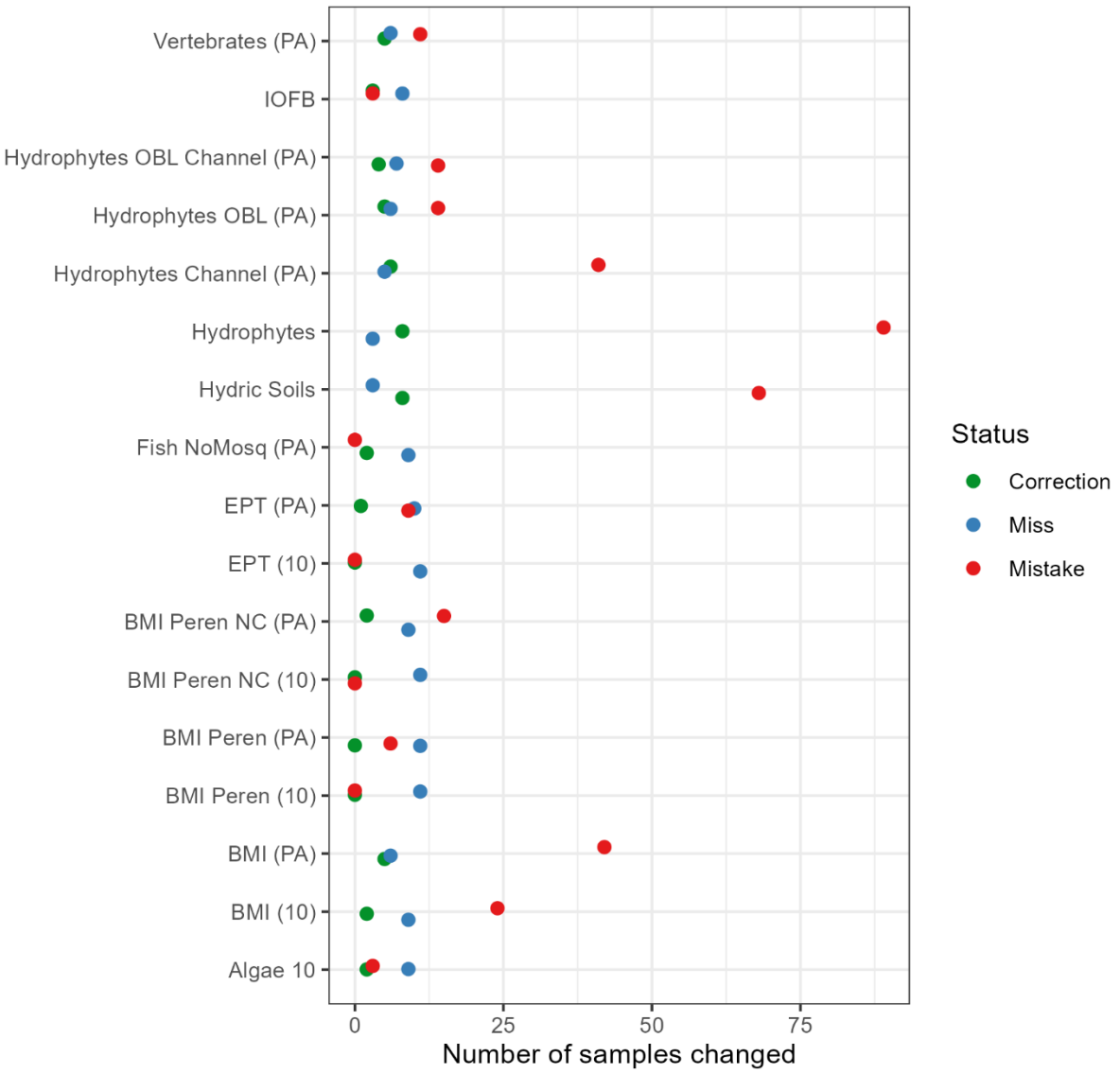


Figure 55. Impact of including single indicators in the Southeast V7 refined model. Green dots count the number of corrections, which were generally fewer than the number of mistakes introduced (red dots). Missed opportunities (blue dots) indicate cases where the single indicator could have helped improve performance but did not.

5.7 Performance assessment of the final models

The Northeast Model includes 7 metrics, 2 of which are GIS metrics. The final model for the Southeast includes 10 metrics, three of which are GIS metrics (ppt.567 was added during model refinement). In both the NE and SE, there were no ephemeral reaches that were classified as perennial, nor were there perennial reaches that were classified as ephemeral ("Big Errors").

The performance of the final Northeast model when evaluated with the test data was an overall PvIvE accuracy of 0.73 (Table 56). EvALI accuracy was significantly higher, with a final testing accuracy of 0.86. PvLTP testing accuracy was also higher than PvIvE (0.84 in the testing dataset). The performance of the final Southeast model when evaluated with the test data was an overall

PvIvE accuracy of 0.67. EvALI accuracy was significantly higher, with a final testing accuracy of 0.91. PvLTP testing accuracy was also higher than PvIvE (0.75 in the testing dataset).

Table 56. Performance of the final NE and SE SDAMs. Accuracy measurements were not weighted because the number of visits per reach were relatively consistent.

Model	Dataset	EvALI	PvLTP	PvIvE	EvIdry	PvIwet	Precision EvALI	Precision PvIvE
NE	Training	0.95	0.93	0.88	0.88	0.96	0.93	0.82
NE	Testing	0.88	0.84	0.73	0.73	0.93	0.93	0.90
SE	Training	0.96	0.96	0.92	0.94	0.98	0.94	0.86
SE	Testing	0.91	0.75	0.67	0.82	0.92	0.94	0.85

Out of 706 reach visits in the NE across testing and unaugmented training datasets, only 30 of them resulted in ambiguous predictions (29 at least intermittent, and 11 less than perennial), as did only 21 of 506 reach visits in the SE (12 at least intermittent and 9 less than perennial). No reach visits in either region resulted in need more information (Table 57).

Table 57. NE and SE final model performance. Green-shaded cells indicate correct predictions for the PvIvE accuracy measure. Blue-shaded cells indicate additional predictions that are correct for the EvALI accuracy measure. Orange-shaded cells indicate additional predictions that are correct for the PvLTP accuracy measure. Purple-shaded cells indicate "big errors" (i.e., ephemeral sites classified as perennial, or perennial sites classified as ephemeral). Pink shaded cells are incorrect for all accuracy measures but are not considered "big errors".

Region	Set	Predicted class	E	I	P
NE	Training	E	161	18	0
NE	Training	I	16	368	27
NE	Training	P	0	13	254
NE	Training	ALI	1	6	18
NE	Training	LTP	3	1	2
NE	Training	NMI	0	0	0
NE	Testing	E	22	1	0
NE	Testing	I	10	58	13
NE	Testing	P	0	5	21
NE	Testing	ALI	0	3	1
NE	Testing	LTP	1	4	0
NE	Testing	NMI	0	0	0
SE	Training	E	137	5	0
SE	Training	I	14	289	13
SE	Training	P	0	4	185
SE	Training	ALI	1	4	4
SE	Training	LTP	2	3	1
SE	Training	NMI	0	0	0
SE	Testing	E	15	0	0
SE	Testing	I	6	37	17
SE	Testing	P	0	4	16
SE	Testing	ALI	0	0	3
SE	Testing	LTP	1	1	1
SE	Testing	NMI	0	0	0

5.7.1 Accuracy at disturbed sites

Disturbances were noted at 102 (14%) site visits in the NE and at 50 (9%) site visits in the SE. In the NE, disturbed site visits were flagged agricultural 30 times, urban 36 times, and developed 45 times. In the SE, disturbed site visits were flagged agricultural 35 times, urban 3 times, and developed 13 times. Accuracy was comparable in disturbed sites compared to undisturbed sites (Table 58).

Table 58. Comparison of accuracy at disturbed and undisturbed reach visits. Training performance does not include oversampled data.

Model	Set	Measure	# disturbed reaches	# disturbed visits	Disturbed accuracy	# undisturbed reaches	# undisturbed visits	Undisturbed accuracy	Difference in accuracy (Disturbed minus undisturbed)
NE	Testing	EvALI	8	16	0.94	36	123	0.88	0.06
NE	Testing	PvLTP	8	16	1	36	123	0.82	0.18
NE	Testing	PvIvE	8	16	0.94	36	123	0.7	0.24
NE	Training	EvALI	51	86	0.99	140	481	0.92	0.07
NE	Training	PvLTP	51	86	0.87	140	481	0.88	-0.01
NE	Training	PvIvE	51	86	0.87	140	481	0.8	0.07
SE	Testing	EvALI	6	11	0.64	28	90	0.92	-0.28
SE	Testing	PvLTP	6	11	0.82	28	90	0.77	0.05
SE	Testing	PvIvE	6	11	0.45	28	90	0.7	-0.25
SE	Training	EvALI	19	39	1	110	416	0.94	0.06
SE	Training	PvLTP	19	39	0.97	110	416	0.94	0.03
SE	Training	PvIvE	19	39	0.97	110	416	0.88	0.09

5.7.2 Sensitivity to variation in indicator values

Thresholds for small versus large changes were determined for each indicator included in the final NE and SE models (Table 59).

Table 59. Threshold for large and small sensitivity.

Indicator	Threshold for large decreases	Threshold for large increases
DRNAREA_mi2	<50% of original value	>150% of original value
Elev_m	*-55m* of original value	*+55m* of original value
PctShading	<50% of original value	>150% of original value
BMI_score_alt4	-1.5 bins from original value	+1.5 bins from original value
BankWidthMean (NE only)	<50% of original value	>150% of original value
BankWidthMean_10_bin (SE only)	-1 bins from original value	+1 bins from original value
Slope_10_bin	-1 bins from original value	+1 bins from original value
fp_entrenchmentratio_mean_1.5_2_bin (NE only)	-1 bins from original value	+1 bins from original value
UplandRootedPlants_score (SE only)	-1.5 bins from original value	+1.5 bins from original value
SubstrateSorting_score (SE only)	-1.25 from original value	+1.25 from original value
FibrousRootedPlants_score (SE only)	-2.5 bins from original value	+2.5 bins from original value
TotalAbundance_0_3_40_bin (SE only)	-1 bins from original value	+1 bins from original value
ppt.567 (SE only)	-10% of original value	+10% of original value

The NE SDAM has very high sensitivity to 2 indicators, high sensitivity to 2 indicators, and moderate sensitivity to 3 indicators. The SE SDAM has high sensitivity to 3 indicators, moderate sensitivity to 2 indicators, and low sensitivity to 5 indicators (Table 60).

Table 60. Sensitivity of indicators in the final NE and SE SDAMs.

Indicator	Model	Large decrease	Small decrease	Small increase	Large increase	Sensitivity
BankWidthMean	SE	Low	Low	Low	Low	Low
BankWidthMean_10_bin	NE	N/A	Very High	Very High	N/A	Very High
BMI_score_alt4	NE	High	High	High	High	High
BMI_score_alt4	SE	Low	Low	Low	Moderate	Moderate
DRNAREA_mi2	NE	High	High	High	Moderate	High
DRNAREA_mi2	SE	Moderate	Low	Low	Small	Moderate
Elev_m	NE	Moderate	Low	Low	Moderate	Moderate
Elev_m	SE	Low	Low	Low	Low	Low
FibrousRootedPlants_score	SE	Low	Low	Low	Low	Low
fp_entrenchmentratio_mean_1.5_2_bin	NE	Moderate	Low	Low	Moderate	Moderate
PctShading	NE	Moderate	High	Low	Low	Moderate
PctShading	SE	Low	Low	Low	Low	Low
ppt.567	SE	High	High	High	High	High
Slope_10_bin	NE	N/A	Very High	Very High	N/A	Very High
SubstrateSorting_score	SE	Moderate	Low	Low	Low	Low
TotalAbundance_0_3_40_bin	SE	High	High	High	High	High
UplandRootedPlants_score	SE	High	Moderate	Moderate	High	High

The following figures (Figure 56, Figure 57) show the relative sensitivity of each indicator. In each figure, the solid vertical lines indicate no change in the indicator value. Dashed vertical lines delineate the threshold between large and small changes in indicator value. Horizontal dashed lines indicate thresholds between small, moderate, and large impacts on the percent of reach visits with unchanged classifications.

5.7.2.1. Northeast Sensitivity Analysis Figures

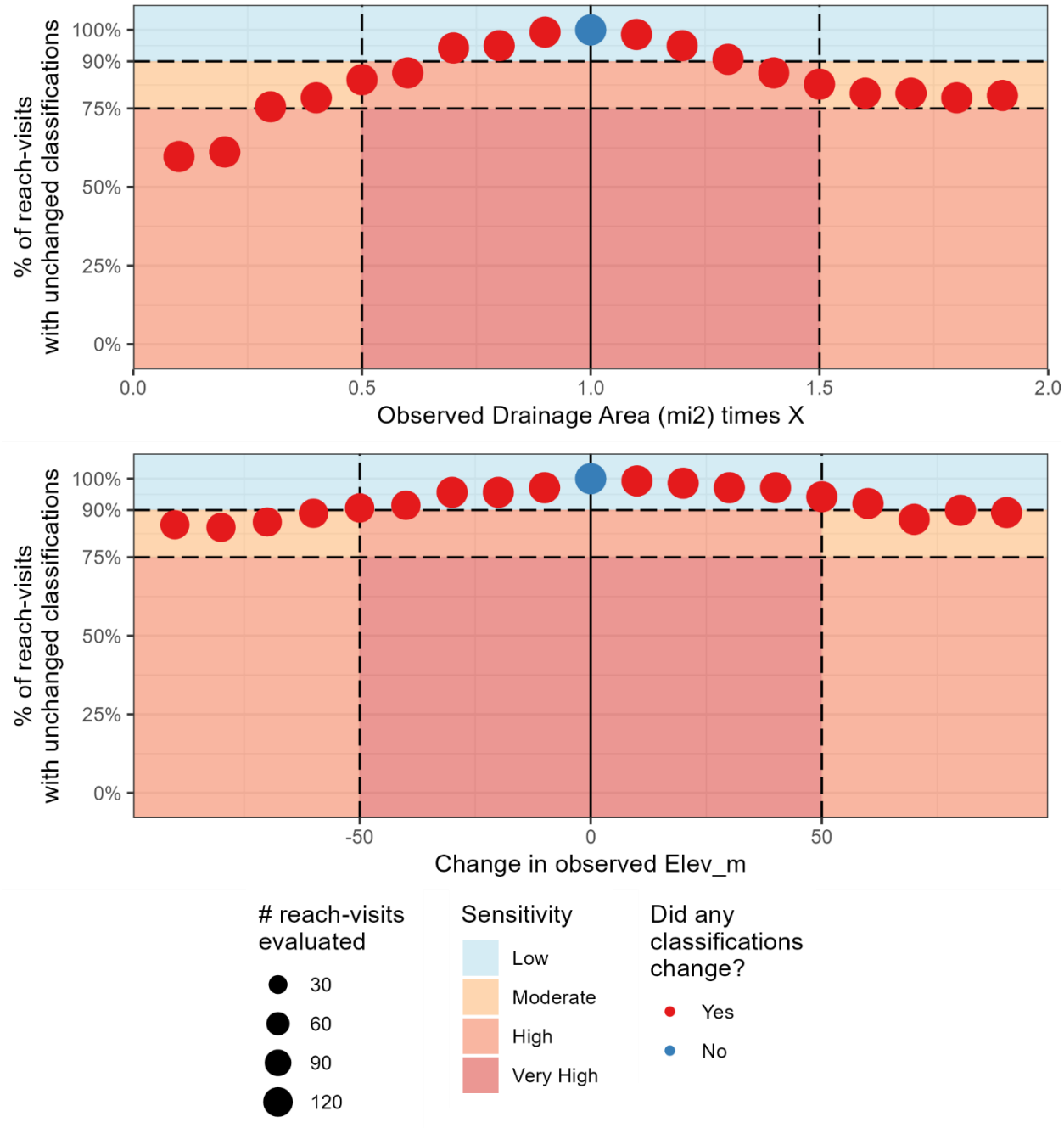


Figure 56a. Sensitivity of the NE SDAM to changes in indicator values for drainage area (top) or elevation (bottom).

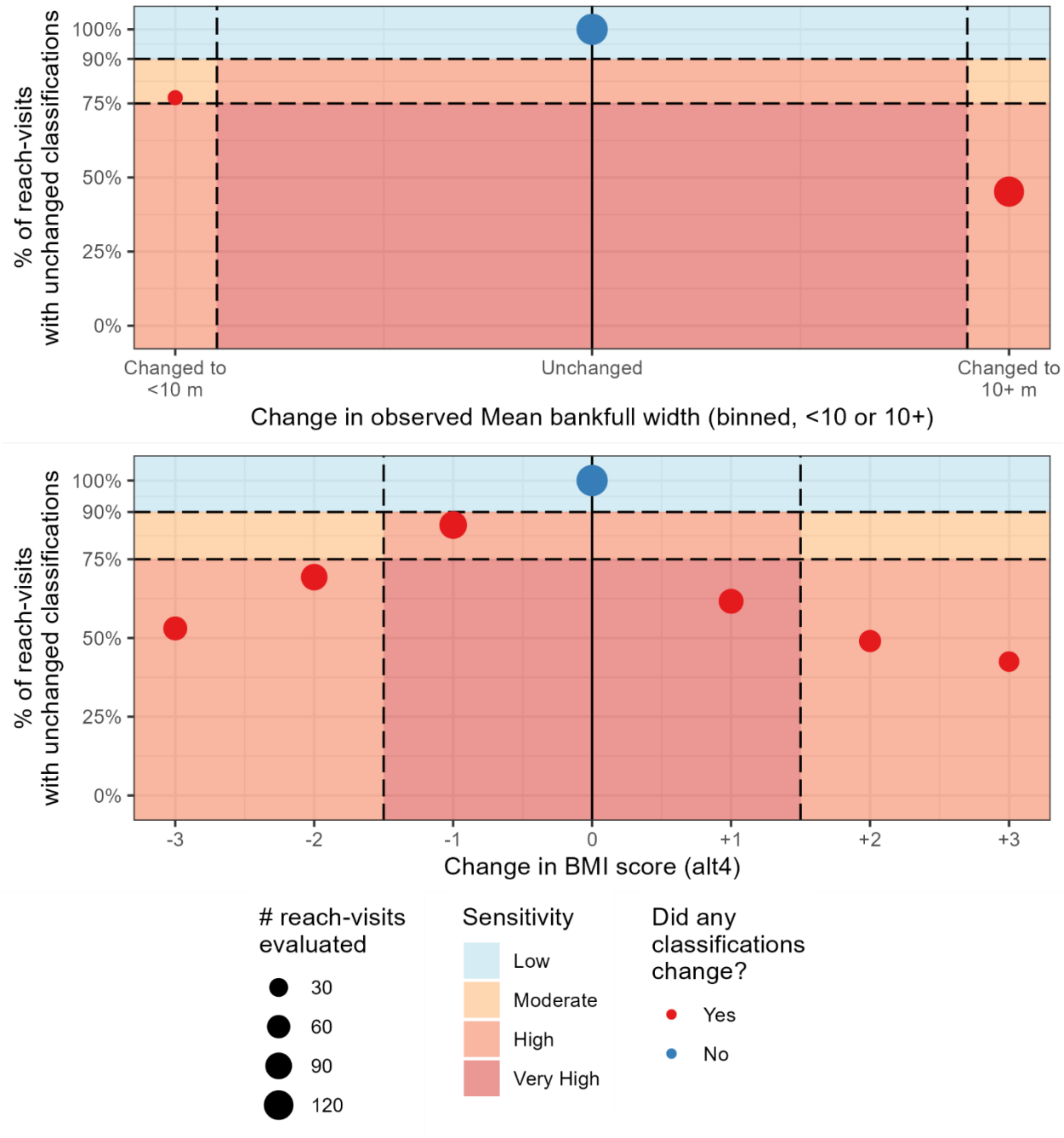


Figure 57ab. Sensitivity of the NE SDAM to changes in indicator values for mean bankfull width (top) or BMI score (bottom).

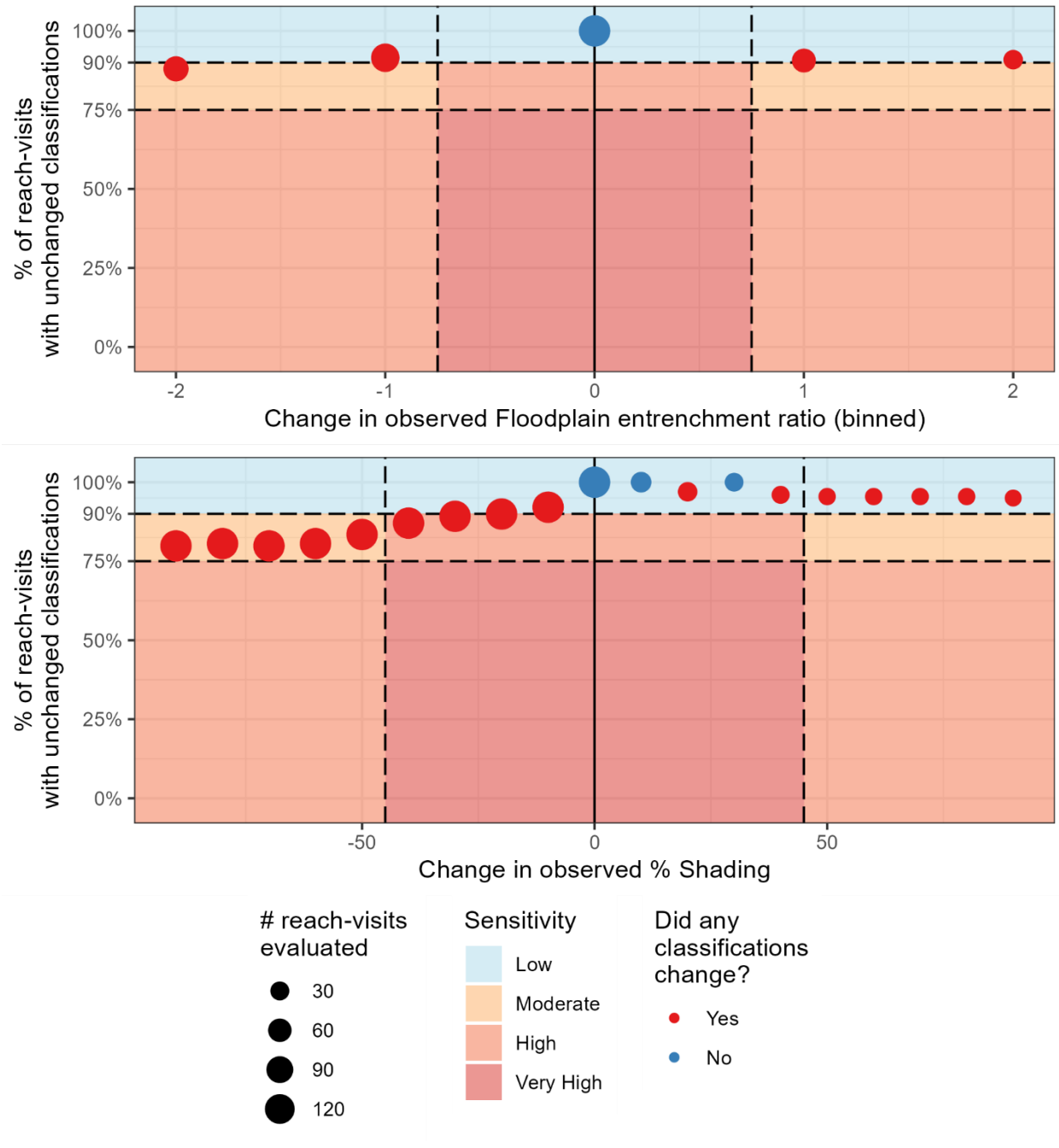


Figure 57ac. Sensitivity of the NE SDAM to changes in indicator values for fp_entrenchmentratio (top) or % shading (bottom).

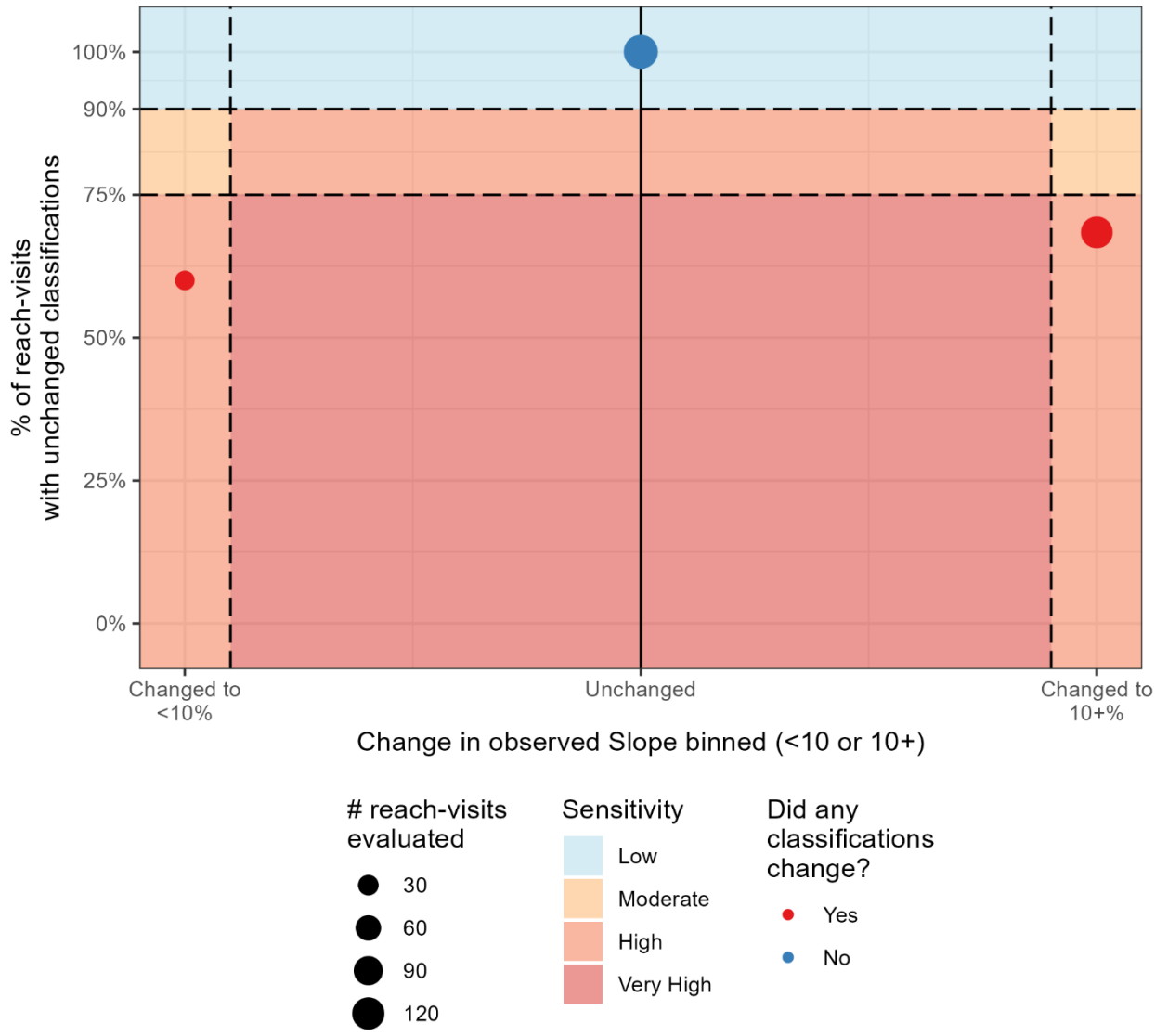


Figure 57ad. Sensitivity of the NE SDAM to changes in indicator values for slope.

5.7.2.2. Southeast Sensitivity Analysis Figures

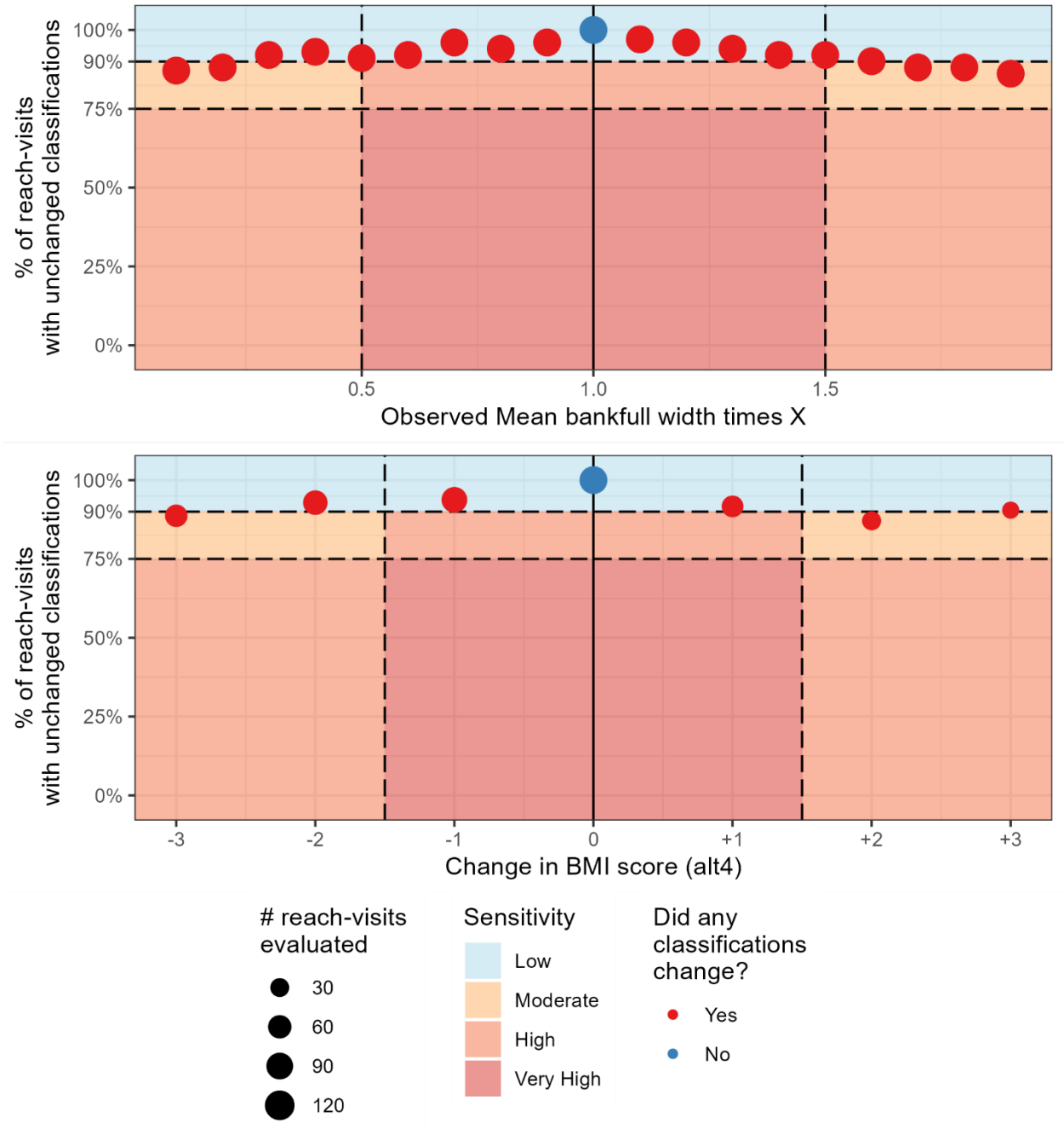


Figure 57a. Sensitivity of the SE SDAM to changes in indicator values for mean bankfull channel width (top) and BMI score (bottom).

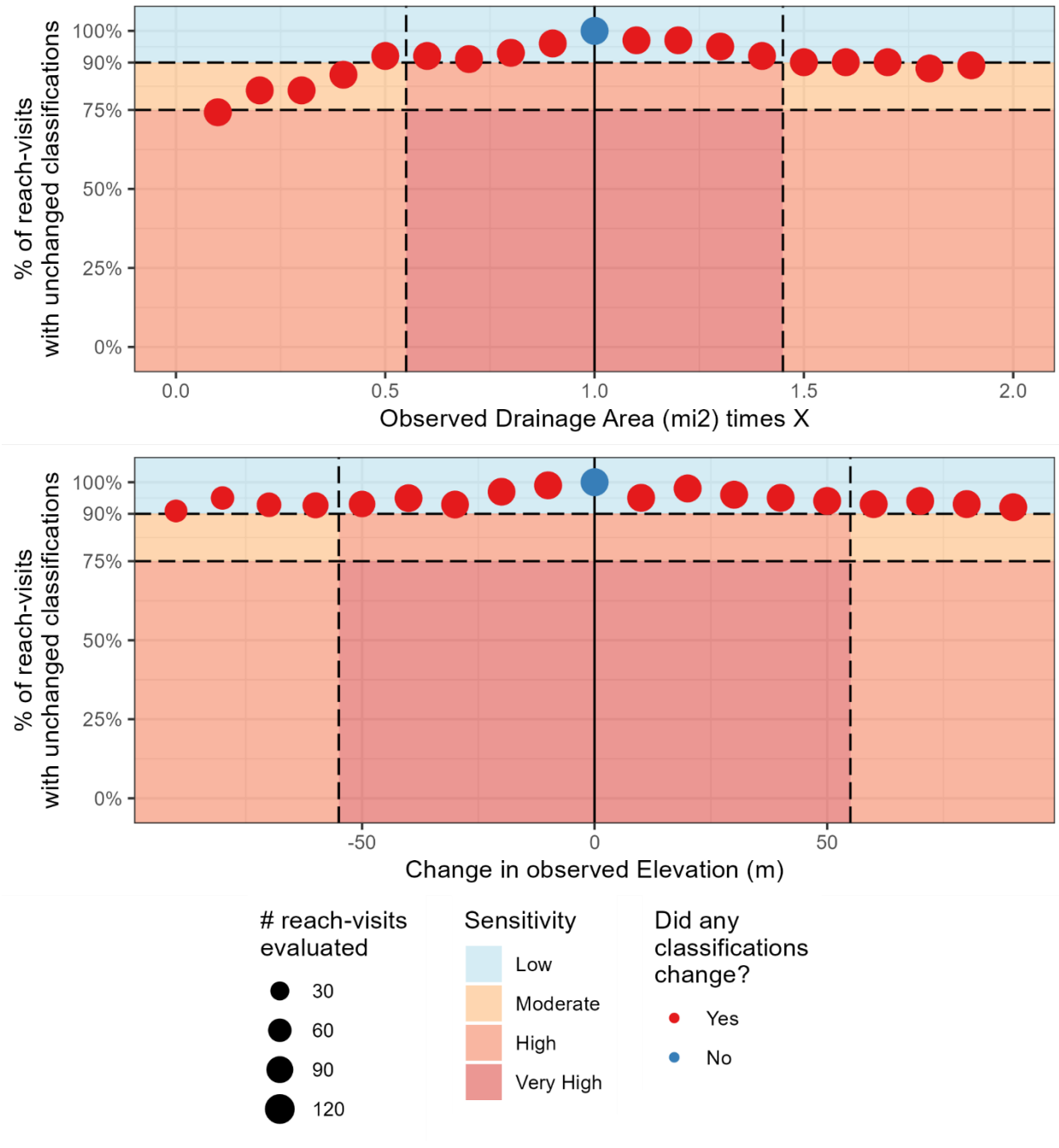


Figure 57b. Sensitivity of the SE SDAM to changes in indicator values for drainage area (top) and elevation (bottom).

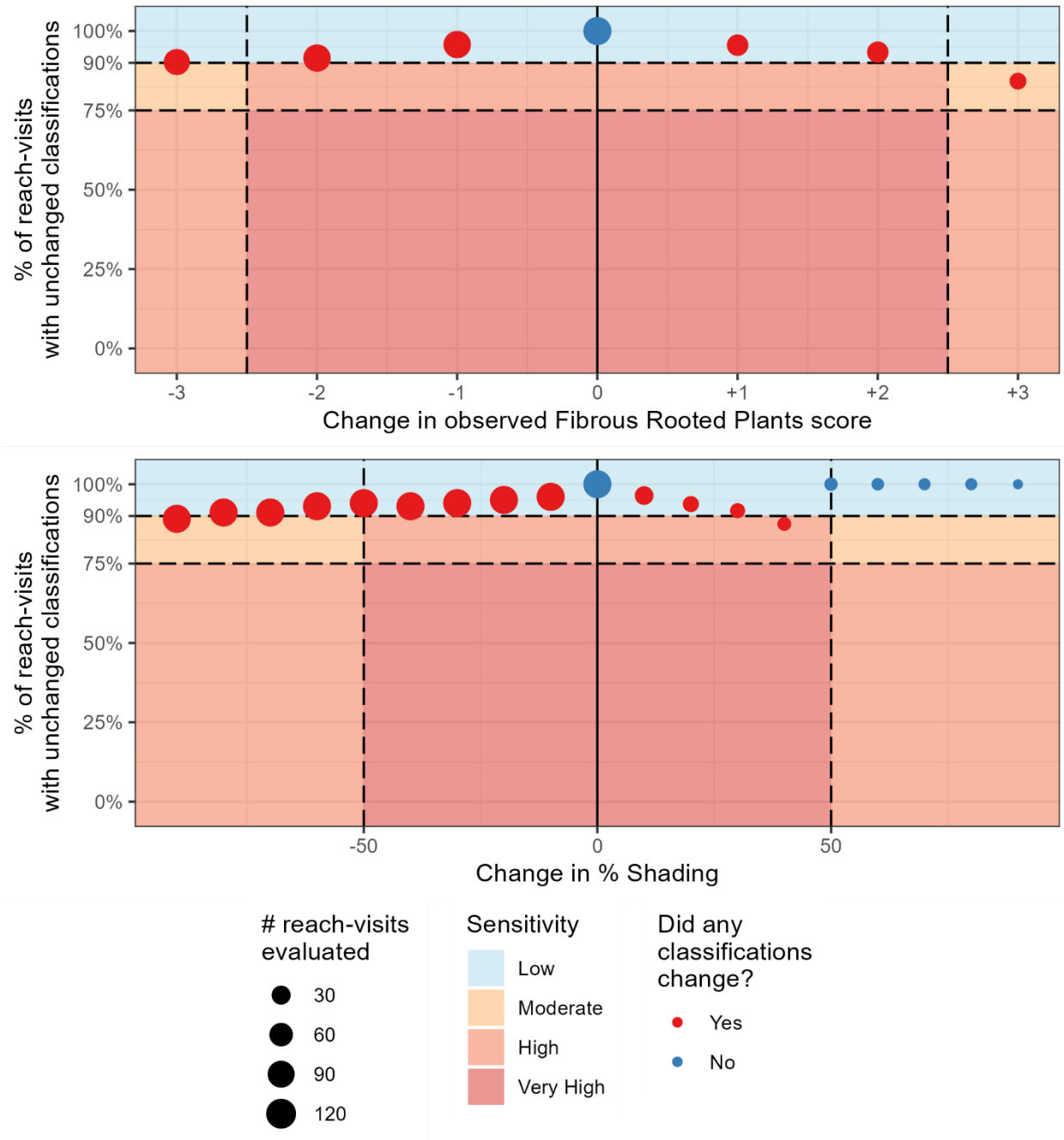


Figure 57c. Sensitivity of the SE SDAM to changes in indicator values for fibrous rooted plants score (top) and % shading (bottom).

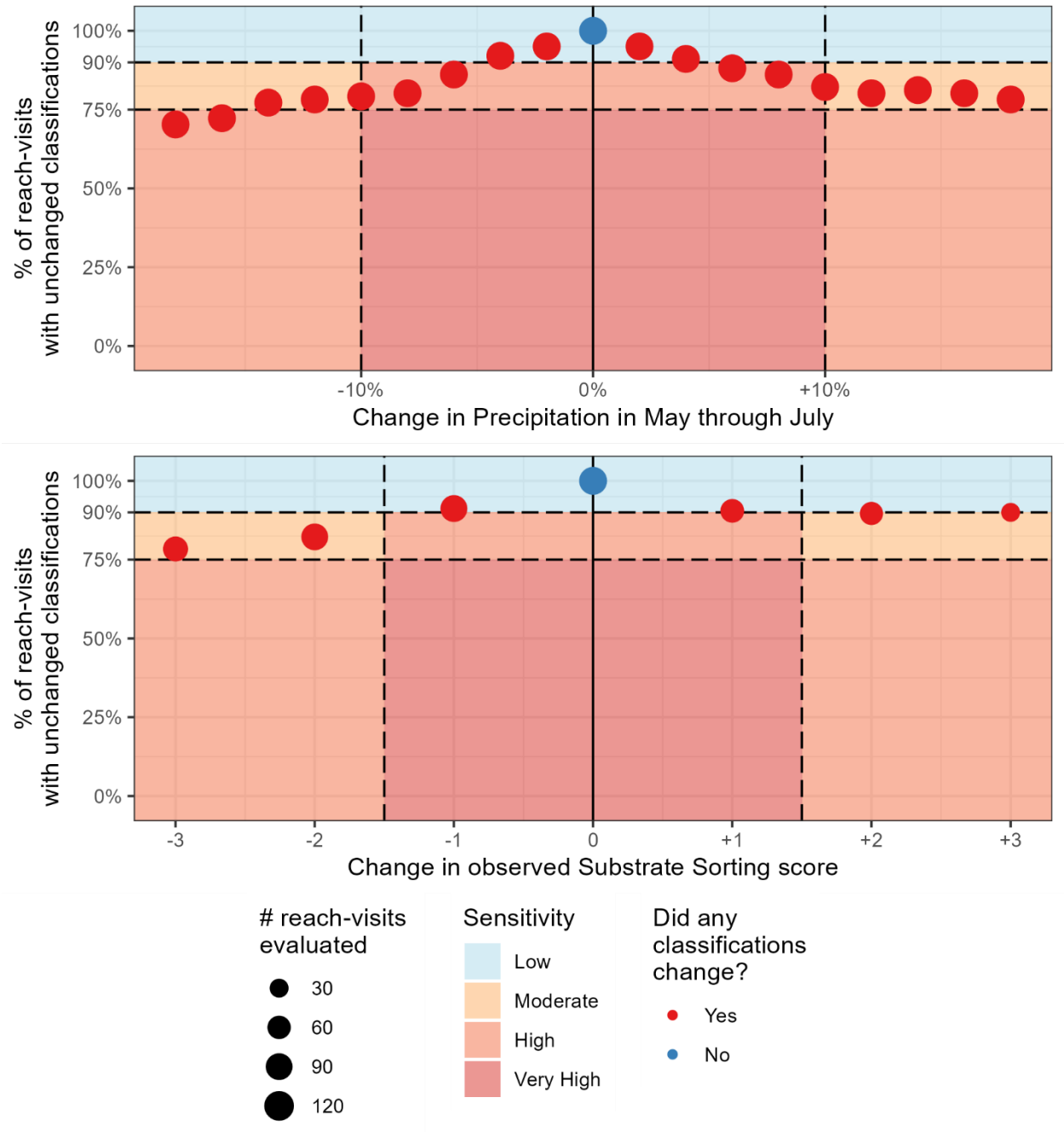


Figure 57d. Sensitivity of the SE SDAM to changes in indicator values for precipitation between May and July (top) and substrate sorting score (bottom).

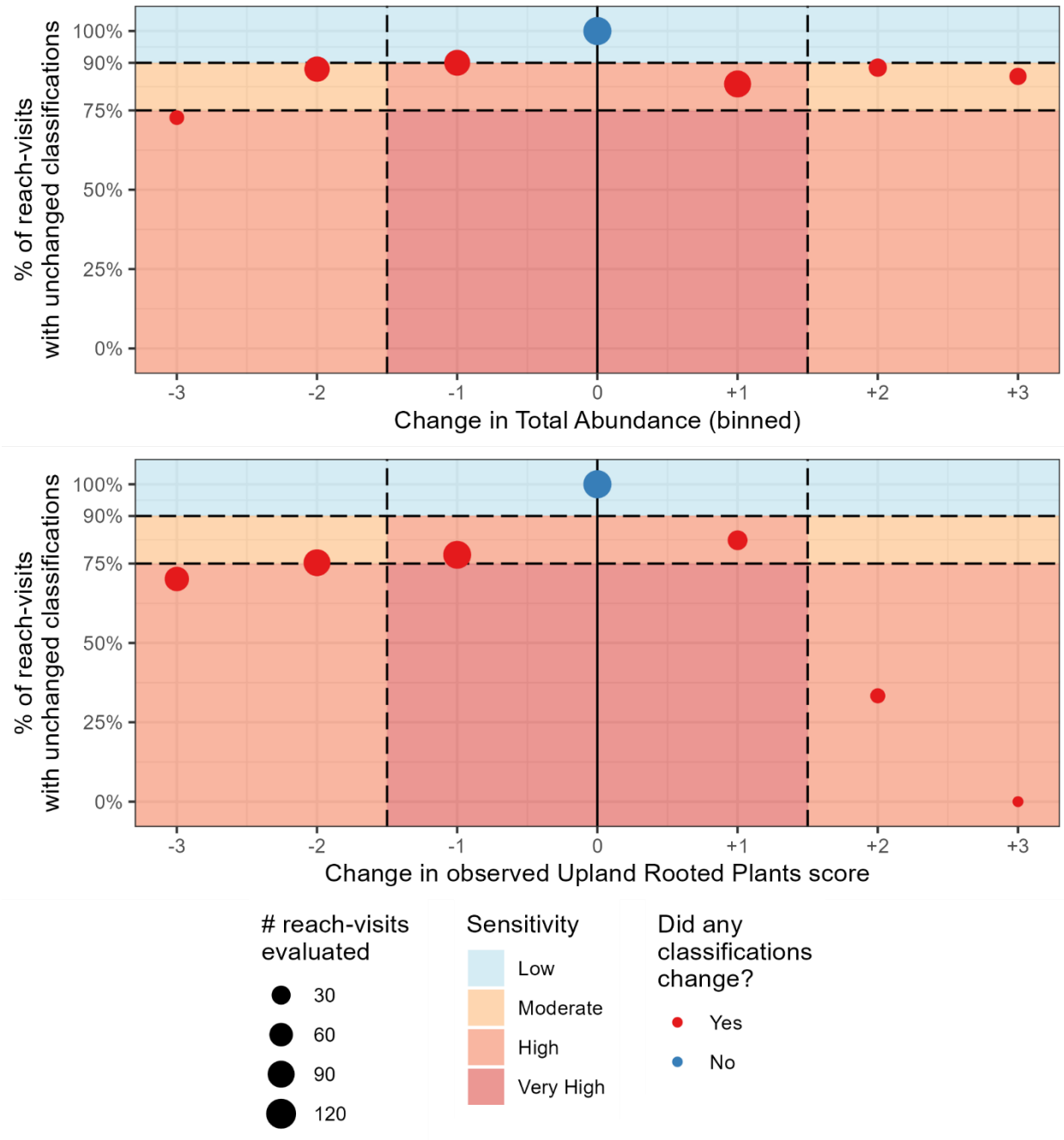


Figure 57e. Sensitivity of the SE SDAM to changes in indicator values for Total abundance score (top) and Upland rooted plants score (bottom).

5.7.3 Sensitivity to time of sampling

Seasonal patterns were relatively consistent across the NE and SE based on reach-specific long-term monthly climatic statistics shown in Appendix B: Indicators and metrics evaluated for SDAM development (Figure 58). Temperature followed a similar pattern in both regions with the warmest months typically occurring in the summer (May to October). Based on these patterns, a consistent seasonal classification was applied to all reaches in the NE and SE (Summer = May-

Oct, Winter = Nov-Apr). Precipitation did not show seasonal patterns that held up across sites in either region. Precipitation was relatively consistent throughout the year in the NE. In the SE summer saw higher precipitation in some sites, while lower precipitation in others but no overall seasonal trend.

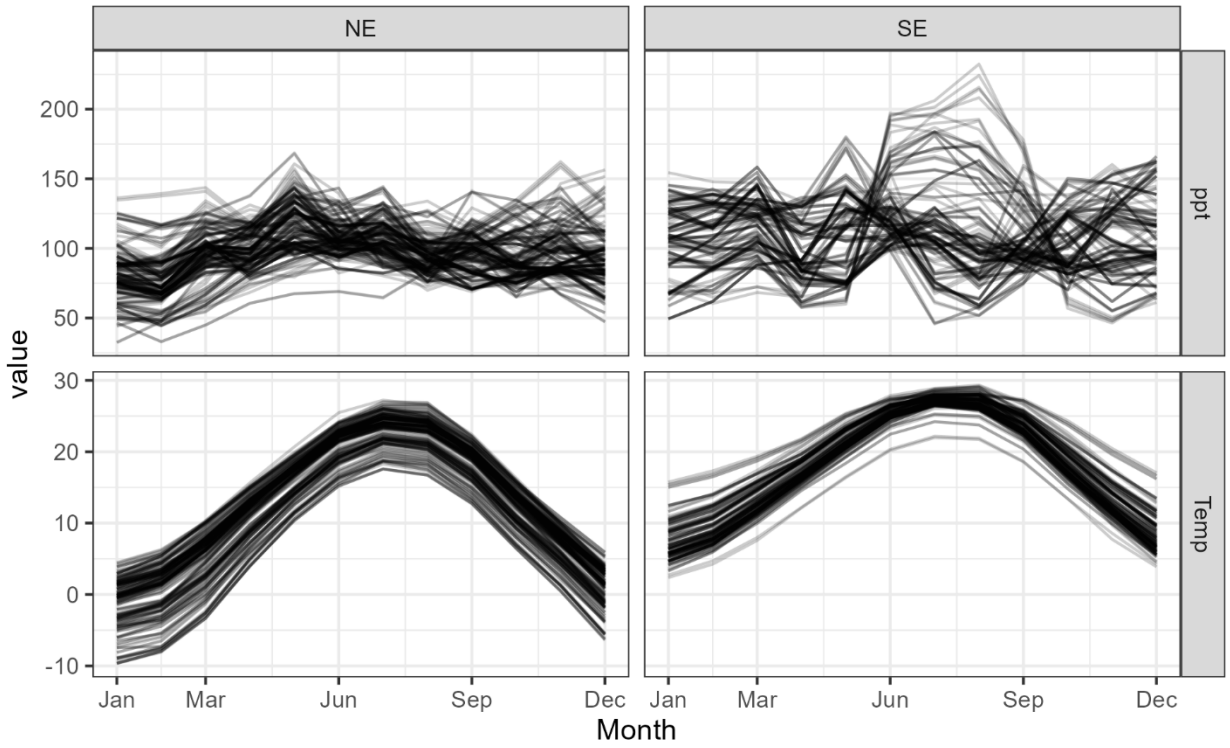


Figure 58. Long-term monthly averages of precipitation (ppt, mm) and temperature (Temp, °C) for NE and SE sites to support approximate "summer" and "winter" delineations (Summer = May-Oct, Winter = Nov-Apr).

There were more reach visits during the Winter than Summer in the NE and SE, especially in the testing datasets (Table 61). Using WebWIMP derived wet and dry seasons, the number of NE reach visits during the wet season was 2.3 and 3.6 times as many as during the dry season within the training and testing sets, respectively (Table 61). There were more NE reach visits classified with normal antecedent precipitation (59.8% training and 64.0% testing) than with wetter than normal (25.6% training and 20.1% testing) or drier than normal antecedent precipitation (14.6% training and 15.8% testing). The APC classes for the SE reach visits were similarly distributed among normal (63.9% training and 58.4% testing), wetter than normal (19.9% training and 23.8% testing) and drier than normal (16.2% training and 17.8% testing). The SE reach visits were also skewed between seasons with 85.0% and 86.1% of training and testing visits, respectively, occurring during the wet season. The inferential power for making accuracy comparisons between seasons and APC classes is limited by the large differences in number of reach visits, especially for the testing sets.

Table 61. Number of reach visits in each season as defined by temperature or precipitation, WebWIMP (WW) and Antecedent Precipitation Condition (APC) in the Northeast (NE) and Southeast (SE) regions training and testing datasets.

Region	Time of sampling comparisons	Classes	Dataset	
			Train	Test
Northeast (NE)	Weather-based seasons (WBS)	Summer	418	43
		Winter	470	96
	WebWIMP season (WW)*	Dry	260	30
		Wet	607	109
	Antecedent Precipitation Condition (APC)	Normal	531	89
		Drier than normal	130	22
		Wetter than normal	227	28
Southeast (SE)	Weather-based seasons (WBS)	Summer	192	25
		Winter	470	76
	WebWIMP season (WW)	Dry	99	14
		Wet	563	87
	Antecedent Precipitation Condition (APC)	Normal	423	59
		Drier than normal	132	24
		Wetter than normal	107	18

*21 reach visits were unclassified by WW because they we located near large water bodies.

Consistent with accuracy across the entire datasets, training and EvALI accuracy (and to less extent PvLTP accuracy) was higher than testing and PvIvE accuracy, respectively, for all time of sampling comparisons (Figure 59 and Figure 60). The SE SDAM was consistently more accurate during the Summer compared to Winter for visits in the SE training data set, but the accuracy of the NE SDAM was comparable for visits during Summer and Winter in the NE training set (Figure 59). Based on the testing data sets, the NE and SE SDAMs performed equally well in EvALI accuracy but had higher PvIvE and PvLTP accuracy during Winter than in Summer (Figure 59).



Figure 59. Accuracy of the Northeast (NE) and Southeast (SE) SDAMs in weather-based seasons (WBS; Summer = May to October and Winter = November to April).

Accuracy of the NE and SE SDAMs was comparable for Dry and Wet season reach visits, except the visits in SE SDAM test dataset had higher PvIvE accuracy for Dry season reach visits than for Wet season reach visits (Figure 60). The EvALI testing accuracy in NE and SE was slightly higher during NC reach visits than DTN or WTN reach visits. Except for testing accuracy for EvALI in the SE, PvIvE and PvLTP for NC reach visits was not higher than accuracy for DTN or WTN reach visits (Figure 60).

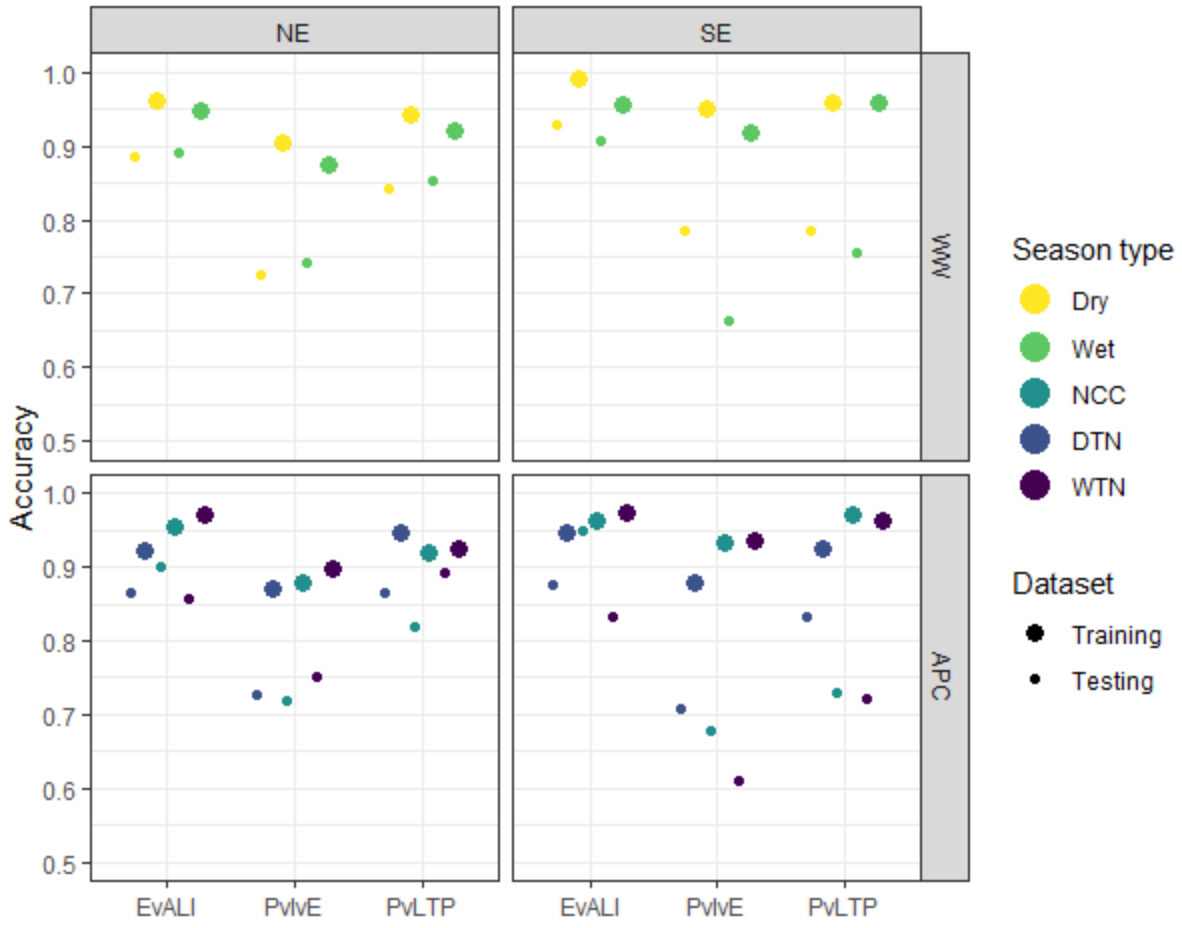


Figure 60. Accuracy of the Northeast (NE) and Southeast (SE) SDAMs in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN).

The proportion misclassifications among the NE and SE reach visits in the training datasets were evenly distributed between predicting longer and shorter streamflow duration classes for Dry and Wet seasons and APCs (Figure 61). For the NE testing dataset, there was a greater proportion of EvALI errors predicting longer than shorter streamflow duration classes for reach visits in the Dry and Wet season, but only for reach visits during the Wet season for the SE. The SE testing dataset also had a slightly greater proportion of PvIvE and PvLTP errors predicting short rather than long streamflow duration classes for reach visits during the Wet and Dry seasons (Figure 61). Apart from EvALI for the SE testing datasets, the proportion of misclassifications in the SE predicting shorter streamflow duration classes was not greater during the Dry season than the Wet season. The proportion of PvIvE and PvLTP misclassifications predicted as short streamflow duration classes was higher than the proportion predicted as long streamflow duration classes when APC was Drier than normal for NE testing reach visits. The proportion of PvIvE and PvLTP misclassifications predicted as short streamflow

duration classes was also higher than the proportion predicted as long streamflow duration classes when APC was NC for SE testing reach visits. There were also no consistent patterns regarding the proportion of misclassifications toward short or long streamflow duration classes based on the APC of reach visits in the training dataset (Figure 61).

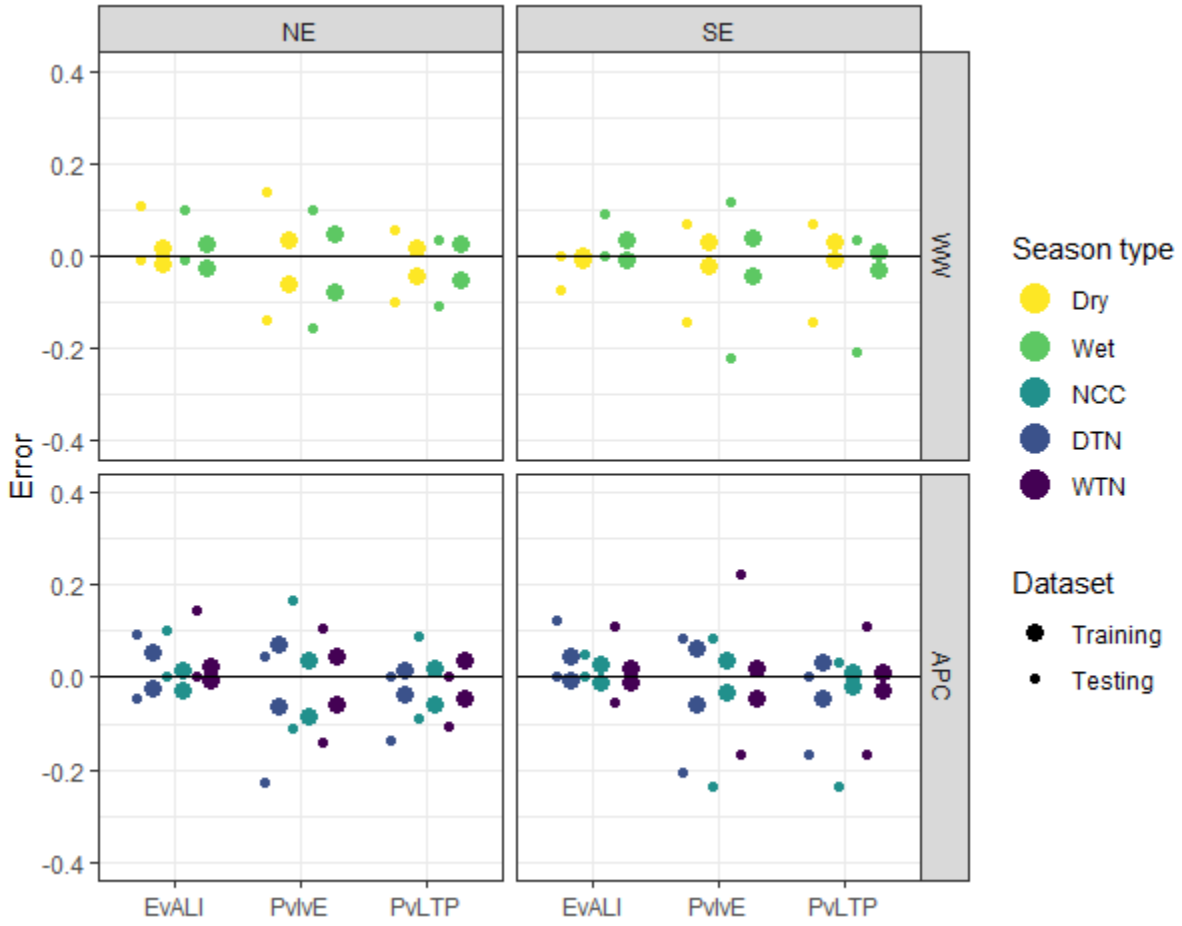


Figure 61. Error by direction of the Northeast (NE) and Southeast (SE) SDAMs in Dry and Wet seasons based on the Antecedent Precipitation Tool use of WebWIMP (WW) estimations for precipitation, snowmelt, and potential evapotranspiration and antecedent precipitation condition (APC) classification of training and testing reach visits as Normal climate condition (NCC), Drier than Normal (DTN), and Wetter than Normal (WTN). Negative values represent the proportion of error predicting shorter flow duration classifications and positive values represent the proportion of error predicting longer flow duration classifications.

5.7.4 Comparison to other SDAMs

5.7.4.1 Comparison to other SDAMs and protocols used within the East or adjacent regions

In Figure 62 performance of the final NE and SE SDAMs are compared to adjacent SDAMs and protocols using the final NE and SE testing datasets (or subsets of those datasets).

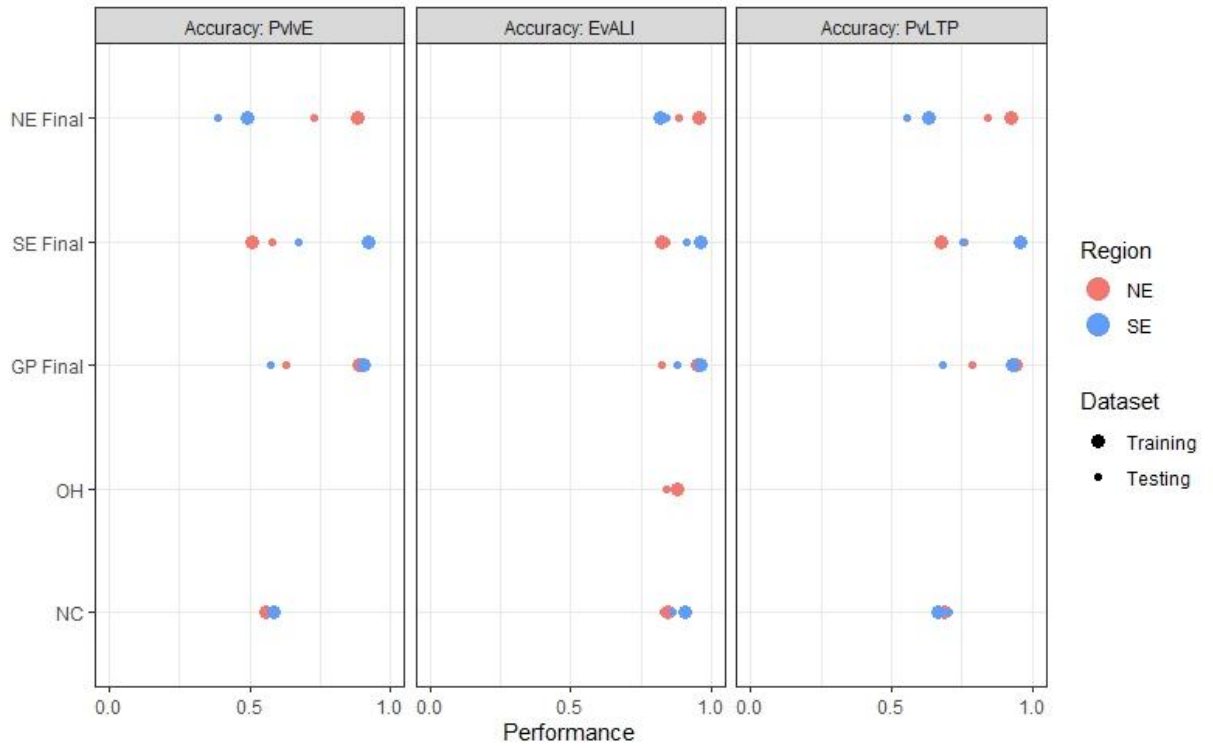


Figure 62. Comparison of the final NE and SE SDAMs to adjacent methods. Red dots represent performance measures calculated from reaches in the NE dataset, and the blue dots represent performance measures calculated from reaches in the SE dataset. The OH protocol was applied to sites in the NE and only evaluated for EvALI accuracy because that method does not distinguish between perennial versus intermittent reaches.

5.7.4.2 Comparison between the NE SDAM and the OH or NC protocols at sites in OH or NC

We compared the accuracy of the final NE SDAM to the OH protocol at sites within OH, as well as to the NC SDAM at sites within NC. This comparison is intended to help practitioners understand the relative performance of the SDAMs available for use in these states. For these analyses, results for both the training and testing datasets are presented (Table 62, Figure 63). The NE SDAM had better accuracy than the NC method. The dataset included no sites in NC that were within the SE, and thus no comparisons to the SE SDAM within NC were possible.

Table 62. Number of reaches within OH and the NE portion of NC available for comparison of the NE SDAM and the NC and OH protocols.

Region	Set	E	I	P	Total
NC (NE portion)	Training	3	3	1	7
NC (NE portion)	Testing	0	0	0	0
OH	Training	6	6	4	16
OH	Testing	2	0	2	4

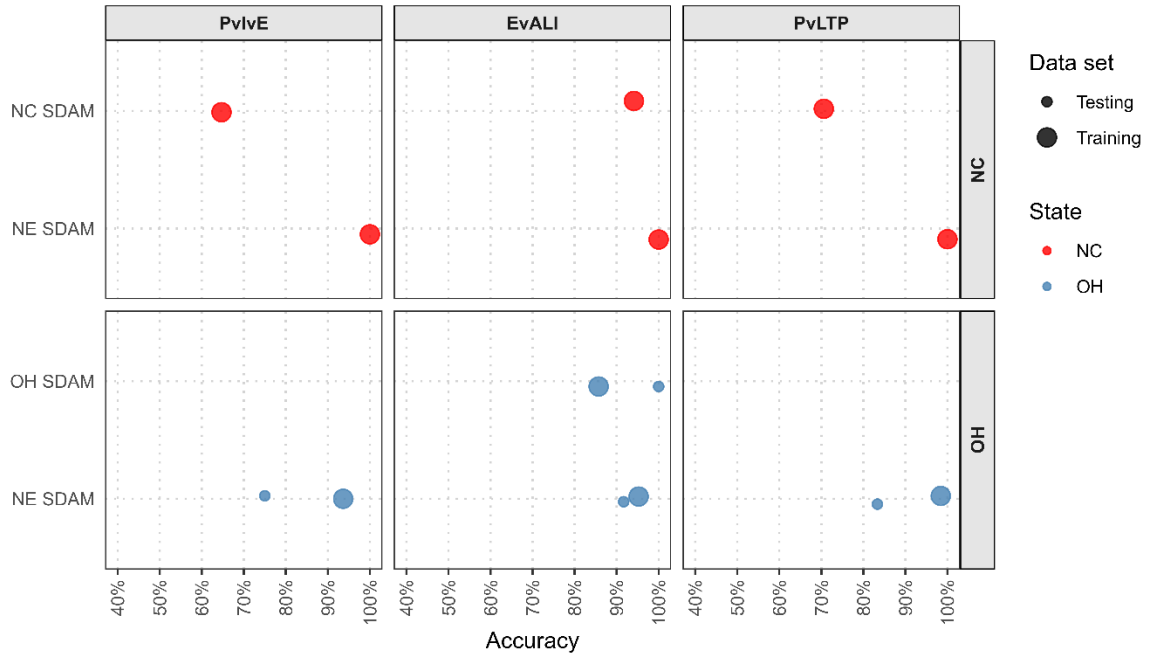


Figure 63. Accuracy of the NE SDAM and the NC and OH protocols at sites within OH and the NE portion of NC.

5.7.4.3 Comparison to NE and SE beta SDAMs

The beta NE SDAM had PvIvE accuracy of 0.68 and an EvALI accuracy of 0.9 (Gross et al. 2023), compared to final NE PvIvE accuracy of 0.73 and EvALI accuracy of 0.89. The beta SE SDAM had PvIvE accuracy of 0.64 and an EvALI accuracy of 0.9 (Gross et al. 2023), compared to final SE PvIvE accuracy of 0.67 and EvALI accuracy of 0.91. PvLTP accuracy was not reported for the beta NE and SE methods. The beta PvIvE accuracies reported (in the National Data Supplement) have been adjusted from the original values reported in 2023 versions of the beta data supplement reports (Eddy et al. 2023 or Gross et al. 2023) to be calculated consistent with the descriptions in section 2.8.4.1 Accuracy.[2]

Performance statistics for the beta methods were calculated from different testing datasets compared to the final SDAMs, which could affect the comparison. Additionally, the beta NE and SE methods were developed from smaller datasets. The NE beta method had 15 fewer reaches and 184 fewer samples prior to rebalancing (oversampling). The SE beta method had 7 fewer reaches and 162 fewer samples prior to rebalancing. In the beta methods, rebalancing (oversampling) targeted 4 samples per reach, while the target for the final methods was 5 per reach.

[2] In development of the beta methods ALI was counted as correct for PvIvE accuracy (treating ALI the same as an I result); however, for the final methods ALI was not counted as correct for PvIvE accuracy.

6. Data and code availability

Data and R code used to develop the regional SDAMs are available at the following repository: https://doi.org/10.23719/d-v0jj

7. Acknowledgements

Thank you to the members of the project delivery team who developed the approach for developing SDAMs across the country as quickly as possible and oversaw the launch of this project, data collection, beta method development, final method development and rollout of everything along the way but moved on to other opportunities before this report was developed; this project would not have been possible without you Tracie Nadeau (EPA R10 retired), Julie Kelso (former ORISE participant), Rachel Harrington (EPA Region 8), and Whitney Beck (former EPA HQ).

Thank you to the team of contractors who helped manage and analyze the data, develop reports, training materials, tools and anything else needed to get the job done, including all those whose contributions did not lead to authorship on this report including: Will Saulnier, Serena Gilson, and Abel Santana.

Thank you to the team of contractors who helped identify study sites, calibrated loggers, identified aquatic macroinvertebrates and collected all the field data at over one thousand sites and three thousand site visits across the country identified in each of the corresponding regional SDAMs. Thank you to the numerous researchers and land managers with local expertise who assisted with the selection of study reaches to calibrate each of the SDAMs identified in each of the corresponding regional SDAMs.

Thank you to the Regional Steering Committee members from the US Army Corps of Engineers (USACE) and Environmental Protection Agency (EPA) offices across the country who helped guide the development the SDAMs from the initial identification of indicators and sites to sample in the field, to prioritizing between models and simplification options for indictors and identifying the key training materials and tools needed for implementation.

Agency	**Member(s)**	**SDAM Region(s)**
EPA Headquarters	Rose Kwok	All
EPA Region 1	Stephanie Tougas, Erica Sachs, and Ray Putnam	NE, SE
EPA Region 2	Marco Finocchiaro, Stephanie Andreescu, Sofía Olivero Lora, Katiana Garcia Rosado (ORISE), and Nancy Rodriguez	NE, SE
EPA Region 3	Christine Mazzarella, Greg Pond, and Natalie Motley	NE, SE
EPA Region 4	Eric Somerville, Cynthia F. Van Der Wiele, and Kacy Sable	NE, SE
EPA Region 5	Melanie Burdick	NE, SE
	Kerryann Weaver and Ed Hammer	GP

Agency	**Member(s)**	**SDAM Region(s)**
EPA Region 6	Loribeth Tanner	AW, WM
	Loribeth Tanner and Chelsey Sherwood	GP
	Raul Gutierrez and Chelsey Sherwood	NE, SE
EPA Region 7	Gabriel DuPree, Jeanette Schafer, Blain McNabb, Shawn Henderson, and Seiji Kuwata	GP
	Blain McNabb, Gabriel DuPree, Jeanette Schafer, and Justin Kensinger	NE, SE
EPA Region 8	Rachel Harrington	AW, WM
	Rachel Harrington and Billy Bunch	GP
EPA Region 9	Joe Morgan and Rob Leidy	AW
	Joe Morgan	WM
USACE Headquarters	Matt Wilson and Tunis McElwain	All
USACE ERDC	Gabrielle David	All
USACE Albuquerque District	Daniel Delgado and Kraig Jashinsky	GP
	Daniel Delgado	WM
USACE Baltimore District	Matthew Gilbert, Matt Hynson, and Nick Ozburn	NE
USACE Charleston District	Jeremy Kinney	SE
USACE Chicago District	Andrew Blackburn	GP, NE
USACE Detroit District	Sabrina Miller	GP, NE
USACE Fort Worth District	Joseph Shelnutt and Elizabeth Shelton	GP, SE
USACE Galveston District	Mark Pattillo, Kara Vick, Wayne Fitzpatrick, and Shawn Patrick Hillen	GP, SE
USACE Huntington District	Robert (Wes) Barnett, Brian Bridgewater, and Justin Elkins	NE
USACE Kansas City District	Matt Mikulecky	GP
	Ryan Langer	NE
USACE Little Rock District	Pablo Bacon and David Rupe	NE
USACE Los Angeles District	Aaron Allen	AW, WM
USACE Louisville District	Patricia Grace-Jarrett, Rusty Retherford, and Sam Werner	NE
USACE Memphis District	Robert (Damon) McDermott	NE, SE
USACE Mobile District	Mike Moxey	SE
USACE Nashville District	Aric Payne, Eric Sinclair, and Mark G. McIntosh	NE, SE

Agency	**Member(s)**	**SDAM Region(s)**
USACE New England District	Taylor Bell and Erin Davis	NE
USACE New Orleans District	Jon Barmore	SE
USACE New York District	William (Tom) Bruno and Rosita Miranda	NE
USACE Norfolk District	Jeanne Richardson, Jennifer Serafin, Tucker Smith, Silvia Gazzera, and Laura Herrmann	NE
USACE North Atlantic Division	Andy Dangler	NE, SE
USACE Omaha District	Jeremy Grauf and Kevin Little	WM
	Karen Lawrence and Jeremy Grauf	GP
USACE Pittsburgh District	Alyssa Barkley, Alani Taylor, and Alex Veto	NE
USACE Rock Island District	Kirsten Brown	GP
USACE Sacramento District	Jaime Robb	AW, WM
USACE Savannah District	Justin Hammonds, Adam White, and Nathan Driggers	SE
USACE Seattle District	(Dale) Jess Jordan	WM
USACE St. Louis District	Chad LaMontagne and Kamren Metzger	GP, NE
USACE St. Paul District	April Marcangeli, Brian Yagle, and Faye Healy	GP
USACE Tulsa District	Rob Hoffman	GP
	Rob Hoffman and Michael Ware	NE, SE
USACE Vicksburg District	Bryton Hixson	SE
USACE Wilmington District	Erin Davis and Tyler Crumbley	NE, SE

8. References

Breiman, L. 2001. Random Forests. Machine Learning 45:5–32.

Buda, M., A. Maki, and M. A. Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks 106:249–259.

Cáceres, M. D., and P. Legendre. 2009. Associations between species and groups of sites: Indices and statistical inference. Ecology 90:3566–3574.

Cao, Y., and C. P. Hawkins. 2011. The comparability of bioassessments: A review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680–701.

Chapin, T. P., A. S. Todd, and M. P. Zeigler. 2014. Robust, low-cost data loggers for stream temperature, flow intermittency, and relative conductivity monitoring. Water Resources Research 50:6542–6548.

Chawla, N. V., N. Japkowicz, and A. Kotcz. 2004. Editorial: Special issue on learning from imbalanced data sets. Special Interest Group on Knowledge Discovery in Data Explorations Newsletter 6:1–6.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

David, G. C. L., K. M. Fritz, T.-L. Nadeau, B. J. Topping, A. O. Allen, P. H. Trier, S. L. Kichefski, L. A. James, E. Wohl, and D. Hamill. 2025. National Ordinary High Water Marks Field Delineation Manual for Rivers and Streams: Final Version, p. 366. Wetlands Regulatory Assistance Program (WRAP) ERDC/CCRREL TR-25-11, US Army Corps of Engineers - Engineer Research and Development Center, Vicksburg, MS. (Available from: https://erdc-library.erdc.dren.mil/entities/publication/76c61f8f-6d75-4a35-aaf3-39aa64918afb)

De Cáceres, M., F. Jansen, and N. Dell. 2010. indicspecies: Relationship Between Species and Groups of Sites. https://cran.r-project.org/web/packages/indicspecies/index.html

Dorney, J., and P. Russell. 2018. North Carolina Division of Water Quality methodology for identification of intermittent and perennial streams and their origins. Pages 273–279 *in* J. Dorney, R. Savage, R.W. Tiner, and P. Adamus (eds.), Wetland and Stream Rapid Assessments: Development, Validation, and Application. Elsevier, San Diego, CA.

Eddy, M., Gross, Shannon, K. M. Fritz, T.-L. Nadeau, B. J. Topping, R. Fertik Edgerton, R. D. Mazor, and J. E. Kelso. 2022. Development and Evaluation of the Beta Streamflow Duration Assessment Methods for the Great Plains (GP) (revised). Data supplement to EPA-840-B-22009, p. 45. EPA-840-R-23003, US Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/development-of-beta-sdam-gp-nov2023update.pdf)

Eng, K., D. M. Wolock, and M. D. Dettinger. 2016. Sensitivity of Intermittent Streams to Climate Variations in the USA. River Research and Applications 32:885–895.

Environmental Systems Research Institute (ESRI). 2024. Hydrology Toolset. Redlands, CA.

Freeman, M. C., and P. A. Marcinek. 2006. Fish assemblage responses to water withdrawals and water supply reservoirs in Piedmont streams. Environmental Management 38:435–450.

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.

Fritz, K. M., T.-L. Nadeau, J. E. Kelso, W. S. Beck, R. D. Mazor, R. A. Harrington, and B. J. Topping. 2020. Classifying streamflow duration: the scientific basis and an operational framework for method development. Water 12:2545.

Gallart, F., P. Llorens, J. Latron, N. Cid, M. Rieradevall, and N. Prat. 2016. Validating alternative methodologies to estimate the regime of temporary rivers when flow data are unavailable. Science of the Total Environment 565: 1001–1010.

Gasith, A., and V. H. Resh. 1999. Streams in Mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. Annual Review of Ecology and Systematics 30:51–81.

Goodrich, D. C., W. G. Kepner, L. R. Levick, and P. J. Wigington. 2018. Southwestern intermittent and ephemeral stream connectivity. Journal of the American Water Resources Association 54:400–422.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857–871.

Grantham, T. E., D. M. Carlisle, J. Howard, B. Lane, R. Lusardi, A. Obester, S. Sandoval-Solis, B. Stanford, E. D. Stein, K. T. Taniguchi-Quan, S. M. Yarnell, and J. K. H. Zimmerman. 2022. Modeling functional flows in California's rivers. Frontiers in Environmental Science 10:787473.

Gross, S., M. Eddy, K. M. Fritz, T.-L. Nadeau, B. J. Topping, R. Fertik Edgerton, R. D. Mazor, and K. Nicholas. 2023. Development and Evaluation of the Beta Streamflow Duration Assessment Methods for the Northeast and Southeast (revised). Data supplement to EPA-843-B-23001, p. 66. EPA-843-R-23003, US Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/development-of-beta-sdam-neandse-nov2023update.pdf)

Hammond, J. C., F. A. Saavedra, and S. K. Kampf. 2017. MODIS MOD10A2 derived snow persistence and no data index for the western U.S. (Available from: https://www.hydroshare.org/resource/1c62269aa802467688d25540caf2467e/)

Hammond, J. C., F. A. Saavedra, and S. K. Kampf. 2018. How does snow persistence relate to annual streamflow in mountain watersheds of the western U.S. with wet maritime and dry continental climates? Water Resources Research 54:2605–2623.

Hammond, J. C., M. Zimmer, M. Shanafield, K. Kaiser, S. E. Godsey, M. C. Mims, S. C. Zipper, R. M. Burrows, S. K. Kampf, W. Dodds, C. N. Jones, C. A. Krabbenhoft, K. S. Boersma, T. Datry, J. D. Olden, G. H. Allen, A. N. Price, K. Costigan, R. Hale, A. S. Ward, and D. C. Allen. 2021. Spatial patterns and drivers of nonperennial flow regimes in the contiguous United States. Geophysical Research Letters 48: e2020GL090794.

Hart, E., and K. Bell. 2015. PRISM: access data from The Oregon State PRISM Climate Project. (Available from: https://docs.ropensci.org/prism/)

Hawkins, C. P., Y. Cao, and B. Roper. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066–1085.

Hedman, E. R., and W. R. Osterkamp. 1982. Streamflow characteristics related to channel geometry of streams in western United States, p. 17. Water Supply Paper 2193, US Geological Survey, Washington, D.C. (Available from: https://pubs.er.usgs.gov/publication/wsp2193)

Hewlett, J. D. 1982. Principles of Forest Hydrology, p. 183. University of Georgia Press, Athens, GA.

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.

James, A., K. M. Fritz, B. J. Topping, R. Fertik Edgerton, K. Nicholas, R. D. Mazor, and T.-L. Nadeau. 2025. Streamflow Duration Assessment Method for the Northeast and Southeast of the United States. Version 2.0.1, p. 108. EPA-843-B-25001, U.S. Environmental Protection Agency, Washington, D.C. (Available from: https://www.epa.gov/system/files/documents/2025-08/ne-se_sdam_august_2025.pdf)

James, A., K. M. Fritz, B. J. Topping, T.-L. Nadeau, R. Fertik Edgerton, K. Nicholas, and R. D. Mazor. 2024. Streamflow Duration Assessment Method for the Great Plains of the United States. Version 2.0, p. 67. EPA-843-B-24001, U.S. Environmental Protection Agency, Washington, D.C. (Available from: https://www.epa.gov/system/files/documents/2024-10/gp-sdam_user_manual_october2024.pdf)

James, A., K. S. McCune, and R. D. Mazor. 2022a. Review of Flow Duration Methods and Indicators of Flow Duration in the Scientific Literature: Great Plains of the United States, p. 56. EPA-840-B-22006, US Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-05/Final%20GP%20Lit%20Review_March2022.pdf)

James, A., K. S. McCune, and R. D. Mazor. 2022b. Review of Flow Duration Methods and Indicators of Flow Duration in the Scientific Literature: Northeast and Southeast of the United States, p. 102. EPA-840-B-22007, US Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-05/Literature-Review-Beta-SDAM-NE-and-SE.pdf)

James, A., T.-L. Nadeau, K. M. Fritz, B. J. Topping, R. Fertik Edgerton, J. E. Kelso, and R. D. Mazor. 2023a. User Manual for a Beta Streamflow Duration Assessment Method for the Great Plains of the United States. Version 1.1, p. 91. EPA-840-B-22009, United States Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/beta-sdam-for-the-gp-user-manual_v1_1_0.pdf)

James, A., T.-L. Nadeau, K. M. Fritz, B. J. Topping, R. Fertik Edgerton, J. E. Kelso, R. D. Mazor, and K. Nicholas. 2023b. User Manual for a Beta Streamflow Duration Assessment Method for the Northeast and Southeast of the United States. Version 1.1, p. 91. EPA-843-B-23001, United States Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/user-manual-beta-sdam-ne-and-se-v1_1_nov2023.pdf)

James City County. 2009. Perennial Stream Protocol Guidance Manual, p. 18. James City County, Williamsburg, VA. (Available from: https://jamescitycountyva.gov/DocumentCenter/View/2158/JCC-Perennial-Stream-Protocol-Manual-PDF?bidId=)

Kanno, Y., and J. C. Vokoun. 2010. Evaluating effects of water withdrawals and impoundments on fish assemblages in southern New England streams, USA. Fisheries Management and Ecology 17:272–283.

Kelso, J. E., and K. M. Fritz. 2021. Standard Operating Procedure: Processing Data and Classifying Streamflow Duration Using Continuous Hydrologic Data. EPA Report J-WECD-ECB-SOP-4425-0, p. 25. US Environmental Protection Agency, Cincinnati, OH.

Kuhn, M. 2020. caret: Classification and Regression Training. (Available from: https://cran.r-project.org/web/packages/caret/index.html)

Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News 2:18–22.

Lichvar, R. W., D. L. Banks, W. N. Kirchner, and N. C. Melvin. 2016. The National Wetland Plant List: 2016 wetland ratings. Phytoneutron 30:1–17.

Mackun, P. 2019. Fast growth in the Desert Southwest continues. America Counts: Stories Behind the Numbers. US Census Bureau. (Available from: https://www.census.gov/library/stories/2019/02/fast-growth-in-desert-southwest-continues.html)

Mazor, R. D., K. M. Fritz, B. J. Topping, T.-L. Nadeau, and J. E. Kelso. 2022. Development and Evaluation of the Beta Streamflow Duration Assessment Method (SDAM) for the Western Mountains (WM). Data supplement, p. 38. EPA 840-B-21008, United States Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/development-wm-beta-sdam_nov2023update.pdf)

Mazor, R. D., A. James, K. M. Fritz, B. J. Topping, T.-L. Nadeau, R. Fertik Edgerton, and K. Nicholas. 2024. Streamflow Duration Assessment Methods for the Arid West and Western Mountains of the United States. Version 2.0, p. 114. EPA-843-B-24002, U.S. Environmental Protection Agency, Washington, D.C. (Available from: https://www.epa.gov/system/files/documents/2024-10/aw-wm-sdam_user_manual_october2024.pdf)

Mazor, R. D., and K. S. McCune. 2021. Review of Flow Duration Methods and Indicators of Flow Duration in the Scientific Literature: Western Mountains, p. 55. 1222, Southern California Coastal Water Research Project, Costa Mesa, CA. (Available from: https://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/1222_FlowDurationLitReview_WesternMountains.pdf)

Mazor, R. D., A. C. Rehn, P. R. Ode, M. Engeln, K. C. Schiff, E. D. Stein, D. J. Gillett, D. B. Herbst, and C. P. Hawkins. 2016. Bioassessment in complex environments: Designing an index for consistent meaning in different settings. Freshwater Science 35:249–271.

Mazor, R. D., B. J. Topping, T.-L. Nadeau, K. M. Fritz, J. E. Kelso, R. A. Harrington, W. S. Beck, K. McCune, H. Lowman, A. Aaron, R. Leidy, J. T. Robb, and G. C. L. David. 2021a. User Manual for a Beta Streamflow Duration Assessment Method for the Arid West of the United States. Version 1.0, p. 83. EPA 800-K-21001. U.S. Environmental Protection Agency, Washington, D.C. (Available from:

https://www.epa.gov/sites/production/files/2021-03/documents/user_manual_beta_sdam_aw.pdf)

Mazor, R. D., B. J. Topping, T.-L. Nadeau, K. M. Fritz, J. E. Kelso, R. A. Harrington, W. S. Beck, K. S. McCune, A. O. Allen, and R. A. Leidy. 2023. Correction Notice for Development of the Beta Streamflow Duration Assessment Method for the Arid West, p. 1. EPA-800-K-21001, U.S. Environmental Protection Agency. (Available from: https://www.epa.gov/system/files/documents/2023-11/aw-betav1_1-dataanalysiscorrections.pdf)

Mazor, R. D., B. J. Topping, T.-L. Nadeau, K. M. Fritz, J. E. Kelso, R. A. Harrington, W. S. Beck, K. S. McCune, A. O. Allen, R. Leidy, J. T. Robb, and G. C. L. David. 2021b. Implementing an operational framework to develop a streamflow duration assessment method: a case study from the Arid West United States. Water 13:3310.

Mazor, R. D., B. J. Topping, T.-L. Nadeau, K. M. Fritz, J. E. Kelso, R. A. Harrington, W. S. Beck, K. S. McCune, A. O. Allen, R. Leidy, J. T. Robb, G. C. L. David, and L. Tanner. 2021c. User Manual for a Beta Streamflow Duration Assessment Method for the Western Mountains of the United States. Version 1.0, p. 116. EPA-840-B-21008, U.S. Environmental Protection Agency, Washington, D.C. (Available from: https://ftp.sccwrp.org/pub/download/PROJECTS/SDAM_for_WM/User%20Manual%20Beta%20SDAM%20WM_FINAL.pdf)

McCune, K., and R. D. Mazor. 2019. Review of Flow Duration Methods and Indicators of Flow Duration in the Scientific Literature: Arid Southwest, p. 84. (Available from: https://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/1063_FlowMethodsReview.pdf)

Messager, M. L., B. Lehner, C. Cockburn, N. Lamouroux, H. Pella, T. Snelder, K. Tockner, T. Trautmann, C. Watt, and T. Datry. 2021. Global prevalence of non-perennial rivers and streams. Nature 594:391–397.

Mohammed, R., J. Rawashdeh, and M. Abdullah. 2020. Machine learning with oversampling and undersampling techniques: overview study and experimental results. Pages 243–248 *in* 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan.

Nadeau, T.-L. 2015. Streamflow Duration Assessment Method for the Pacific Northwest, p. 28. EPA 910-K-14-001, U.S. Environmental Protection Agency, Washington, D.C. (Available from: https://www.epa.gov/system/files/documents/2022-03/sdam-pnw_nov-2015-final.pdf)

Nadeau, T.-L., S. G. Leibowitz, P. J. Wigington, J. L. Ebersole, K. M. Fritz, R. A. Coulombe, R. L. Comeleo, and K. A. 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 M. C. Rains. 2007. Hydrological connectivity between headwater streams and downstream waters: how science can inform policy. Journal of the American Water Resources Association 43:118–133.

North Carolina Division of Water Quality (NCDWQ). 2010. Methodology for Identification of Intermittent and Perennial Streams and their Origins (Version 4.11). North Carolina Department of Environment and Natural Resources, p. 41. (Available from:

https://files.nc.gov/ncdeq/Water%20Quality/Surface%20Water%20Protection/401/Policies_Guides_Manuals/StreamID_v_4point11_Final_sept_01_2010.pdf)
New Mexico Environment Department (NMED). 2011. Hydrology Protocol for the Determination of Uses Supported by Ephemeral, Intermittent, and Perennial Waters, p. 35. Surface Water Quality Bureau, New Mexico Environment Department, Albuquerque, NM. (Available from: https://www.env.nm.gov/surface-water-quality/wp-content/uploads/sites/25/2019/11/WQMP-CPP-Appendix-C-Hydrology-Protocol-20201023-APPROVED.pdf)
Ohio Environmental Protection Agency (Ohio EPA). 2020. Field Evaluation Manual for Ohio's Primary Headwater Habitat Streams. Version 4.1, p. 130. Ohio EPA Division of Surface Water, Columbus, OH. (Available from: https://epa.ohio.gov/static/Portals/35/credibledata/PHWHManual_2020_Ver_4_1_May_2020_Final.pdf)
Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. M. Stevens, E. Szoecs, and H. Wagner. 2019. vegan: Community Ecology Package. (Available from: https://cran.r-project.org/web/packages/vegan/index.html)
Perkin, J. S., K. B. Gido, J. A. Falke, K. D. Fausch, H. Crockett, E. R. Johnson, and J. Sanderson. 2017. Groundwater declines are linked to changes in Great Plains stream fish assemblages. Proceedings of the National Academy of Sciences 114:7373–7378.
R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rugel, K., C. R. Jackson, J. J. Romeis, S. W. Golladay, D. W. Hicks, and J. F. Dowd. 2012. Effects of irrigation withdrawals on streamflows in a karst environment: Lower Flint River Basin, Georgia, USA. Hydrological Processes 26:523–534.
Schumacher, C., and K. M. Fritz. 2019. Standard Operating Procedure: Verifying/Calibrating, Deploying, Retrieving Stream Temperature, Intermittency, and Conductivity (STIC) Data Loggers, and Downloading and Converting Data, p. 12. EPA Report D-WQD-ECB-024-SOP-02. Environmental Protection Agency, Washington, D.C.
Sleeter, B. M., T. S. Wilson, C. E. Soulard, and J. Liu. 2011. Estimation of late twentieth century land-cover change in California. Environmental Monitoring and Assessment 173:251–266.
Sparrow, K.H., S.W. Brown, C.E. French, K. Gordon, J.L. Gutenson, C.O. Hamilton, and J.C. Deters. 2025. Antecedent Precipitation Tool (APT) Version 3.0: Technical and User Guide, p. 34. ERDC/TN WRAP-25-1. U.S. Army Engineer Research and Development Center, Vicksburg, MS. (Available from: https://erdc-library.erdc.dren.mil/items/af14290c-ed08-411b-ae5d-effa5b5b947d)
Strickler, G. S. 1959. Use of the densiometer to estimate density of forest canopy on permanent sample plots. USDA, Forest Service, Pacific Northwest Forest and Range Experiment Station Research Note No. 180, Pages 1–5. No. 180, U.S. Forest Service, Portland, OR. (Available from: https://www.fs.usda.gov/research/treesearch/25832)
Tarboton, D. G., R. L. Bras, and I. Rodriguez-Iturbe. 1991. On the extraction of channel networks from digital elevation data. Hydrological Processes 5:81–100.

Tennessee Department of Environment and Conservation (TDEC). 2020. Guidance for Making Hydrologic Determinations, version 1.5. Division of Water Pollution Control. (Available from: https://www.tn.gov/content/dam/tn/environment/water/policy-and-guidance/dwr-nr-g-03-hydrologic-determinations%E2%80%9304012020.pdf)

Thien, S. J. 1979. A flow diagram for teaching texture-by-feel analysis. Journal of Agronomic Education 8:54–55.

United States Environmental Protection Agency (USEPA). 2019. Arid West, Western Mountains, Great Plains Flow Duration Protocol version 2.1, p. 38.

United States Environmental Protection Agency (USEPA). 2020. Northeast and Southeast Flow Duration Protocol, p. 57.

U.S. Army Corps of Engineers. 2008. Regional Supplement to the Corps of Engineers Wetland Delineation Manual: Arid West Region (Version 2.0), p. 135. J. S. Wakeley, R. W. Lichvar, and C. V. Noble (eds.). ERDC/EL TR-08-28, U.S. Army Engineer Research and Development Center, Vicksburg, MS: U.S. Army Engineer Research and Development Center. (Available from: https://usace.contentdm.oclc.org/utils/getfile/collection/p266001coll1/id/7627)

U.S. Army Corps of Engineers. 2010. Regional Supplement to the Corps of Engineers Wetland Delineation Manual: Western Mountains, Valleys, and Coast Region (Version 2.0), p. 153. J. S. Wakeley, R. W. Lichvar, and C. V. Noble (eds.). ERDC/EL TR-10-3, U.S. Army Engineer Research and Development Center, Vicksburg, MS: U.S. Army Engineer Research and Development Center. (Available from: https://usace.contentdm.oclc.org/utils/getfile/collection/p266001coll1/id/7646)

Vengosh, A., R. B. Jackson, N. Warner, T. H. Darrah, and A. Kondash. 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. Environmental Science & Technology 48:8334–8348.

Wohl, E., M. K. Mersel, A. O. Allen, K. M. Fritz, S. L. Kichefski, R. W. Lichvar, T.-L. Nadeau, B. J. Topping, P. H. Tier, and F. B. Vanderbilt. 2016. Synthesizing the Scientific Foundation for Ordinary High Water Mark Delineation in Fluvial Systems, p. 198. Wetlands Regulatory Assistance Program ERDC/CCREL SR-16-5, U.S. Army Corps of Engineers Engineer Research and Development Center. (Available from: https://erdc-library.erdc.dren.mil/entities/publication/81b728f7-9e4f-4ef8-e053-411ac80adeb3)

Yap, B. W., K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah. 2014. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. Pages 13–22 *in* T. Herawan, M. M. Deris, and J. Abawajy (eds.). Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013). Springer, Singapore.

Zimmerman, N., and J. R. Olson. 2019. Flow Duration Protocol for the Development of Regional Streamflow Duration Assessment Methods (SDAMs). U.S. Environmental Protection Agency. (Available from: https://ftp.sccwrp.org/pub/download/PROJECTS/Attachment 1_Flow Duration Protocol_version2.zip)

Appendix A. Glossary of terms

Term	Definition
Accuracy	The ability of an SDAM to correctly predict the streamflow duration of a reach.
Active or relict floodplain	Ordinal score for the extent of evidence of an active floodplain or recent incision with a relict floodplain. Higher scores indicate greater evidence for an active floodplain or recent incision with a relict floodplain.
Algae	Ordinal estimates for the extent of streambed covered by live macroalgae and visible biofilm, separately estimated for live and dead algae (excluding mats that appeared to be deposited from upstream) into: Not detected, <2%, 2-10%, 10-40%, and >40%.
Alluvial deposits	Ordinal score for the extent of fresh alluvium (from recent high flows) deposited within the stream or on the floodplain. Higher scores indicate greater extent of fresh alluvium deposited within or along the reach.
Amphibians	Abundance of amphibians within reach channel. See Appendix B for description of additional amphibian metrics included in Northeast and Southeast regions.
Antecedent Precipitation Condition (APC)	The wetness status at a particular location and date based on the previous 30-d periods relative to the previous 30-y precipitation climatology using US Army Corps of Engineers' Antecedent Precipitation Tool (APT) Version 3.0 (Sparrow et al. 2025). Estimated precipitation amounts are scored for each rolling 30-d period to represent Normal (between 30th-70th%), Dry (<30th%), and Wet (>70th%) and given more weight based on the imminency of the rolling 30-d period relative to the reach visit. The APC is the combined score (range 6 – 18) across the three rolling 30-d periods and is assigned an APC class as either Drier than Normal (APC score < 10), Normal (10 ≤ APC score < 15), or Wetter than normal (≥ 15 APC score) classes.
Aquatic macroinvertebrates	Invertebrate organisms that require aquatic environments for parts or all of their life cycle and are visible without the use of a microscope (i.e., > 0.5 mm body length). Includes bottom dwelling or benthic macroinvertebrates. See Appendix B for descriptions of different aquatic macroinvertebrate metrics.
Assessment reach	The length of reach, ranging from 40 m to 200 m, where SDAM indicators are measured.
At least intermittent (ALI)	A Regional SDAM classification outcome that identifies a stream reach that may be intermittent or perennial, but not ephemeral.
Augmentation	The process of replicating observations from sites with relatively few visits so that they have the same number of observations as sites with more visits. Augmentation has the same meaning as oversampling.
Bank	The side of an active channel, typically associated with a steeper side gradient than the adjacent streambed, floodplain, or valley bottom.
Bankfull elevation	The elevation associated with a shift in the hydraulic geometry of the channel and the transition point between the channel and the floodplain. In unconstrained settings this is the height of the water in the channel just when it begins to flow onto the floodplain.
Bankfull width	Width of the stream channel at bankfull elevation.
Binning	The transformation of a continuous metric (e.g., bank width) into ordinal bins (e.g., <1 m, between 1 and 5 m, ≥ 5m).

Term	Definition
Capping	The transformation of a continuous metric (e.g., number of hydrophytic plant species) into a semi-continuous metric that has a maximum value (e.g., 1 taxon, 2 taxa, 3 or more taxa).
Catchment	An area of land bounded by a drainage divide, which drains to a channel or waterbody. Synonymous with watershed.
Channel	A feature in fluvial systems consisting of a streambed and its opposing banks which confines and conveys surface water flow. A braided system consists of multiple channels, which may include inactive or abandoned channels.
Coleoptera	Order of insects commonly called beetles.
Continuity of channel bed and bank	Ordinal score for the evidence of bank continuity and streambed development. Higher scores indicate greater degree of channel development and continuity of bed and banks.
Depositional bars and benches	Ordinal score for the extent of alluvial bars and/or benches present in the channel. Higher scores indicate greater prevalence of the features.
Difference in vegetation (riparian corridor)	Ordinal score for the distinctiveness of vegetation in the riparian corridor compared to surrounding upland vegetation. Higher scores indicate a more distinct riparian corridor.
Diptera	Order of insects commonly called the true flies.
Downsampling	The process of withholding observations from sites with many visits so that they have the same number of observations as sites with fewer visits. Contrast with oversampling or augmentation. See also augmentation, oversampling, rebalancing.
Drainage area	The total surface area (miles^2) that contributes runoff to the reach coordinates from USGS StreamStats (DRNAREA) or National Mapper (area measurement tool).
Elevation	Mean watershed elevation (m) draining to reach coordinates from StreamCat.
Entrenchment ratio	The ratio of floodprone width to bankfull width, used to characterize floodplain channel dimensions for estimating extent of channel entrenchment and connectivity to the floodplain.
Ephemeral	Channels that flow only in direct response to precipitation. Water typically flows at the surface only during and/or shortly after large precipitation events, the streambed is always above the water table, and stormwater runoff is the primary water source.
Ephemeroptera	Order of aquatic insects commonly called mayflies.
FAC	Facultative plants. They are equally likely to occur in wetlands and non-wetlands.
FACU	Facultative upland plants. They usually occur in non-wetlands but are occasionally found in wetlands.
FACW	Facultative wetland plants. They usually occur in wetlands but may occur in non-wetlands.
Fibrous plant roots	Ordinal score for the abundance of fibrous roots (non-woody, <2.5 mm diameter) in the streambed of the reach. Lower scores indicate greater abundance of fibrous roots in the streambed.
Fish	Ordinal score for the abundance of fish observed (excluding any non-native *Gambusia* within the reach. Higher scores indicate greater abundance of fish observed.

Term	Definition
Floodprone width	The width around a stream channel that is prone to flooding, measured at twice the bankfull channel height.
Flow Class	An outcome of a streamflow duration assessment (Ephemeral, Intermittent, Perennial, At Least Intermittent, Less than Perennial)
Gastropoda	Class of invertebrates that are commonly called snails.
Grade control	Ordinal score for the extent and kinds of grade control features in the channel. Higher scores indicate larger size, more prevalence, and/or permanence of structures acting as grade controls in the channel.
Groundwater	Water found underground in soil, pores, or crevices in rocks.
Headcuts	Ordinal score for the size and number of headcuts in the channel. Higher scores indicate presence and greater vertical drop in bed associated with headcut(s) in the channel.
Hemiptera	Order of insects commonly called true bugs.
Hydric soils	Presence/absence of soils showing redoximorphic features indicative of saturated conditions at the toe of stream banks in the reach.
Hydrophyte	Plants that are adapted to inundated conditions found in wetlands and riparian areas. As used for SDAMs, plants identified as having FACW or OBL wetland indicator status in regional National Wetland Plant Lists. The number of hydrophyte species growing within the reach channel and ½ channel width on each side were recorded.
Hyporheic	The saturated zone under a river or stream, including the substrate and water-filled spaces between the particles.
In-channel structure / riffle-pool sequence	Ordinal score of diversity and distinctiveness of riffles, pools, and other flow-based microhabitats. Higher scores indicate more distinct riffles, pools, and other flow habitats with clear transitions within the reach.
Indicator	Measurable properties of an ecosystem that provide inference for a more complex characteristic (e.g., streamflow duration) that is either too difficult or resource intensive to measure directly. For the Regional SDAMs, indicators are rapid, generally field-based measurements that predict streamflow duration class. For SDAM development, the term “indicator” is used in two ways: First, it refers to the general characteristic of the stream that is measured during assessment (e.g., aquatic invertebrates). In addition, it refers to the metrics that are used in the final SDAM (e.g., total abundance of aquatic invertebrates). See metric.
Indicator species	A taxon determined through indicator species analysis (Cáceres and Legendre 2009) to be associated with habitats sharing a common environmental characteristic (such as perennial reaches).
Intermittent	Channels that contain sustained flowing surface water for only part of the year, typically during the wet season, where the streambed may be below the water table and/or where the snowmelt from surrounding uplands provides sustained flow. The flow may vary greatly with stormwater runoff.
Iron-oxidizing bacteria or fungi	Presence/absence of visual evidence of iron-oxidizing bacteria or fungi (as rusty floc or oily sheen that fractures) in the reach channel.

Term	Definition
Isolated pools	The number of pools in the channel of the reach without any connection to flowing surface water.
Latitude	Coordinate measuring the north-south position in decimal degrees for the downstream end of the reach (WGS84).
Leaf litter	Ordinal score of the extent of leaf litter distributed on the active channel of the reach. Lower scores indicate greater percent of the active channel covered with leaves (less redistribution by existing or recent flowing water).
Less than perennial (LTP)	A Regional SDAM classification outcome that identifies a stream reach that may be intermittent or ephemeral, but not perennial.
Liverworts	Ordinal score of liverwort percent cover of the active channel in the reach: not detected, ≤2%, 2-10%, and >10%.
Longitude	Coordinate measuring the east-west position in decimal degrees for the downstream end of the reach (WGS84).
Maximum pool depth	Maximum water depth (nearest cm) within the channel of the reach.
Metric	A specific quantification an indicator of streamflow duration of a reach (e.g., number of hydrophytic plant species within the channel, presence/absence of hydrophytic plant species within the channel). See indicator.
Metric screening	The process of eliminating candidate metrics for an SDAM that do not meet specified criteria (e.g., eliminating a metric that has the same value in >99% of samples).
Model refinement	The manual and iterative process of modifying a model to improve ease of use by diverse practitioners. For the SDAMs, refinement included the substitution, elimination, or modification of metrics, as well as the evaluation of single indicators.
Mosses	Ordinal score of streamer moss (prostrate, long-formed, branched pleurocarps) percent cover of the active channel in the reach: not detected, ≤2%, 2-10%, and >10%.
Natural valley	Ordinal score for the extent of valley development (proportion of catchment area sloping to the valley bottom). Higher scores indicate greater proportion of the catchment area slopes to the valley bottom or channel.
Need more information (NMI)	A Regional SDAM classification outcome that cannot confidently identify a stream reach as ephemeral, intermittent, or perennial.
NI	Plants that have no assigned wetland indicator (e.g., FACW, FACU) in a specific National Wetland Plant List region.
Noninsects	Aquatic macroinvertebrates that are not insects, includes aquatic crustaceans, snails, clams, worms, mites, and other groups.
Nonmetric multidimensional scaling (NMS)	A multivariate ordination method of visualizing complex data (e.g., species abundances) in a small number of dimensions that preserve ecological similarities among samples.
OBL	Obligate wetland plants. They almost always occur in wetlands.
Odonata	Order of aquatic insects commonly called damselflies and dragonflies.
Oligochaeta	Subclass of invertebrates that are commonly called segmented worms.

Term	Definition
Ordinary high-water mark (OHWM)	The line on the shore established by the fluctuations of water and indicated by physical characteristics, such as a clear natural line impressed on the bank, shelving, changes in the character of the soil, destruction of terrestrial vegetation, the presence of litter and debris, or other appropriate means that consider the characteristics of the surrounding areas. See 33 CFR 328.3. An OHWM is required to establish lateral extent of U.S. Army Corps of Engineers jurisdiction in non-tidal streams. See 33 CFR 328.4.
Oversampling	The process of replicating observations from sites with relatively few visits so that they have the same number of observations as sites with more visits. Oversampling has the same meaning as augmentation. See also downsampling, rebalancing.
Particle size / stream substrate sorting	Ordinal score for the extent of evidence of substrate sorting within the channel and/or similarity of particle size in the channel relative to outside of channel. Higher scores indicate greater sorting of substrate within the channel relative to surrounding uplands.
Perennial	Channels that contain flowing surface water continuously during a year of normal rainfall, often with the streambed located below the water table for most of the year. Groundwater typically supplies the baseflow for perennial reaches, but the baseflow may also be supplemented by stormwater runoff and/or snowmelt.
Plecoptera	Order of aquatic insects commonly called stoneflies.
Pool	A depression in a channel where water velocity is slow and suspended particles tend to deposit. Pools typically retain surface water longer than other portions of intermittent or ephemeral reaches.
Precipitation	30-y normal mean annual or monthly precipitation (PRISM climate data; Hart and Bell 2015).
Precision	The ability of an SDAM to produce consistent classifications of a reach over multiple visits.
Random forest	A machine learning algorithm that combines numerous decision trees to predict a variable or membership within a class.
Reach	A length of stream that generally has consistent geomorphological and biological characteristics.
Reach visit	Set of indicator values measured on a single date from a single study reach. Sample unit for the development of SDAMs.
Rebalancing	The process of increasing the number of observations at sites with few visits (through oversampling or augmentation) and/or decreasing the number of observations at sites with many visits (through downsampling) to produce a dataset where all sites have the same number of observations.
Recursive feature elimination (RFE)	A predictor selection method used in random forest to simplify complex models that use many predictors. Complex models are simplified by iteratively removing the least important predictors and then retraining the model until a simpler but sufficiently accurate model is produced.
Responsiveness	The ability for a metric to reflect differences among reaches of different streamflow durations.

Term	Definition
Riffle	A shallow portion of a channel where water velocity and turbulence are high, typically with coarse substrate (cobble and gravels). Riffles typically dry out earlier than other portions of intermittent or ephemeral reaches, and harbor higher abundance and diversity of aquatic macroinvertebrates.
Riparian	A transitional area between the channel and adjacent upland ecosystems.
Rooted upland plants	Ordinal score for the percent cover of plants rooted in the streambed that have wetland indicator statuses of FAC, FACU, UPL, and NI. Lower scores indicate greater percent of the streambed covered with rooted upland plants.
Runoff	Surface flow of water caused by precipitation or irrigation over saturated or impervious surfaces.
Sediment deposition on plants and debris	Ordinal score for the extent of evidence of sediment deposition on plants and on debris within the floodplain. Higher scores indicate the sediment deposition was more prevalent throughout the reach.
Sensitivity	The readiness with which an SDAM's classification changes in response to different sampling conditions or input data.
Seeps and springs	Presence/absence of seeps and/or springs observed within ½ channel width of the reach channel.
Single indicator	An indicator whose observation by itself can result in a determination that a reach is at least intermittent.
Sinuosity	Ordinal score for channel sinuosity. Higher scores indicate more sinuous channels.
Slope	Reach slope measured with a handheld clinometer or other device.
Snakes	Presence/absence of snakes with water-dependent life stages in the reach channel (not assessed in Northeast and Southeast).
Snow persistence	The percent of time snow is visible on the ground via remote sensing between the dates of January 1 and July 3 (Hammond et al. 2017).
Soil moisture	Measured in reaches without surface water, the degree of sediment saturation (at 15 cm depth into streambed) is qualitatively measured as "saturated", "partly saturated" or "dry" in three locations of the reach. Reaches with surface water were classified as "not assessed," and dry reaches without sediment (i.e., bedrock, concrete) were classified as "dry."
Soil texture	Measured texture classes done concurrently with soil moisture following Thien (1979).
Stratum (plural: Strata)	A subset of a population that is being sampled or analyzed. Strata may be used to ensure that sampling efforts cover major environmental gradients (i.e., sampling strata), or to assess the performance of a model in major subregions (e.g., assessment strata).
Streambed	The bottom of a stream channel between the banks over which water and sediment are transported during periods of flow.
Streambed substrate characteristics	Classification of streambed substrate types as dominant (top one or two in percent cover), significant (at least 1% cover, max of 8 types), or not significant with in the study reach.
Streamflow Duration Assessment Method (SDAM)	A method that uses a small number of indicators to predict the streamflow duration class of stream reaches.

Term	Definition
Stream order	Strahler stream orders 1, 2 or 3+ based on synthetic network from StreamStats or blue lines on USGS 1:24,000 quadrangle. Assigned stream order 1 if reach was upgradient of synthetic network or blue line network.
Stream shading	Percent shade from objects (vegetation, buildings, walls) blocking sunlight from reaching the streambed of the reach. Measured following Strickler (1959) in three locations at 0.3 m above the water surface facing four directions (upstream, downstream, right bank, and left bank).
Subsurface flow	In a reach with discontinuous surface flow, this is the percentage of the reach length determined (visually estimated or measured with a tape) that water is observed or heard moving below the streambed surface (see hyporheic).
Surface flow	The percentage of the reach length determined (visually estimated or measured with a tape) that water is observed to be moving over the streambed.
Temperature	30-y normal mean, maximum, and minimum annual air temperature (PRISM climate data; Hart and Bell 2015).
Testing	The process of evaluating the performance of a model with data that was not used in model training (known as the testing dataset).
Thalweg	The line along the deepest flowpath within the channel.
Training	The process of teaching a model or machine learning algorithm (such as random forest) how to make predictions. Training is conducted with a portion of the development dataset known as the training data.
Trichoptera	Order of aquatic insects commonly called caddisflies.
UPL	Upland plants. They almost always occur in non-wetlands.
Uplands	Any portion of a drainage basin outside the river corridor.
Valley width	The portion of the valley within which the fluvial channel is able to migrate without cutting into hill slopes, terraces, or artificial structures.
Water in channel (Presence of baseflow)	Ordinal score for the surface water observed in the reach. Higher scores indicate reaches with higher levels of surface water flow or presence within the channel.
Watershed	An area of land, bounded by a drainage divide, which drains to a channel or waterbody. Synonymous with catchment.
WebWIMP season (WW)	Web-based, Water-Budget, Interactive, Modelling Program classifies dates for locations as occuring in the dry season or wet season based on climatic balance estimates (0.5º gridded raster). Dates for locations are identified as being in the dry season when the estimated historical (1950-1999) monthly average potential evapotranspiration (PET) is greater than the historical monthly average precipitation plus snowmelt. Months for locations are identified as being in the wet season when the historical monthly average precipitation plus snowmelt is greater than the historical monthly average PET. Because WW estimates climatic water balance for land, season classifications cannot be made for locations within 0.5º raster cells with mostly water (i.e., near coast). WW is a web application developed by the University of Delaware and was accessed using US Army Corps of Engineers' Antecedent Precipitation Tool (APT) Version 3.0 (Sparrow et al. 2025).

Term	Definition
Woody jams	Number of woody jams present within the channel reach or up to 10 m outside of reach. Woody jams are accumulations of at least 3 large pieces (>1 m long and > 10 cm diameter) wood that completely span that active channel and is in contact with the streambed.

Appendix B: Indicators and metrics evaluated for SDAM development

Candidate metrics calculated from indicators shown in Table 4. Bio: biological indicators. Geom: Geomorphological indicators. Hydro: Hydrological indicators. Geos: Geospatial indicators. "ISA" metrics were based on indicator species analysis conducted on the development data set (described in Section 2.6.1.1 Indicator species analysis). Coding O = ordinal (number of levels, half scores counted as a level), I = integer, C = continuous. NM from NMED 2011, NC from NCDWQ 2011, OH from Ohio EPA 2020.

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Bio	Algae cover	AlgalCover_LiveOrDead_NoUpstream	Live or dead algal cover not from upstream	O(5)	W, GP, E
Bio	Algae cover	AlgalCover_Live_NoUpstream	Live algal cover not from upstream	O(5)	W, GP, E
Bio	Iron-oxidizing bacateria and fungi	ironox_bfscore_NC	Iron oxidizing bacteria/fungi abundance[NC]	O(7)	E
Bio	Iron-oxidizing bacateria and fungi	ironox_bfscore_NM	Iron oxidizing bacteria/fungi present[NM]	O(2)	W, GP
Bio	Aquatic invertebrates	BMI_presence	Aquatic macroinvertebrates presence	O(2)	W, GP, E
Bio	Aquatic invertebrates	BMI_score_alt4	Aquatic macroinvertebrate score[NC]	O(4)	E
Bio	Aquatic invertebrates	EPT_abundance	Abundance of Ephemeroptera, Plecoptera, and Trichoptera	I	W, GP, E
Bio	Aquatic invertebrates	EPT_taxa	Ephemeroptera, Plecoptera, and Trichoptera family richness	I	W, GP, E
Bio	Aquatic invertebrates	Ephemeroptera_abundance	Abundance of Ephemeroptera	I	W, GP, E
Bio	Aquatic invertebrates	Ephemeroptera_taxa	Ephemeroptera family richness	I	W, GP, E
Bio	Aquatic invertebrates	GOLD_abundance	Abundance of Gastropoda, Oligochaeta, and Diptera	I	W, GP, E
Bio	Aquatic invertebrates	Noninsect_abundance	Abundance of non-insects	I	W, GP, E
Bio	Aquatic invertebrates	OCH_abundance	Abundance of Odonata, Coleoptera, and Heteroptera	I	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Bio	Aquatic invertebrates	OCH_taxa	Odonata, Coleoptera, and Heteroptera family richness	I	W, GP, E
Bio	Aquatic invertebrates	Plecoptera_abundance	Abundance of Plecoptera	I	W, GP, E
Bio	Aquatic invertebrates	Plecoptera_taxa	Plecoptera family richness	I	W, GP, E
Bio	Aquatic invertebrates	Richness	Aquatic invertebrate taxa richness	I	W, GP, E
Bio	Aquatic invertebrates	TotalAbundance	Abundance of aquatic invertebrates	I	W, GP, E
Bio	Aquatic invertebrates	Trichoptera_abundance	Abundance of Trichoptera	I	W, GP, E
Bio	Aquatic invertebrates	Trichoptera_taxa	Trichoptera family richness	I	W, GP, E
Bio	Aquatic invertebrates	ephemeral_ISA_abundance	Abundance of taxa listed for ephemeral in Table 8	I	GP
Bio	Aquatic invertebrates	ephemeral_ISA_taxa	Richness of taxa listed for ephemeral in Table 10	I	GP
Bio	Aquatic invertebrates	ephemeral_ISAsubregion_abundance	Abundance of taxa listed for ephemeral in Table 8	I	GP
Bio	Aquatic invertebrates	ephemeral_ISAsubregion_taxa	Richness of taxa listed for ephemeral in Table 8	I	GP
Bio	Aquatic invertebrates	ephinteph_ISA_abundance	Abundance of taxa listed for ephemeral and ephemeral + intermittent in Table 8	I	GP, E
Bio	Aquatic invertebrates	ephinteph_ISA_taxa	Richness of taxa listed for ephemeral and ephemeral + intermittent in Table 8	I	GP, E
Bio	Aquatic invertebrates	ephinteph_ISAsubregion_abundance	Abundance of taxa listed for ephemeral and ephemeral + intermittent in Table 8	I	GP
Bio	Aquatic invertebrates	ephinteph_ISAsubregion_taxa	Richness of taxa listed for ephemeral and ephemeral + intermittent in Table 8	I	GP
Bio	Aquatic invertebrates	perennial_ISA_abundance	Abundance of taxa listed for perennial in Table 8	I	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Bio	Aquatic invertebrates	perennial_ISA_taxa	Richness of taxa listed for perennial in Table 8	I	W, GP, E
Bio	Aquatic invertebrates	perennial_ISAsubregion_abundance	Abundance of taxa listed for perennial in Table 8	I	W, GP
Bio	Aquatic invertebrates	perennial_ISAsubregion_taxa	Richness of taxa listed for perennial in Table 8	I	W, GP
Bio	Aquatic invertebrates	perennial_NC_abundance	Abundance of NC perennial indicator aquatic invertebrate taxa in Table 9	I	GP, E
Bio	Aquatic invertebrates	perennial_NC_live_abundance	Abundance of living NC perennial indicator aquatic invertebrate taxa in Table 9	I	GP, E
Bio	Aquatic invertebrates	perennial_NC_taxa	Richness of NC perennial indicator aquatic invertebrate taxa in Table 9	I	GP, E
Bio	Aquatic invertebrates	perennial_PNW_abundance	Abundance of PNW perennial indicator aquatic invertebrate taxa in Table 9	I	W
Bio	Aquatic invertebrates	perennial_PNW_live_abundance	Abundance of living PNW perennial indicator aquatic invertebrate taxa in Table 9	I	W
Bio	Aquatic invertebrates	perennial_PNW_taxa	Richness of PNW perennial indicator aquatic invertebrate taxa in Table 9	I	W
Bio	Aquatic invertebrates	perintper_ISA_abundance	Abundance of taxa listed for perennial and intermittent + perennial in Table 8	I	W, E
Bio	Aquatic invertebrates	perintper_ISA_taxa	Richness of taxa listed for perennial and intermittent + perennial in Table 8	I	W, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Bio	Aquatic invertebrates	perintper_ISAsubregion_abundance	Abundance of taxa listed for perennial and intermittent + perennial in Table 8	I	W, GP
Bio	Aquatic invertebrates	perintper_ISAsubregion_taxa	Richness of taxa listed for perennial and intermittent + perennial in Table 8	I	W, GP
Bio	Differences in vegetation	DifferencesInVegetation_score	Differences in vegetation score[NM]	O(7)	W, GP, E
Bio	Fibrous rooted plants	FibrousRootedPlants_score	Fibrous roots in streambed score[NC]	O(7)	E
Bio	Mosses and liverworts	Liverwort_cover	Liverwort cover on the streambed	O(4)	W, GP, E
Bio	Mosses and liverworts	Moss_cover	Streamer moss cover on the streambed	O(4)	W, GP, E
Bio	Shading	PctShading	Percent shading of streambed (densiometer)	C	W, GP, E
Bio	Upland rooted plants	UplandRootedPlants_score	Absence of rooted upland plants in streambed[NM, NC]	O(7,4)[3]	W, GP, E
Bio	Hydrophytes	hydrophytes_obl_present	Number of OBL hydrophytic plant species observed within the reach channel and 1/2 channel width on either bank	I	E
Bio	Hydrophytes	hydrophytes_obl_present_inchannel	Number of OBL hydrophytic plant species observed only within the reach channel	I	E
Bio	Hydrophytes	hydrophytes_present	Number of hydrophytic plant species (FACW, OBL or SAV) observed within the reach channel and 1/2 channel width on either bank	I	W, GP, E

[3] NM scoring used in the AW, WM and GP allows half scores and has 7 possible values, NC scoring used in the NE and SE does not allow half scores and has 4 possible values.

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Bio	Hydrophytes	hydrophytes_present_inchannel	Number of hydrophytic plant species (FACW, OBL or SAV) observed only within the reach channel	I	E
Bio	Vertebrates	AmphSnake_PA	Presence of aquatic life stage of amphibians or snakes	O (2)	W, GP, E
Bio	Amphibians	Amphib_Juvenile_abundance	Abundance of aquatic juvenile amphibians	I	E
Bio	Amphibians	Amphib_Juvenile_richness	Richness of aquatic juvenile amphibian taxa	I	E
Bio	Amphibians	Amphib_abundance	Abundance of amphibians	I	E
Bio	Amphibians	Amphib_richness	Richness of amphibians	I	E
Bio	Amphibians	Amphibian_presence	Presence aquatic life stage of amphibians	O(2)	W, GP, E
Bio	Fish	Fish_PA	Presence of fish	O(2)	W, GP, E
Bio	Amphibians	Frog_abundance	Abundance of frogs	I	E
Bio	Amphibians	Frog_richness	Richness of frogs	I	E
Bio	Amphibians	Multiyear_tadpole_abundance	Abundance of tadpoles requiring >1 year for lifestage	I	E
Bio	Amphibians	Salamander_Juvenile_abundance	Abundance of salamanders	I	E
Bio	Amphibians	Salamander_abundance	Richness of salamanders	I	E
Bio	Amphibians	Salamander_richness	Abundance of juvenile salamanders	I	E
Bio	Reptiles	snake_score	Presence aquatic lifestage of snakes	O(2)	W, GP
Geom	Active or relict floodplain	ActiveFloodplain_score	Active floodplain indicator score[NC]	O(7)	E
Geom	Alluvial deposits	Alluvial deposits	Recent alluvial deposits indicator score[NC]	O(7)	E
Geom	Bankfull channel width	BankWidthMean	Mean bankfull width (from ≤3 measurements, nearest 0.1m)	C	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Geom	Floodplain channel dimensions	ChannelDimensions_score_NM	Floodplain and channel dimensions score[NM]	O(3)	W, GP
Geom	Bed and bank continuity	Continuity_score	Continuity of channel bed and bank score[NC]	O(7)	E
Geom	Depositional bars and benches	Depositional_score	Depositional bars or benches score[NC]	O(7)	E
Geom	Grade control	GradeControl_score	Grade control score	O(7)	E
Geom	Headcuts	Headcut_score	Headcut score	O(7)	E
Geom	Natural valley	NaturalValley_score	Natural valley score	O(6)	E
Geom	In-channel structure/riffle-pool sequence	RifflePoolSeq_score	In-channel structure and riffle-pool sequence score[NM, NC]	O(7)	W, GP, E
Geom	Sediment deposition on plants and debris	SedimentOnPlantsDebris_score	Sediment on plants and debris score[NM, NC]	O(7)	W, GP, E
Geom	Sinuosity	Sinuosity_score	Channel sinuosity score[NM, NC]	O(4)	W, GP, E
Geom	Slope	Slope	Chanel slope (clinometer or viewfinder, nearest 0.5%)	C	W, GP, E
Geom	Particle size/stream substrate sorting	SubstrateSorting_score	Particle size or stream substrate sorting score[NM, NC]	O(5,7)[4]	W, GP, E
Geom	Floodplain and channel dimensions	fp_entrchementratio_mean	Mean entrenchment ratio (max 2.5), (floodprone:bankfull)[OH]	C	E
Geom	Streambed substrate characteristics	oh_substrate_dominance_score	Weighted substrate size score[OH]	I	E
Geom	Streambed substrate characteristics	oh_substrate_significance_score	Number of biologically significant substrates (max 8)[OH]	I	E
Hydro	Hydric soils	HydricSoils_score	Presence of hydric soils in channel[NM, NC]	O(2)	W, GP, E

[4] NM scoring used in the AW, WM and GP allows half scores and has 5 possible values, NC scoring used in the NE and SE allows half scores and has 7 possible values.

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Hydro	Leaf litter	LeafLitter_score	Leaf litter score[NC]	O(5)	E
Hydro	Organic drift lines	ODL_score	Organic drift lines score[NC]	O(7)	E
Hydro	Woody jams	WoodyJams_number	Number of woody jams within or ≤10m from reach	I	W, GP, E
Geos	Climate—Snow persistence	MeanSnowPersistence_01	Mean percent of year snow persistence within 1-km radius	C	W, GP, E
Geos	Climate—Snow persistence	MeanSnowPersistence_05	Mean percent of year snow persistence within 5-km radius	C	W, GP, E
Geos	Climate—Snow persistence	MeanSnowPersistence_10	Mean percent of year snow persistence within 10-km radius	C	W, GP, E
Geos	Climate—Precipitation	ppt	Mean annual precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.11121	Mean November, December & January precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.234	Mean February, March & April precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.567	Mean May, June & July precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.8910	Mean August, September & October precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m01	Mean January precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m02	Mean February precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m03	Mean March precipitation (mm; PRISM 30-y normal)	C	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Geos	Climate—Precipitation	ppt.m04	Mean April precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m05	Mean May precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m06	Mean June precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m07	Mean July precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m08	Mean August precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m09	Mean September precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m10	Mean October precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m11	Mean November precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Precipitation	ppt.m12	Mean December precipitation (mm; PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.11121	Mean November, December & January air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.234	Mean February, March & April air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.567	Mean May, June & July air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.8910	Mean August, September & October (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m01	Mean January air temperature (°C, PRISM 30-y normal)	C	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Geos	Climate—Temperature	temp.m02	Mean February air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m03	Mean March air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m04	Mean April air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m05	Mean May air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m06	Mean June air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m07	Mean July air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m08	Mean August air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m09	Mean September air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m10	Mean October air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m11	Mean November air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	temp.m12	Mean December air temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	tmax	Maximum annual temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	tmean	Mean annual temperature (°C, PRISM 30-y normal)	C	W, GP, E
Geos	Climate—Temperature	tmin	Minimum annual temperature (°C, PRISM 30-y normal)	C	W, GP, E

Type	Indicator	Metric	Definition	Coding	Aggregate regions
Geos	Watershed and location	DRNAREA_mi2	Drainage area (nearest 0.1 mi^2)[OH]	C	W, GP, E
Geos	Watershed and location	Elev_m	Mean watershed elevation (m, StreamCat)	C	W, GP, E
Geos	Watershed and location	StreamOrder	Strahler stream order from synthetic network (max 3, StreamStats)[NC]	O(3)	E