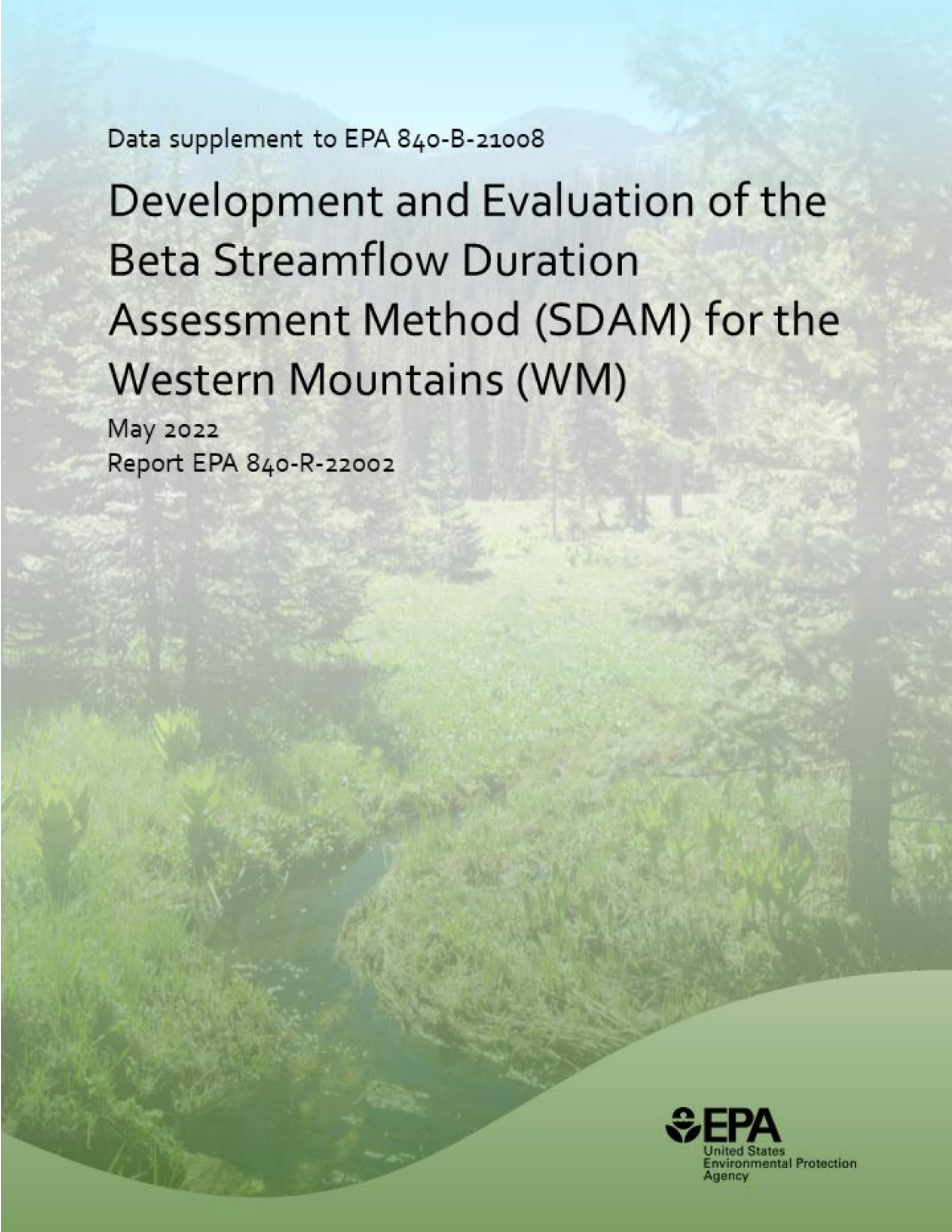

Data supplement to EPA 840-B-21008

Development and Evaluation of the Beta Streamflow Duration Assessment Method (SDAM) for the Western Mountains (WM)

May 2022

Report EPA 840-R-22002

EPA
United States
Environmental Protection
Agency

Development and Evaluation of the Beta Streamflow Duration Assessment Method for the Western Mountains

Data supplement

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

In collaboration with the U.S. Environmental Protection Agency's Streamflow Duration Assessment Method Project Delivery Team:

Ken Fritz
Office of Research and Development
Cincinnati, OH 45268

Brain Topping
Office of Wetlands, Oceans, and Watersheds
Washington, DC 20004

Tracie-Lynn Nadeau
Office of Wetlands, Oceans, and Watersheds
Portland, OR 97205

Julie Kelso, ORISE Fellow
Office of Wetlands, Oceans, and Watersheds
Washington, DC 20004

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. 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. Funding was provided under contract EP-C-17-001 for data management and analysis and EP-C-16-006 for data collection. The views expressed in this report are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Suggested citation:

Mazor, R.D., Fritz, K.M., Topping, B., Nadeau, T.-L., and Kelso, J. 2022. Development and Evaluation of the Beta Streamflow Duration Assessment Method for the Western Mountains. Document No. EPA 840-R-22002.

Introduction

Streamflow duration assessment methods (SDAMs) are rapid, field-based methods to determine flow duration class at the reach scale. The conceptual framework and process steps presented by Fritz and others (2020) were followed to integrate the three key components of an SDAM development study (hydrological data, indicators, and study reaches) and develop a beta SDAM for the Western Mountains (WM; Mazor et al. 2021c).

This supplemental document describes the data collection, data analysis, and evaluation steps that resulted in the beta SDAM WM. The SDAM Project Delivery Team is making this document available to inform public review and comment on the beta method. For a complete description of the beta SDAM WM protocol, please see the User Manual (Mazor et al. 2021c). 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, please see here.

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 followed 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 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, or confinement). A channel is an area that is confined by banks and a bed and contains flowing water (continuously or not).

Overview of the beta method for the Western Mountains

The beta SDAM for the WM uses a small number of indicators to predict the streamflow duration class of stream reaches in the WM. Some indicators are measured through desktop analysis, while others are quantified during a single field visit. The beta SDAM WM results in one of four possible classifications: *ephemeral, intermittent, perennial,* and *at least intermittent*. The *at least intermittent* category occurs when an *intermittent* or *perennial* classification cannot be made with high confidence, but an *ephemeral* classification can be ruled out.

The tool uses a machine learning model known as random forest. Random forest models are increasingly common in the environmental sciences because of their superior performance in handling complex relationships among indicators used to predict classifications. We previously used this approach to develop regional SDAMs for the Arid West (AW; Mazor et al. 2021a, 2021b) and Pacific Northwest (PNW; Nadeau et al. 2015, Nadeau 2015). Because the beta method for the WM includes continuous indicators, the random forest model was not able to be simplified into a decision tree or table, as was done with the beta SDAM AW (Mazor et al. 2021b) and SDAM PNW (Nadeau et al. 2015). Consequently, the random forest model for the beta SDAM WM requires specialized software to run, so we developed an online open-access, user-friendly web application to facilitate efficient and consistent use of the beta SDAM WM protocol for those that do not have access to specialized software.

The degree of snow influence at an assessment reach was used to stratify the WM region (snow-influenced and non-snow influenced areas) because persistent snow can be an important water source affecting flow duration in streams. Snow influence is measured as the mean snow persistence within a 10-km radius of the assessment reach (Hammond et al. 2017). Snow persistence is the fraction of time that snow is present on the ground between January 1 and July 3; for the beta SDAM WM, snow persistence is calculated as the average of the years between 2000 and 2020. Assessment reaches where the mean snow persistence is greater than 25% are 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). Although climate change and annual variation may change the degree of snow influence affecting a reach in any given year, the stratification for this beta method is based on a fixed 21-year time period that should be robust to short-term changes in climate. Snow-influenced areas are prevalent in the Rocky Mountains, as well as at higher elevations in Arizona and the Sierra Nevada of California. Non-snow influenced areas 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 1).

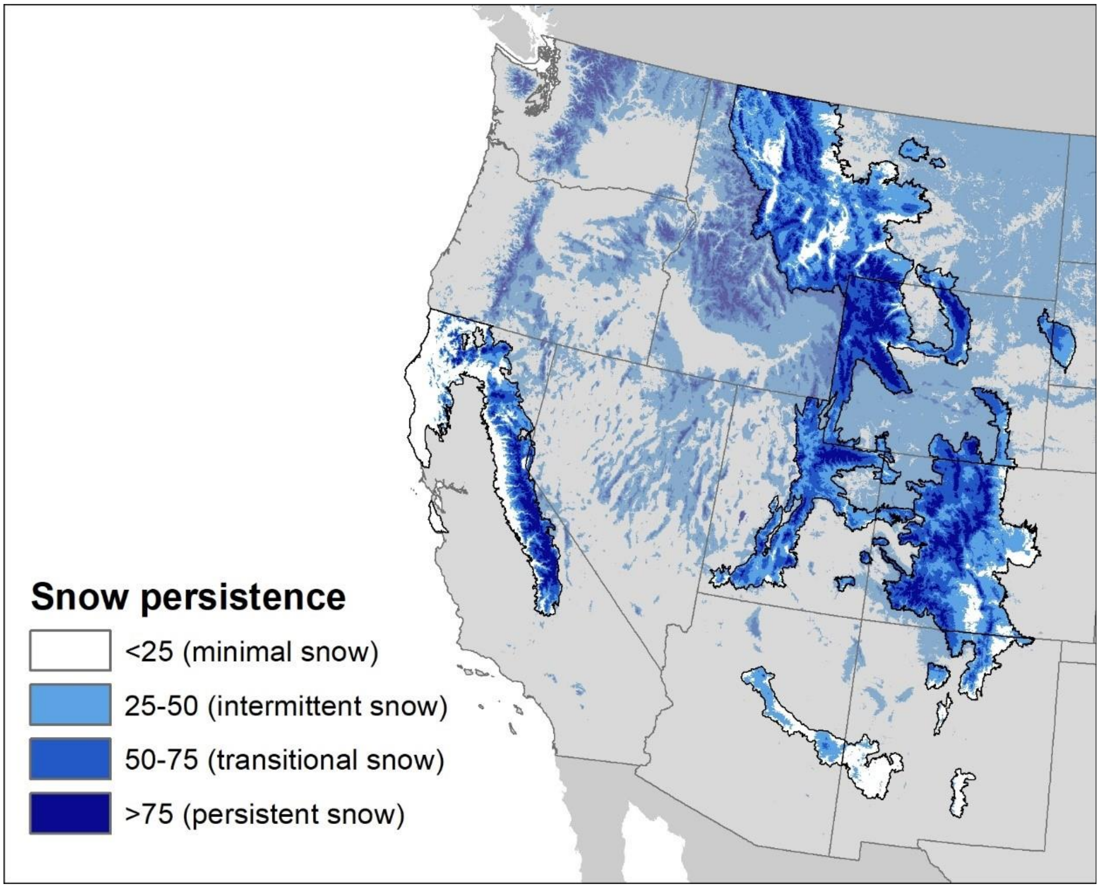

Figure 1. Average snow persistence in the western United States. Data accessed from Hammond et al. (2017). Snow-influenced areas are defined as those with mean snow persistence greater than 25 (i.e., on average, snow is on the ground more than 25% of the time between January 1 and July 3). Portions of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

Methods and Results

Study area

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; for purposes of the current study, portions of the WM region that overlap with the states of Washington, Oregon, and Idaho were excluded (Figure 2; 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 in bordering portions of the AW and 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 (New Mexico Environment Department (NMED) 2011). However, prior to the current study, the rest of the region lacked any tool to classify streamflow duration. Our effort focused on the portion of the WM outside the PNW (Figure 2).

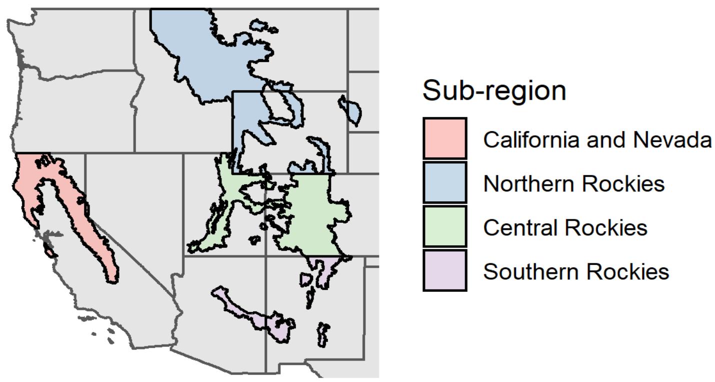

Figure 2. Sub-regions of the WM.

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, Lichvar et al. 2016), excluding the WM region that overlaps with the states of Washington, Oregon, and Idaho. 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.

Development of the Beta SDAM WM

To develop this method, the steps described in Fritz et al. (2020) were followed, as detailed below.

Preparation

At the outset of the project, we assembled a regional steering committee (RSC) consisting of technical staff at Corps Districts and EPA Regional Offices in the WM region that manage programs where streamflow duration information is often needed (e.g., Clean Water Act programs, including permits and enforcement). RSC members were selected based on their expertise in both scientific and programmatic elements relevant to streamflow duration classification needs. The RSC served several functions in the development process, such as reviewing technical products, facilitating connections with local experts, and identifying resources such as sources of hydrologic data.

We identified candidate indicators that were supported by the scientific literature (reviewed in (Mazor and McCune 2021) or used in existing SDAMs developed for portions of the WM; specifically, the New Mexico SDAM (NM method; (NMED 2011), and the SDAM PNW (PNW method; (Nadeau 2015). Following input from the RSC, these candidate indicators were then screened using the criteria described by Fritz and others (2020), including:

- *Consistency*: Does the indicator consistently discriminate among flow duration classes (e.g., demonstrated in multiple studies)?
- *Repeatability:* Can different practitioners take similar measurements, given sufficient training and standardization?
- *Defensibility:* Does the indicator have a rational mechanistic relationship with flow 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?
- *Robustness:* Does human activity complicate indicator measurement or interpretation (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 1) if they met all of the above criteria or were included in the NM or PNW SDAMs to facilitate comparison across the methods (McCune and Mazor 2019).

Identify candidate reaches

We had two objectives in selecting candidate reaches for the WM region covered by this study: first, to include a sufficient number of reaches in each streamflow duration class to characterize variability in indicator measurements; and second, to select reaches representing the range of key natural and disturbance gradients within the region to aid applicability of the method in anticipated conditions across the WM region. To support our goal of geographic representativeness, we established four sub-regional strata in the WM (Figure 2): one stratum for California and Nevada (comprising both the cold Sierra Nevada mountains, and the warmer North Coast of California) and one each for the Southern, Central, and Northern Rocky Mountains. We aimed to select 150 publicly accessible stream-reaches (one assessed location per reach) with equal representation of perennial, intermittent, and ephemeral flow duration classes among and within the four WM sub-regions.

Table 1. Candidate indicators evaluated in the present study. Indicators with "NM" in the Origin column were measured following the NM method protocol (NMED 2011) and indicators marked with "PNW" were measured following the PNW protocol (Nadeau 2015); other indicators (OTH) were measured with protocols developed for this study (available here) and come from sources reviewed in a study by Mazor and McCune (2021) or recommendations from the RSC. Asterisks (*) indicate hydrologic indicators that are considered direct measures of water presence.

Candidate indicator	Description	Origin
Geomorphic indicators		
Sinuosity	Visual estimate of the curviness of the stream channel	NM
Bankfull width	Width of the channel at bankfull height	PNW
Floodplain channel dimensions	Visual estimate of the extent of channel entrenchment and connectivity to the floodplain	NM
Particle size/stream substrate sorting	Visual estimate of the extent of evidence of substrate sorting within the channel	NM
In-channel structure/riffle pool sequence	Visual estimate of the diversity and distinctiveness of riffles, pools, and other flow-based microhabitats	NM
Sediment deposition on plants and debris	Visual estimate of the extent of evidence of sediment deposition on plants and on debris within the floodplain	NM
Hydrologic indicators		
Surface and subsurface flow*	Estimate of the percent of the reach-length with surface and subsurface flow	PNW
Isolated pools*	Number of pools in the channel without any connection to flowing surface water	PNW
Water in channel*	Visual estimate of the extent of surface flow in the channel	NM
Seeps and springs*	Presence/absence of springs or seeps within one-half channel width of the channel	NM
Hydric soils	Presence/absence of hydric soils within the channel, measured at up to three locations	NM
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
Biological indicators		
Live and dead algal cover	Visual estimate of the percent of streambed covered by live or dead algal growth	OTH
Filamentous algal abundance	Estimate of the overall abundance of filamentous algae within the channel	NM
Stream shading	Percent shade-providing cover above the streambed measured with a densiometer at three locations	OTH

Candidate indicator	Description	Origin
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
Fish	Estimate of the overall abundance of fish (other than non-native mosquitofish) in the channel	NM
Aquatic invertebrates	Abundance and richness of aquatic invertebrate families collected from the channel	PNW
Aquatic invertebrates	Estimate of the overall abundance of aquatic invertebrates within the channel	NM
Amphibians	Estimate of the overall abundance of amphibians within the channel	NM
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)	Visual estimate of the distinctiveness of vegetation in the riparian corridor compared to surrounding upland vegetation	NM
Absence of upland rooted plants in the streambed	Visual estimate of the extent of upland rooted plants growing within the streambed	NM
Presence of iron-oxidizing fungi or bacteria	Presence of oily sheens indicative of iron-oxidizing fungi or bacteria within the assessment reach	NM
Presence of aquatic or semi-aquatic snakes	Presence of aquatic or semi-aquatic snakes (e.g., most garter snake species) in the channel	PNW
Geospatial indicators		
Location and watershed characteristics	Latitude, longitude, and elevation	OTH
Long-term normal precipitation and temperature	30-y normal mean annual and monthly precipitation and 30-y normal mean, maximum, and minimum annual temperature (PRISM climate data; Hart and Bell 2015).	OTH
Long-term mean snow persistence between January 1 and July 3	Snow persistence (Hammond et al. 2017)	OTH

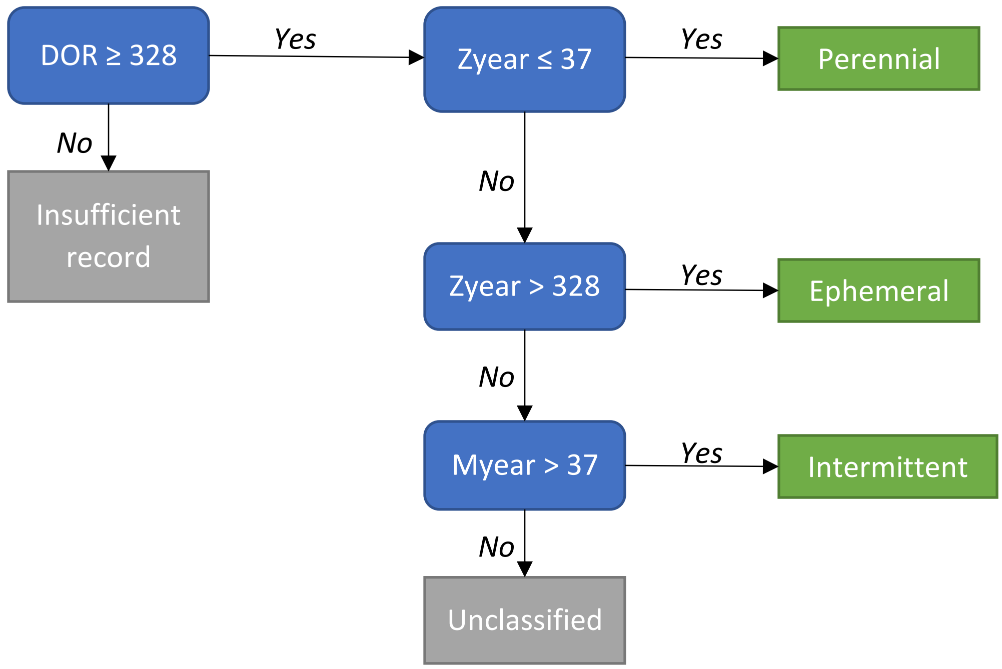

Figure 3. Flowchart used to classify reaches based on continuous measures of water presence (e.g., USGS stream gages). DOR: days of record. Zyear: Average number of dry days per year. Myear: Average length of longest continuous wet period per year, in days. For USGS gages, at least 20 years of data were analyzed whenever possible.

To screen reaches for use in method development, we first compiled a list of 1166 candidate study reaches based on existing hydrologic data records (e.g., U.S. Geological Survey (USGS) stream gages, water presence logger, wildlife cameras, field photos), published studies, and interviews with local experts familiar with the specific reach's hydrology. Most of these reaches (858) were derived from the database of gages operated by the USGS and nearly all of them were perennial (as determined by applying the flowchart in Figure 3). Consequently, other sources were required to identify candidate ephemeral and intermittent reaches. Hydrologic data collected for other purposes (e.g., gages maintained by local flood control agencies, or local natural resource managers) provided another 239 reaches. Published studies and public land management plans yielded 49 candidate reaches and consultation with local experts provided another 30. Whenever possible, multiple sources of hydrologic information were used to confirm classifications. In the resulting set of reaches, 9.6% were determined to be ephemeral, 15.6% were intermittent, and 74.7% were perennial.

Classified reaches were prioritized for study inclusion based on the number and type of data sources available to determine actual streamflow duration classification. Reaches where flow duration could be determined based on multiple data sources (e.g., water presence loggers 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 sub-regions where an insufficient number of “preferred” and “USGS gage” reaches classified as intermittent or ephemeral could be identified.

Of these 1166 reaches, 149 reaches were sampled (31 ephemeral, 66 intermittent, and 52 perennial reaches) in a sampling campaign that ran from July 2019 to October 2020. Post-sampling site classifications were reviewed in light of the data collected, including the Stream Temperature, Intermittence, and Conductance (STIC; Chapin et al. 2014) logger data collected at 48 “baseline” sites that were revisited multiple times over a year (baseline sites are described under Data collection below). If sampling events produced direct observations of stream hydrology inconsistent with the initial classification (e.g., ephemeral reaches flowing during site visits without antecedent precipitation), then field notes and field photos were used to determine reach flow duration. Each of these cases triggered case-by-case review of all available materials by the project delivery team and the RSC to determine if the original classification should remain the same, be updated, or excluded from analysis.

In the final data set of 149 sampled reaches, streamflow duration class was directly determined from USGS stream gage records at 48% of reaches (41 perennial and 30 intermittent reaches, but no ephemeral reaches; **Error! Reference source not found.**, Figure 4**Error! Reference source not found.**). Other sources of hydrologic data used to directly classify study reaches include continuous data loggers (48 reaches), trail cameras, published studies, and consultation with local experts. Multiple sources of hydrologic data were used to classify 47 of the ungaged assessment reaches and a single source was used at 33 ungaged study reaches. In general, more hydrologic data were available at perennial reaches than at intermittent or ephemeral reaches.

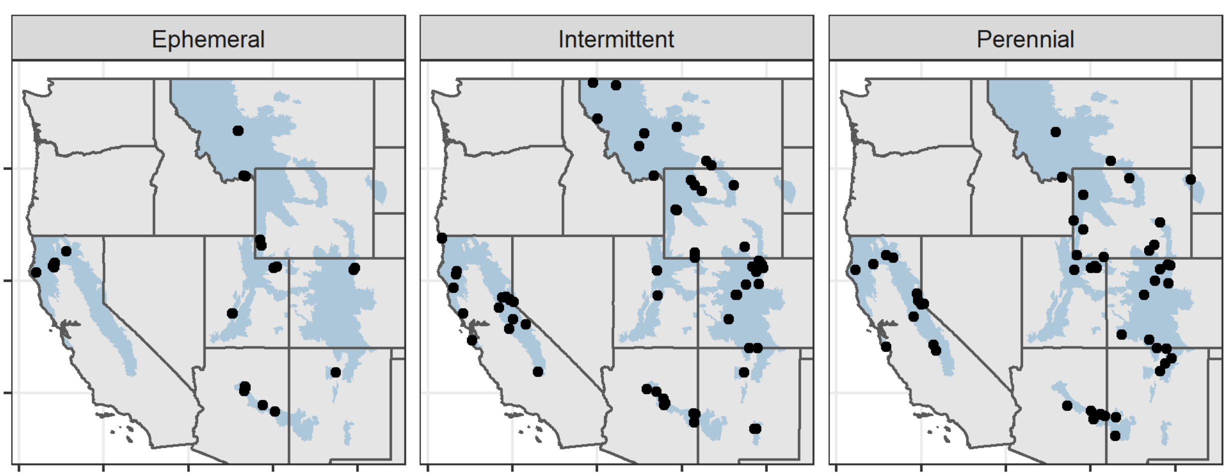

Figure 4. Locations of 31 ephemeral, 66 intermittent, and 52 perennial study stream reaches used to develop the beta SDAM WM.

Table 2. Distribution of sites used to develop the beta SDAM WM. Baseline sites were visited three times throughout the study and had water presence loggers installed and validation sites were visited once throughout the study and did not have loggers installed.

	Validation			Baseline		
Class	Gaged	Preferred	Acceptable	Gaged	Preferred	Total
Ephemeral	*0*	*5*	*22*	*0*	*4*	*31*
-California and Nevada	0	0	8	0	2	10
-Central Rockies	0	2	4	0	1	7
-Northern Rockies	0	0	6	0	0	6
-Southern Rockies	0	3	4	0	1	8
Intermittent	*16*	*10*	*10*	*12*	*18*	*66*
-California and Nevada	5	2	1	5	5	18
-Central Rockies	2	4	3	0	8	17
-Northern Rockies	6	0	6	2	4	18
-Southern Rockies	3	4	0	5	1	13
Perennial	*31*	*6*	*1*	*10*	*4*	*52*
-California and Nevada	9	0	0	4	0	13
-Central Rockies	4	5	1	0	2	12
-Northern Rockies	9	1	0	3	1	14
-Southern Rockies	9	0	0	3	1	13

Data collection

Reaches were sampled following the development protocol (available here and in the supplementary material of Mazor et al. 2021c), which covers measurement of indicators identified in Mazor and McCune (2021), as well as “Level 1” indicators of the NM method (NMED 2011), and all indicators of the PNW method (Nadeau 2015). STIC loggers (Chapin et al. 2014) were deployed at 48 “baseline” reaches and were revisited a total of three times each over a year; “validation” sites were visited once and did not have loggers. For further details on STIC data loggers and their verification/calibration, deployment, and data retrieval, see Schumacher and Fritz (2019). The sampling protocol used in this study was identical to that used to develop the beta SDAM AW. Mazor et al. (2021a) provides a summary of these data collection protocols. Sampled study sites are shown in Figure 4. Forty-two of these study sites were noted as disturbed by human activity (e.g., channelization, discharges, diversions) by field crews.

Data analysis

Metric calculation

Candidate indicator data were used to calculate 72 candidate metrics: 37 biological metrics, 7 geomorphological metrics, 8 hydrologic metrics (7 of which were direct measures of water presence), and 20 geospatial metrics (Table 3).

Table 3. Metrics evaluated for the development of the beta SDAM WM. PctDom: Percent of observations with the most common value (typically zero). PvIvE: F-statistic from a comparison of mean values at perennial, intermittent, and ephemeral reaches. Absolute t-statistic from a comparison of mean values at ephemeral and at least intermittent reaches (EvALI), at perennial and non-perennial reaches (PvNP), at flowing intermittent and perennial reaches (PvIwet), and at non-flowing intermittent and ephemeral reaches (EvIdry). rf_MDA: Variable importance from a random forest model, measured as mean decrease in accuracy. Screen: Indicates if the metric passed or failed screening criteria in Table 4. Ord: Ordinal metrics. Bin: Binary metrics. Con: Continuous metrics. Asterisks (*) indicate hydrologic metrics that directly measure the presence of water. NM: Metrics derived from candidate indicators used in the SDAM NM. OBL and FACW: Obligate and facultative-wet wetland indicator plants, respectively (Lichvar et al. 2016). EPT: Ephemeroptera, Plecoptera, and Trichoptera insect orders. GOLD: Gastropoda, Oligochaeta, and Diptera invertebrate groups. OCH: Odonata, Coleoptera, and Heteroptera insect orders.

Metric	Description	Form	PctDom	Range	PvIvE	EvALI	PvNP	PvIwet	EvIdry	rf_MDA	Screen
Biological											
fishabund_score2	Abundance of fish, excluding mosquitofish (NM)	Ord	73%	3	8.91	7.09	3.01	0.51	1.44	0.0004	Pass
DifferencesInVegetation_score	Differences in vegetation between the riparian corridor and adjacent uplands score (NM)	Ord	34%	3	31.10	6.28	6.35	1.47	1.93	0.0027	Pass
UplandRootedPlants_score	Absence of upland rooted plants in the streambed score (NM)	Ord	47%	3	6.04	2.60	3.29	0.20	0.53	-0.0003	Pass
iofb_score	Presence of iron-oxidizing bacteria and fungi score (NM)	Bin	85%	1.5	6.30	5.18	2.74	0.96	1.00	0.0003	Pass
mayfly_abundance	Abundance of mayflies	Con	47%	66	52.47	10.78	8.41	4.18	1.16	0.0111	Pass
perennial_abundance	Abundance of perennial indicator taxa	Con	58%	90	16.06	6.48	4.72	2.23	1.56	0.0062	Pass
perennial_taxa	Number of perennial indicator taxa	Con	58%	14	16.27	7.21	4.75	1.67	1.02	0.0007	Pass
perennial_live_abundance	Abundance of perennial indicator taxa (living specimens only)	Con	58%	90	15.90	6.44	4.71	2.19	1.17	0.0063	Pass
snake_score	Presence of aquatic snakes	Bin	97%	1	0.31	0.27	0.72	0.10	1.00	0.0000	Fail
vert_score	Presence of aquatic vertebrates	Bin	86%	1	1.66	0.88	1.67	0.32	1.79	0.0001	Fail
vert_sumscore	Number of aquatic vertebrate types present (fish, amphibians, snakes, turtles)	Ord	92%	2	0.48	0.16	0.80	0.15	1.36	0.0001	Fail
hydrophytes_present	Number of OBL and FACW plant species present in the channel or within a half-channel width of the channel	Ord	20%	13	15.71	5.41	4.65	0.76	0.24	0.0012	Pass

Metric	Description	Form	PctDom	Range	PvIvE	EvALI	PvNP	PvIwet	EvIdry	rf_MDA	Screen
hydrophytes_present_noflag	Number of OBL and FACW plant species present in the channel or within a half-channel width of the channel (excluding those with a flagged unusual distribution)	Ord	20%	13	14.79	5.37	4.44	0.58	0.09	0.0001	Pass
alglivedead_cover_score	Cover of live or dead algae on the streambed	Ord	34%	4	45.84	10.23	8.03	2.06	2.51	0.0049	Pass
moss_cover_score	Moss cover on the streambed	Ord	63%	3	0.21	0.26	0.65	0.85	0.35	0.0000	Fail
liverwort_cover_score	Liverwort cover on the streambed	Ord	88%	3	1.38	2.46	1.20	0.57	1.06	-0.0001	Pass
PctShading	Percent shading on the streambed	Con	8%	1	2.54	0.66	2.25	1.90	0.13	0.0001	Pass
TotalAbundance	Total abundance of aquatic invertebrates	Con	21%	287	35.93	9.72	7.09	2.73	0.11	0.0077	Pass
Richness	Total richness of aquatic invertebrate families	Con	21%	36	45.87	10.24	8.53	2.87	0.13	0.0067	Pass
EPT_abundance	Abundance of EPT	Con	34%	150	37.86	9.62	7.27	3.16	0.84	0.0107	Pass
EPT_taxa	Number of EPT families	Con	34%	27	37.14	9.86	7.38	2.82	0.90	0.0095	Pass
EPT_relabd	Relative abundance of EPT families	Con	34%	1	13.78	3.86	5.36	2.27	0.86	0.0021	Pass
EPT_reltaxa	Relative richness of EPT families	Con	34%	2	16.42	5.83	5.42	1.94	1.39	0.0013	Pass
GOLD_abundance	Abundance of GOLD	Con	33%	91	31.93	9.39	6.62	2.17	0.30	0.0025	Pass
GOLD_taxa	Number of GOLD families	Con	33%	14	32.04	10.28	6.59	1.70	0.27	0.0012	Pass
OCH_abundance	Abundance of OCH	Con	44%	74	10.19	5.26	3.86	1.19	1.03	-0.0002	Pass
OCH_taxa	Numer of OCH families	Con	44%	11	9.61	4.39	3.82	0.62	1.03	-0.0010	Pass
GOLD_relabd	Relative abundance of GOLD taxa	Con	33%	1	4.11	2.40	2.12	1.56	0.18	0.0030	Pass
GOLD_reltaxa	Relative richness of GOLD taxa	Con	33%	1	6.66	3.44	2.48	1.66	0.06	0.0022	Pass
OCH_relabd	Relative abundance of OCH taxa	Con	44%	1	0.06	0.04	0.36	0.02	0.32	-0.0001	Fail
OCH_reltaxa	Relative richness of OCH taxa	Con	44%	1	0.03	0.18	0.06	0.76	0.21	0.0002	Fail
GOLDOCH_relabd	Relative abundance of GOLD and OCH taxa	Con	27%	1	2.93	2.00	1.58	1.51	0.06	0.0011	Pass
GOLDOCH_reltaxa	Relative richness of GOLD and OCH taxa	Con	27%	1.4	4.75	2.74	2.01	1.94	0.09	0.0008	Pass
Noninsect_abundance	Abundance of non-insect taxa	Con	50%	87	6.76	5.02	2.87	0.40	0.37	0.0001	Pass

Metric		Description	Form	PctDom	Range	PvIvE	EvALI	PvNP	PvIwet	EvIdry	rf_MDA	Screen
Noninsect_taxa		Richness of non-insect taxa	Con	50%	11	7.34	5.68	2.81	0.15	0.05	0.0001	Pass
Noninsect_relabund		Relative abundance of non-insect taxa	Con	50%	1	0.37	0.30	0.69	1.21	0.21	0.0003	Fail
Noninsect_reltaxa		Relative richness of non-insect taxa	Con	50%	1	0.54	0.62	0.53	1.37	0.22	-0.0009	Fail
Geomorphological												
Sinuosity_score		Channel sinuosity score (NM)	Ord	33%	3	4.30	1.52	2.76	1.42	0.68	0.00	Pass
ChannelDimensions_score		Channel dimensions score (NM)	Ord	37%	3	0.52	0.97	0.31	0.57	0.28	0.00	Fail
RifflePoolSeq_score		Riffle-pool sequence score (NM)	Ord	31%	3	11.92	2.66	5.07	2.48	0.09	0.00	Pass
SubstrateSorting_score		Substrate sorting score (NM)	Ord	33%	3	8.64	2.78	4.14	2.11	0.56	0.00	Pass
SedimentOnPlantsDebris_score		Sediment on plants and debris score (NM)	Ord	91%	1.5	0.43	0.70	0.97	0.42	0.38	0.00	Fail
BankWidthMean		Mean bank-width	Ord	2%	48	8.54	5.29	3.44	1.67	1.32	0.01	Pass
Slope		Valley slope	Ord	15%	26	1.24	1.48	0.61	0.69	0.43	0.00	Fail
Hydrologic												
WaterInChannel_score	*	Water in channel score (NM)	Ord	48%	6	110.02	17.82	13.98	3.88	1.85	0.03	Pass
HydricSoils_score		Presence of hydric soils in the channel score (NM)	Bin	76%	3	8.20	3.77	3.42	1.89	0.97	0.00	Pass
springs_score	*	Presence of springs or seeps in the channel score (NM)	Bin	98%	3	0.63	1.00	1.00	1.00	0.00	0.00	Fail
SurfaceFlow_pct	*	Percent of reach with flowing surface water	Ord	50%	100	102.77	19.20	13.77	3.53	1.00	0.03	Pass
SurfaceSubsurfaceFlow_pct	*	Percent of reach with flowing surface or subsurface water	Ord	86%	100	6.66	4.20	2.04	3.14	1.97	0.00	Pass
IsolatedPools_number	*	Number of isolated pools (no connection to flowing surface water)	Ord	89%	9	3.49	0.78	2.92	1.53	1.84	0.00	Pass
WoodyJams_number		Number of woody jams in the reach	Ord	79%	10	0.98	0.22	1.42	0.49	1.10	0.00	Fail
SoilMoist_MaxScore	*	Maximum soil moisture score in the reach	Ord	72%	2	55.23	8.53	8.44	0.00	1.94	0.01	Pass
Geospatial												
Elev_m		Elevation	Con	3%	3250	2.11	1.97	1.49	0.41	0.89	0.00	Pass
tmean		Mean annual temperature	Con	3%	17	1.66	1.95	0.92	0.34	0.67	0.00	Fail
tmax		Maximum annual temperature	Con	2%	18	1.97	2.17	0.39	1.11	0.76	0.00	Pass
tmin		Minimum annual temperature	Con	2%	17	1.56	1.63	1.40	0.43	0.54	0.00	Fail

Metric	Description	Form	PctDom	Range	PvIvE	EvALI	PvNP	PvIwet	EvIdry	rf_MDA	Screen
MeanSnowPersistence_10	Mean snow persistence within a 10-km radius of the reach	Con	1%	82	2.69	2.52	1.45	0.00	0.91	0.00	Pass
MeanSnowPersistence_05	Mean snow persistence within a 5-km radius of the reach	Con	1%	86	2.97	2.67	1.35	0.17	1.11	0.00	Pass
MeanSnowPersistence_01	Mean snow persistence within a 1-km radius of the reach	Con	1%	84	2.53	2.51	1.18	0.25	1.02	0.00	Pass
ppt	Mean annual precipitation	Con	2%	1603	0.80	0.23	1.38	0.76	0.54	0.00	Fail
ppt.m01	Mean January precipitation	Con	2%	337	0.90	0.80	1.44	0.38	0.33	0.00	Fail
ppt.m02	Mean February precipitation	Con	2%	293	0.50	0.35	1.09	0.33	0.53	0.00	Fail
ppt.m03	Mean March precipitation	Con	2%	254	0.49	0.41	1.06	0.42	0.37	0.00	Fail
ppt.m04	Mean April precipitation	Con	2%	143	1.08	0.82	0.99	1.33	0.66	0.00	Fail
ppt.m05	Mean May precipitation	Con	2%	107	1.93	1.29	1.07	2.28	0.36	0.00	Pass
ppt.m06	Mean June precipitation	Con	3%	129	2.20	1.51	0.99	2.15	0.53	0.00	Pass
ppt.m07	Mean July precipitation	Con	2%	102	0.37	0.86	0.57	0.27	0.53	0.00	Fail
ppt.m08	Mean August precipitation	Con	2%	131	0.05	0.06	0.31	0.50	0.17	0.00	Fail
ppt.m09	Mean September precipitation	Con	2%	80	0.49	0.66	0.93	0.57	0.17	0.00	Fail
ppt.m10	Mean October precipitation	Con	2%	102	0.08	0.33	0.20	0.16	0.38	0.00	Pass
ppt.m11	Mean November precipitation	Con	2%	247	0.80	0.44	1.44	0.73	0.35	0.00	Fail
ppt.m12	Mean December precipitation	Con	2%	367	0.74	0.68	1.34	0.39	0.32	0.00	Fail

Metric screening

As an initial data exploration step, we visualized the relationships between streamflow duration class (hereafter “flow class”) and indicators by ordinating all 72 metrics for all samples in the data set in a nonmetric multidimensional scaling using Gowers’ distance. Convex hulls were drawn around each streamflow duration class to help visualize their distributions in ordination space. The 2-axis ordination was computed using the metaMDS function in the *vegan* R package (Oksanen et al. 2019). Correlation coefficients (Spearman’s rho) were calculated between ordination axes and metric values. Wet and dry reaches were plotted separately to evaluate the role of flow conditions at the time of the visit on flow duration indicators; streams with scores 4 and higher for the “Water in channel” indicator (WaterInChannel_score) from the NM SDAM were considered wet and scores 3 or lower were considered dry.

The ordination showed that perennial and ephemeral reaches were quite distinct, but intermittent reaches overlapped considerably with the other classes (Figure 5). In general, intermittent reaches that were dry on collection dates were similar to ephemeral reaches and

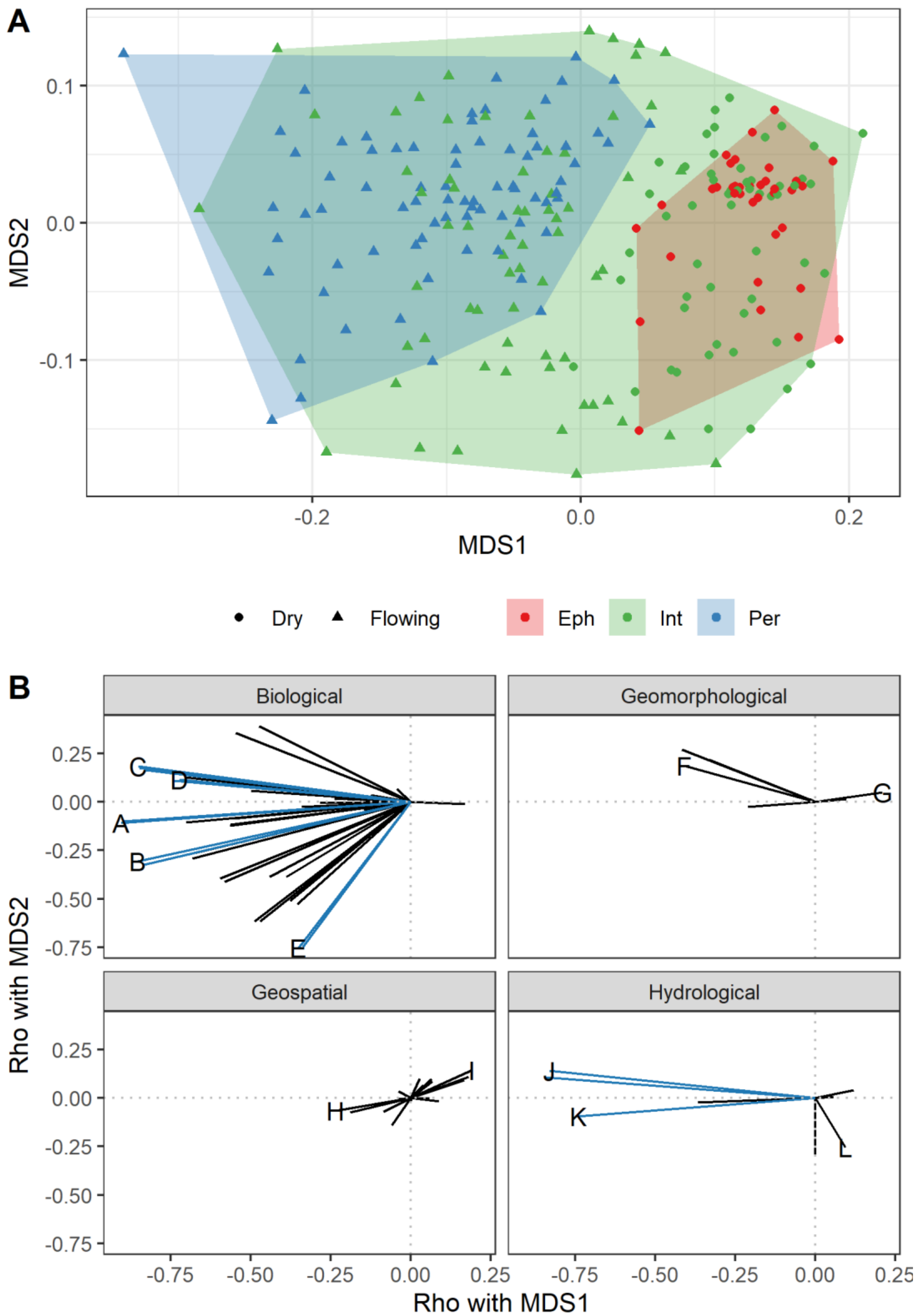

Figure 5. A two-axis nonmetric multidimensional scaling of metrics based on biological, geomorphic, geospatial, and hydrologic indicators. Panel A shows individual reaches. MDS: Multidimensional scaling axis 1 or 2. Eph: Ephemeral reaches. Int: Intermittent reaches. Per: Perennial reaches. Circle: Reaches were dry during the site visit. Triangle: reaches were flowing during the site visit. Panel B shows correlations (Spearman's rho) between selected metrics and ordination axis scores; metrics with rho^2 > 0.5 are highlighted in blue (no geomorphological or geospatial metrics had rho^2 > 0.5, nor did any metric have rho^2> 0.5 with

the second axis). Selected metrics are labeled: Biological metrics: A: Total aquatic invertebrate abundance. B: GOLD abundance. C: EPT abundance. D: Perennial indicator taxa abundance. E: GOLDOCH relative richness. Geomorphological metrics: F: Bank width. G: Slope. Geospatial metrics: H: Mean snow persistence within 10 km. I: Mean annual maximum temperature. Hydrologic metrics. J: Percent of reach with surface flow. K: Soil moisture. L: Number of isolated pools.

intermittent reaches that had surface flow on collection dates were similar to perennial reaches. Hydrologic and biological metrics were among the most strongly correlated with ordination axes and no geomorphological or geospatial metric correlated with an ordination axis with a rho2 greater than 0.5.

Metrics were evaluated using several criteria for inclusion in the beta SDAM (Table 4). We developed criteria following approaches for screening metrics in bioassessment indices (e.g., Stoddard et al. 2008) and applied them to data from initial reach-visits (i.e., data from revisits were withheld from analysis). One criterion was a distribution statistic, calculated as percent dominance of the most common value (which was typically zero); all metrics had to meet this criterion. The remaining criteria measured responsiveness of metrics (i.e., ability to discriminate across flow classes). Most of these measures were based on 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). Another responsiveness statistic was based on variable importance (specifically, mean decrease in accuracy) from a random forest model to predict streamflow duration class from all candidate metrics; the model was calibrated using the default option from the randomForest function in the *randomForest* package in R (Liaw and Wiener 2002). Metrics had to meet at least one responsiveness criterion to be considered in further analyses. A total of 47 of the 72 candidate metrics met these criteria and were considered as screened metrics.

Table 4. 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 data set
Responsiveness criteria		
PvIvE	F>2	F-statistic in a comparison of values at perennial versus intermittent versus ephemeral reaches
EvALI	t>2	t-statistic in a comparison of values at ephemeral versus at least intermittent reaches
PvNP	t>2	t-statistic in a comparison of values at perennial versus non-perennial reaches
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

Metric selection

The screened metrics were reduced to a final set of metrics for the beta SDAM based on their importance in random forest models using the recursive feature elimination (rfe) function in the R *caret* package (Kuhn 2020). 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 iteratively by eliminating the least important metrics. We considered the most complex model (i.e., 47 candidate metrics included) then iteratively eliminating 5 variables at a time in each step based on low variable importance until a 20-variable model was identified; after this point, only one variable was eliminated in each step. The best performing model (i.e., highest accuracy in predicting streamflow duration class) was identified. Then, the simplest model (i.e., the one with the fewest variables) with accuracy within 1% of the best was selected to identify the final set of metrics. If the best-performing model selected by this approach had more than 20 variables, the 20-variable model was selected. For this analysis, accuracy was measured with Cohen's Kappa statistic —a measure of accuracy that accounts for uneven distribution among the three streamflow duration classes.

We applied this modeling process to different subsets of the dataset, including:

- the full region-wide dataset;
- datasets stratified by sub-regions shown in Figure 2 (4 total); and
- datasets stratified into snow-influenced and non-snow influenced sites, based on mean snow persistence greater than 25% calculated for a 1-km, 5-km, and 10-km buffer from the sampling reach (2 strata for each of 3 buffers).

For each subset, the modeling process was implemented:

- with or without considering geospatial metrics; and
- with or without considering metrics based on direct measures of water presence.

There are advantages and disadvantages to including these metrics in an SDAM and thus we evaluated options with and without them. Geospatial metrics may improve SDAM performance but would require GIS analysis to use the resulting method. Direct measures of water presence can also greatly increase performance, but this introduces circularity (because water presence was used to confirm and update streamflow duration classes in the development data set) and may degrade the ability of the SDAM to work during atypical conditions, such as drought. See (Mazor et al. (2021b) for a discussion of the implications of including geospatial metrics and direct measures of water presence in SDAMs.

To explore all these options, we developed 20 sets of models for different subsets of reaches and combinations of predictors, with sets including between 1 and 5 models (44 models total; Table 5). Analyses were conducted on data from the initial reach visits alone. For each of the 20 models, data were split into 80% training and 20% testing data sets, stratified by the 4 sub-regions and 3 streamflow duration classes. Model design characteristics and optimal number of metrics selected by rfe are shown in Table 5 and the selected metrics for each model are shown in Figure 6.

Table 5. Design characteristics of the 44 models. H2O: included direct measures of water presence. GIS: included geospatial metrics. n sites: number of sites used in model training, testing, and evaluated for repeatability (revisit). rfe accuracy: accuracy of best model produced by recursive feature elimination (rfe), measured as Cohen's Kappa or as out-of-bag (OOB) accuracy.

				n reaches				rfe accuracy	
Model set	Stratum	H2O	GIS	training	testing	revisit	# metrics	Kappa	OOB
Unstratified models									
Unstrat	None			117	32	84	19	0.41	0.45
Unstrat GIS	None		Yes	117	32	84	15	0.41	0.38
Unstrat H2O	None	Yes		117	32	84	16	0.47	0.38
Unstrat H2O GIS	None	Yes	Yes	117	32	84	3	0.47	0.28
Models stratified by region									
Strat	California & Nevada			32	9	25	20	0.36	0.47
Strat	Central Rockies			27	9	21	18	0.52	0.26
Strat	Northern Rockies			29	9	19	19	0.27	0.41
Strat	Southern Rockies			26	8	19	3	0.51	0.31
Strat GIS	California & Nevada		Yes	32	9	25	16	0.38	0.28
Strat GIS	Central Rockies		Yes	27	9	21	20	0.52	0.41
Strat GIS	Northern Rockies		Yes	29	9	19	16	0.42	0.41
Strat GIS	Southern Rockies		Yes	26	8	19	3	0.45	0.38
Strat H2O	California & Nevada	Yes		32	9	25	3	0.59	0.28
Strat H2O	Central Rockies	Yes		27	9	21	14	0.54	0.26
Strat H2O	Northern Rockies	Yes		29	9	19	10	0.39	0.45
Strat H2O	Southern Rockies	Yes		26	8	19	3	0.52	0.31
Strat H2O GIS	California & Nevada	Yes	Yes	32	9	25	3	0.51	0.25
Strat H2O GIS	Central Rockies	Yes	Yes	27	9	21	20	0.51	0.3
Strat H2O GIS	Northern Rockies	Yes	Yes	29	9	19	20	0.25	0.31
Strat H2O GIS	Southern Rockies	Yes	Yes	26	8	19	8	0.49	0.31
Models stratified by snow influence									
Snow influence within 1 km									
Snow01	Not snow-dominated			46	13	36	20	0.42	0.43
Snow01	Snow-dominated			71	19	48	8	0.35	0.35
Snow01 GIS	Not snow-dominated		Yes	46	13	36	20	0.37	0.41
Snow01 GIS	Snow-dominated		Yes	71	19	48	18	0.35	0.32
Snow01 H2O	Not snow-dominated	Yes		46	13	36	3	0.57	0.3
Snow01 H2O	Snow-dominated	Yes		71	19	48	20	0.33	0.38

				n reaches				rfe accuracy	
Model set	Stratum	H2O	GIS	training	testing	revisit	# metrics	Kappa	OOB
Snow01 H2O GIS	Not snow-dominated	Yes	Yes	46	13	36	6	0.6	0.28
Snow01 H2O GIS	Snow-dominated	Yes	Yes	71	19	48	10	0.48	0.35
Snow influence within 5 km									
Snow05	Not snow-dominated			40	11	28	20	0.38	0.43
Snow05	Snow-dominated			77	21	56	15	0.48	0.36
Snow05 GIS	Not snow-dominated		Yes	40	11	28	14	0.33	0.4
Snow05 GIS	Snow-dominated		Yes	77	21	56	13	0.43	0.29
Snow05 H2O	Not snow-dominated	Yes		40	11	28	13	0.48	0.25
Snow05 H2O	Snow-dominated	Yes		77	21	56	20	0.43	0.3
Snow05 H2O GIS	Not snow-dominated	Yes	Yes	40	11	28	4	0.54	0.3
Snow05 H2O GIS	Snow-dominated	Yes	Yes	77	21	56	17	0.48	0.3
Snow influence within 10 km									
Snow10	Not snow-dominated			39	11	31	13	0.24	0.49
Snow10	Snow-dominated			78	21	53	6	0.43	0.45
Snow10 GIS	Not snow-dominated		Yes	39	11	31	17	0.22	0.31
Snow10 GIS	Snow-dominated		Yes	78	21	53	11	0.52	0.33
Snow10 H2O	Not snow-dominated	Yes		39	11	31	5	0.54	0.31
Snow10 H2O	Snow-dominated	Yes		78	21	53	8	0.41	0.33
Snow10 H2O GIS	Not snow-dominated	Yes	Yes	39	11	31	4	0.42	0.31
Snow10 H2O GIS	Snow-dominated	Yes	Yes	78	21	53	13	0.52	0.24

Biological metrics (particularly those based on aquatic invertebrates) were among the most widely selected metrics across model sets. Among non-biological metrics, mean bankfull width was the only frequently selected geomorphological metric. Direct measures of water presence were selected every time these measures were eligible for selection. Among geospatial metrics, October precipitation was the most frequently selected metric (Figure 6).

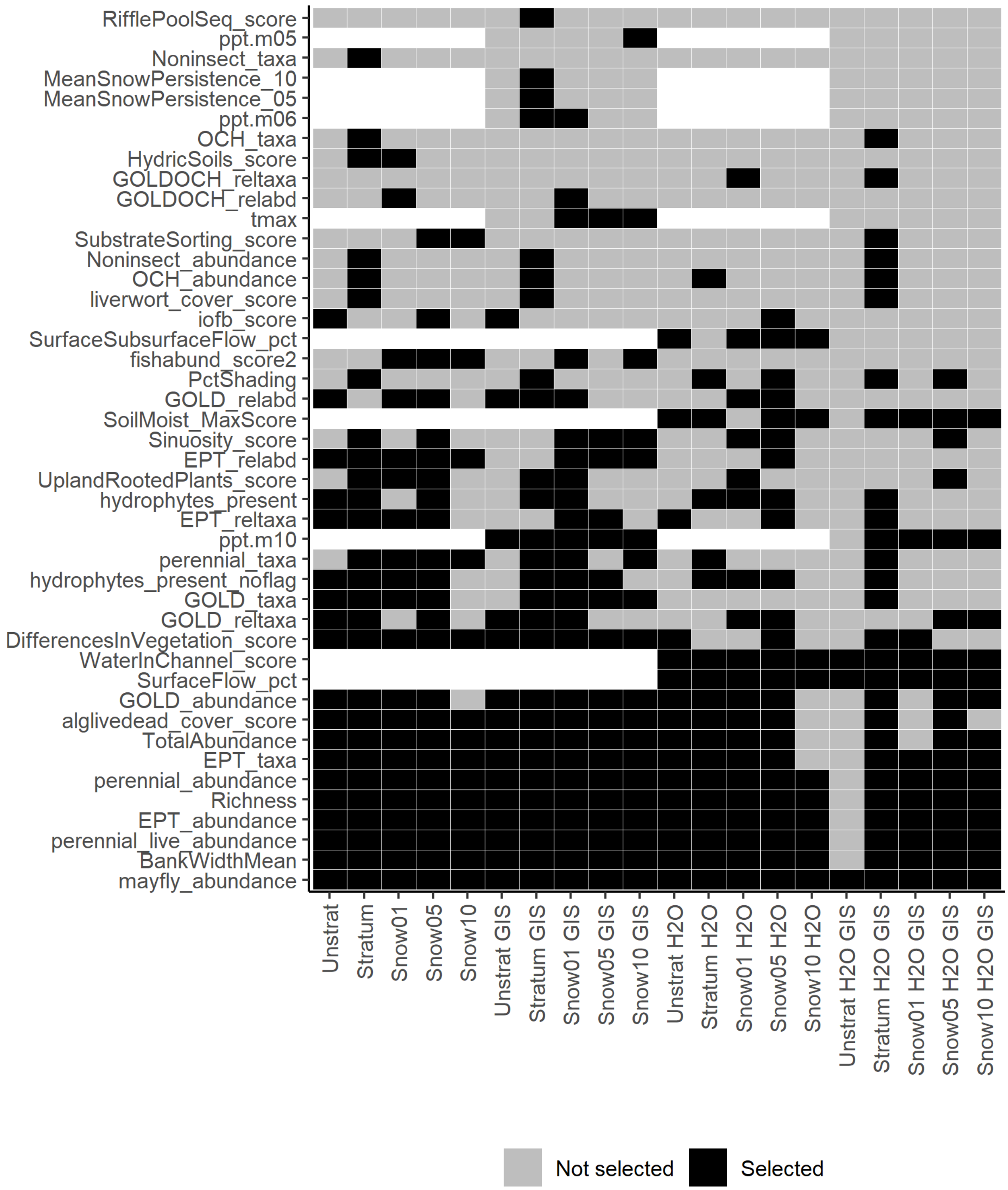

Figure 6. Metrics (left) selected by RFE for each model set (bottom). White tiles indicate that a metric was ineligible for selection in that model set (e.g., the water in channel score was ineligible for models that did not allow direct measures of water presence). X-axis labels refer to model sets described in Table 5; Y-axis labels refer to metrics described in Table 3.

Preliminary model calibration and performance assessment

Random forest models were then fit for each of the 20 options using the randomForest function in the *randomForest* package in R (Liaw and Wiener 2002) using default parameters, except that the number of trees was set to 1500 instead of the default 500. Only the initial visit for reaches in the calibration data set was used for model fitting.

Model performance evaluation focused on two aspects: accuracy and repeatability. Accuracy was assessed by calculating the same comparisons used to evaluate metric responsiveness during the metric screening phase (e.g., ephemeral versus at least intermittent reaches, perennial versus wet intermittent reaches, etc.; Table 4). Accuracy was measured using the initial reach-visit in both the calibration training and testing data sets independently. We compared training and testing measures to see if models validated poorly, suggesting that they may be overfit.

Repeatability was assessed using data from the 48 reaches that were revisited (i.e., Baseline sites; **Error! Reference source not found.**) and was calculated as the percent of reaches where model classifications from visits were the same (regardless of classification accuracy). Due to the limited amount of data, repeatability was only assessed on a region-wide basis and not within each subregion; it was not analyzed separately for calibration and validation reaches. Performance of the beta SDAM AW, SDAM PNW, and SDAM NM was also evaluated within the training data set.

SDAM models newly developed through the current effort had better performance than previously developed SDAMs (especially the beta SDAM AW), but among the new models, performance was similar and there was no clear best model set (Table 6, Figure 7 and Figure 8). Stratified model sets performed slightly better than the unstratified models and there were modest improvements in accuracy achieved by including geospatial metrics, as well as direct measures based on water presence. The RSC recommended the model set stratified by snow influence calculated within a 10-km radius; furthermore, the RSC opted for the models that included geospatial metrics (i.e., model set Snow 10 GIS) but did not recommend including direct measures of water presence due to the potential introduction of circularity (water presence during field visits was sometimes used to inform or verify the direct flow classification of stream reaches), as described above.

Table 6. Performance of the 20 model sets evaluated. PvIvE: Percent of reaches classified correctly as perennial, intermittent, or ephemeral. EvALI: Percent of reaches classified correctly as ephemeral or at least intermittent. PvNP: Percent of reaches classified correctly as perennial or non-perennial. PvIwet: Percent of flowing reaches classified correctly as perennial or intermittent. IvEdry: Percent of dry reaches correctly classified as intermittent or ephemeral. Train: Result for training data. Test: Result for testing data. Model sets are described in Table 5. AW: Results for the beta SDAM AW. NM: Results for the SDAM NM. PNW: Results for the SDAM PNW.

	Accuracy										
	PvIvE		EvALI		PvNP		PvIwet		IvEdry		
Model set	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Precision
AW	0.39		0.79		0.45		0.48		0.25		0.67
NM	0.58		0.8		0.72		0.66		0.46		0.87
PNW	0.57		0.79		0.78		0.64		0.46		0.82
Snow10 H2O GIS	0.74	0.59	0.88	0.81	0.85	0.78	0.76	0.63	0.7	0.54	0.81
Snow05 H2O GIS	0.7	0.75	0.86	0.88	0.84	0.88	0.72	0.81	0.67	0.64	0.83
Snow01 H2O GIS	0.68	0.78	0.88	0.88	0.79	0.91	0.66	0.84	0.7	0.69	0.82
Stratum H2O GIS	0.71	0.6	0.82	0.83	0.89	0.77	0.79	0.65	0.6	0.5	0.84
Unstrat H2O GIS	0.72	0.69	0.87	0.84	0.85	0.84	0.75	0.74	0.67	0.62	0.8
Snow10 H2O	0.68	0.66	0.86	0.78	0.81	0.88	0.69	0.78	0.64	0.5	0.8
Snow05 H2O	0.72	0.5	0.9	0.81	0.82	0.69	0.7	0.5	0.74	0.5	0.83
Snow01 H2O	0.65	0.69	0.85	0.81	0.79	0.88	0.66	0.79	0.63	0.54	0.82
Stratum H2O	0.68	0.6	0.83	0.77	0.84	0.83	0.76	0.7	0.55	0.47	0.83
Unstrat H2O	0.62	0.75	0.85	0.88	0.77	0.88	0.62	0.79	0.63	0.69	0.8
Snow10 GIS	0.68	0.63	0.89	0.81	0.78	0.81	0.66	0.7	0.7	0.5	0.83
Snow05 GIS	0.68	0.69	0.85	0.88	0.83	0.81	0.72	0.68	0.61	0.69	0.84
Snow01 GIS	0.64	0.59	0.85	0.81	0.79	0.78	0.65	0.67	0.62	0.5	0.84
Stratum GIS	0.63	0.57	0.82	0.8	0.81	0.77	0.69	0.65	0.55	0.42	0.8
Unstrat GIS	0.62	0.53	0.81	0.81	0.8	0.72	0.68	0.47	0.5	0.6	0.84
Snow10	0.54	0.69	0.79	0.84	0.74	0.84	0.59	0.63	0.44	0.75	0.73
Snow05	0.62	0.69	0.85	0.88	0.75	0.81	0.59	0.68	0.65	0.69	0.77
Snow01	0.62	0.69	0.84	0.88	0.78	0.81	0.63	0.65	0.6	0.75	0.78
Stratum	0.63	0.46	0.82	0.8	0.8	0.66	0.66	0.42	0.58	0.5	0.83
Unstrat	0.55	0.63	0.79	0.81	0.74	0.81	0.58	0.67	0.49	0.57	0.79

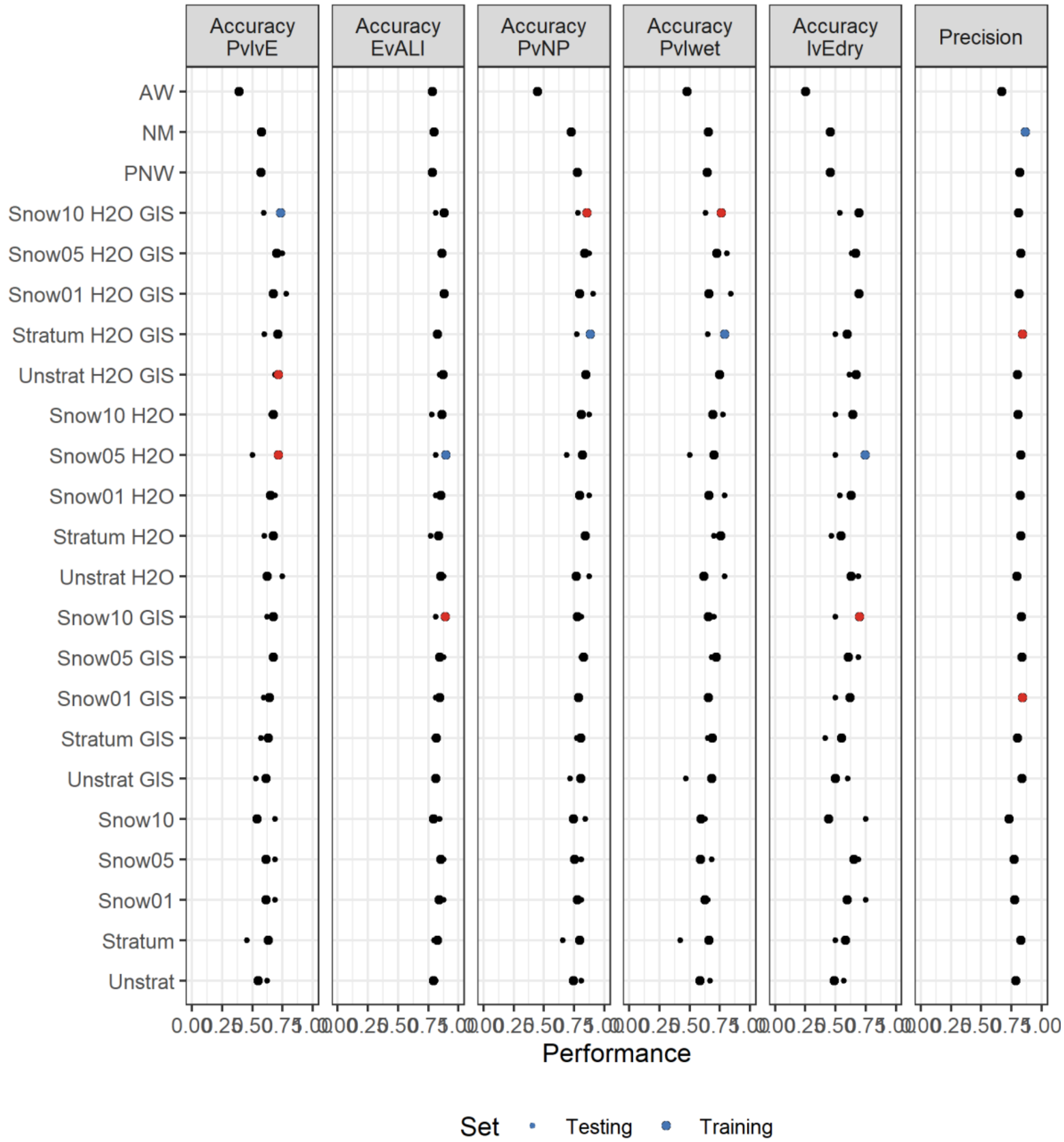

Figure 7. Performance of the 20 model sets evaluated. Blue dots indicate the highest-performing model sets and red dots indicate the next-best performing model sets. PvIvE: Percent of reaches classified correctly as perennial, intermittent, or ephemeral. EvALI: Percent of reaches classified correctly as ephemeral or at least intermittent. PvNP: Percent of reaches classified correctly as perennial or non-perennial. PvIwet: Percent of flowing reaches classified correctly as perennial or intermittent. IvEdry: Percent of dry reaches correctly classified as intermittent or ephemeral. Unstrat: Unstratified models. Stratum: Models stratified by subregion. Snow10: Models stratified by snow persistence. Model sets are described in Table 5. AW: Results for the beta SDAM AW (Mazor et al. 2021a). NM: Results for the SDAM NM. PNW: Results for the SDAM PNW.

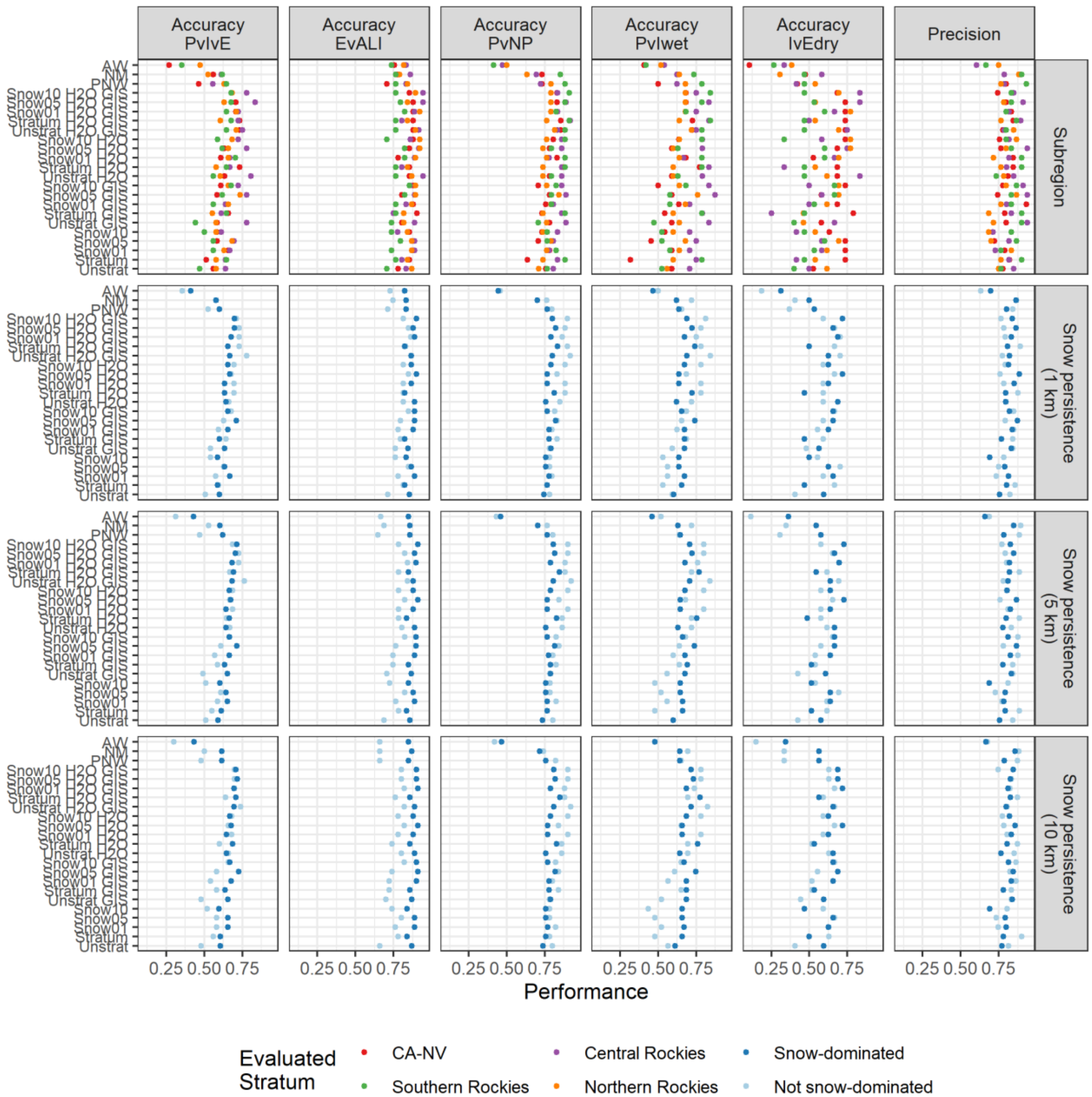

Figure 8. Performance of the 20 model sets evaluated within strata defined by sub-region or snow influence. The y-axis labels on the left indicate the stratifications used to develop the models (if any) and the panel labels on the right indicate the stratifications used to assess performance. PvIvE: Percent of reaches classified correctly as perennial, intermittent, or ephemeral. EvALI: Percent of reaches classified correctly as ephemeral or at least intermittent. PvNP: Percent of reaches classified correctly as perennial or non-perennial. PvIwet: Percent of flowing reaches classified correctly as perennial or intermittent. IvEdry: Percent of dry reaches correctly classified as intermittent or ephemeral. Model sets are described in Table 5. AW: Results for the beta SDAM AW (Mazor et al. 2021a). NM: Results for the SDAM NM. PNW: Results for the SDAM PNW.

Simplification of the selected model set

Upon selection of the final model set (i.e., models that included geospatial metrics and were stratified by snow influence calculated within a 10-km radius), we attempted to simplify the selected model set in three steps to make the SDAM easier to implement in the field while improving (or at least not sacrificing) performance. Simplification occurred in three steps:

1. Refinement of metrics
2. Increased confidence required for classifications
3. Addition of single indicators of at least intermittent flow

Refinement of metrics

The metric selection process described above identified an optimal set of metrics to use in the SDAM, but it did so without considering difficulties in measuring each metric or effort required to measure all of the metrics. For example, rfe may have selected a metric based on the total number of aquatic invertebrates, even if there was little new information provided once 20 individuals were recorded. That is, SDAM users might be able to cease counting aquatic invertebrates once 20 individuals were recorded. Simplifying metrics was intended to reduce the burden on SDAM users and facilitate method use (e.g., avoid reliance on access to statistical software). Some metrics were eliminated because they were closely related to another metric in the selected model set (i.e., they described similar stream characteristics, such as mayfly abundance and EPT abundance). 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 random forest partial dependence curves (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 snow-influenced and non-snow influenced models were refined in parallel steps. At each step, metrics were either eliminated, classified into categorical bins, or otherwise modified. The impact on performance was assessed and the highest performing modification was selected for further refinement. Performance was assessed in terms of three accuracy measures: PvIvE (i.e., proportion of reaches classified corrected as perennial, intermittent, or ephemeral), EvALI (% of reaches classified correctly as ephemeral or at least intermittent), and Cohen's Kappa. The metric refinement steps are described below. Asterisks (*) indicate the selected refinement at each step; if no asterisk is shown, none of the refinements considered at that step were selected and the selected option from the previous step was used for further analysis.

Snow-influenced model:

1. Select two aquatic invertebrate metrics:
    a. Total abundance and richness
    b. Total abundance and perennial indicator abundance*
    c. Total abundance and richness of perennial indicator taxa
    d. Total abundance and EPT abundance
    e. Total abundance and richness of EPT taxa
    f. Total abundance and GOLD abundance
    g. Total abundance and richness of GOLD taxa
2. Add a third aquatic invertebrate metric
    a. Richness of EPT taxa

    b. Richness of perennial indicator taxa *
    c. Total richness
    d. Richness of GOLD taxa
3. Bin richness of perennial indicator taxa metric
    a. Two categories* (0 to 3, ≥4)
    b. Three categories (0, 1 to 3, ≥4)
4. Bin total and perennial indicator abundance
    a. Three categories for total abundance (0, 1 to 19, 20+) and perennial indicator abundance (0, 1 to 5, ≥6)*
5. Bin mean bankfull width
    a. Three categories (<2, 2 to 6, ≥6)*
6. Bin streambed algal cover
    a. Two categories (<10%, ≥10%)*
7. Bin or drop geospatial metrics (NONE SELECTED)
    a. Bin October precipitation at quartiles
    b. Bin October precipitation at quintiles
    c. Drop October precipitation

Refinements to the snow-influenced model improved model performance at most steps (Figure 9). These refinements included eliminating several variables and binning those that remained into two or three categories. Unfortunately, no satisfactory way to bin the single geospatial metric in this model (October precipitation) was identified, so it was retained as a continuous variable for the beta SDAM WM.

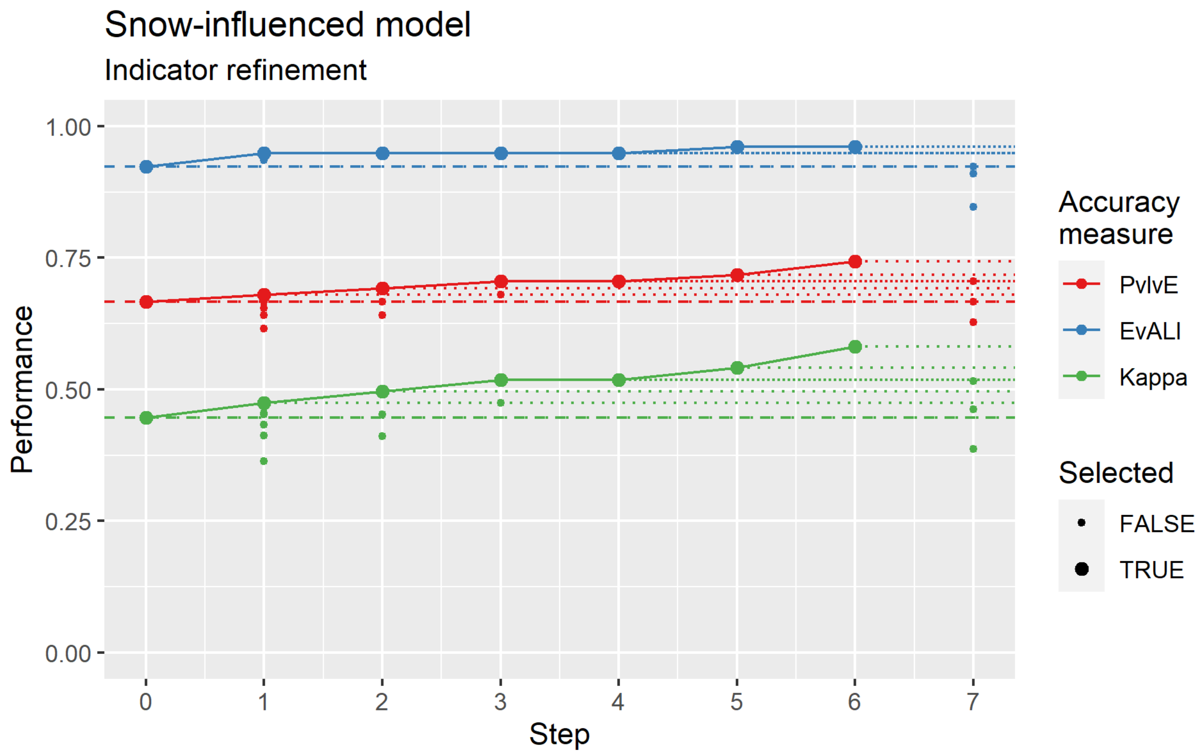

Figure 9. Impact of indicator refinement on the accuracy of the snow-influenced model. Solid lines show the performance of the best model from each step. Dotted lines show the performance of model selected at each step. Dashed lines show performance of the original model.

Non-snow influenced model

1. Select 2 aquatic invertebrate metrics
    a. Total abundance and richness
    b. Total abundance and abundance of perennial indicators
    c. Total abundance and richness of perennial indicator taxa
    d. Total abundance and EPT abundance
    e. Total abundance and richness of EPT taxa
    f. Total abundance and mayfly abundance
    g. Total abundance and GOLD abundance
    h. Total abundance and richness of GOLD taxa
    i. Abundance and richness of EPT taxa
    j. Abundance and richness of perennial indicator taxa
    k. Mayfly abundance and total richness
    l. Mayfly abundance and richness of perennial indicator taxa*
2. Add a third aquatic invertebrate metric (NONE SELECTED)
    a. Total abundance
3. Remove an additional metric (NONE SELECTED)
    a. Sinuosity
    b. Mean bankfull width

c. Fish abundance
4. Bin mayfly abundance
    a. Five categories (0, 1 to 5, 6 to 10, 11 to 15, ≥16)*
5. Bin richness of perennial indicator taxa
    a. Four categories (0, 1, 2, ≥3)*
6. Bin mean bankfull width (NONE SELECTED)
    a. Three categories (<2, 2 to 6, ≥6)
    b. Bin at quartiles
7. Bin geospatial metrics (NONE SELECTED)
    a. Bin May precipitation at three categories (<45, 45 to 50, 50+)
    b. Bin May precipitation at quartiles
    c. Bin maximum temperature at quartiles
    d. Bin maximum temperature in two categories (<18, ≥18)
    e. Bin maximum temperature and May precipitation based on quartiles

Refinements to the non-snow influenced model rarely improved model performance and most refinements were rejected (Figure 10). The only refinement to substantially improve performance was the binning of the mayfly abundance metric (step 4). Thus, the non-snow influenced model retained more metrics in continuous forms than the snow-influenced model.

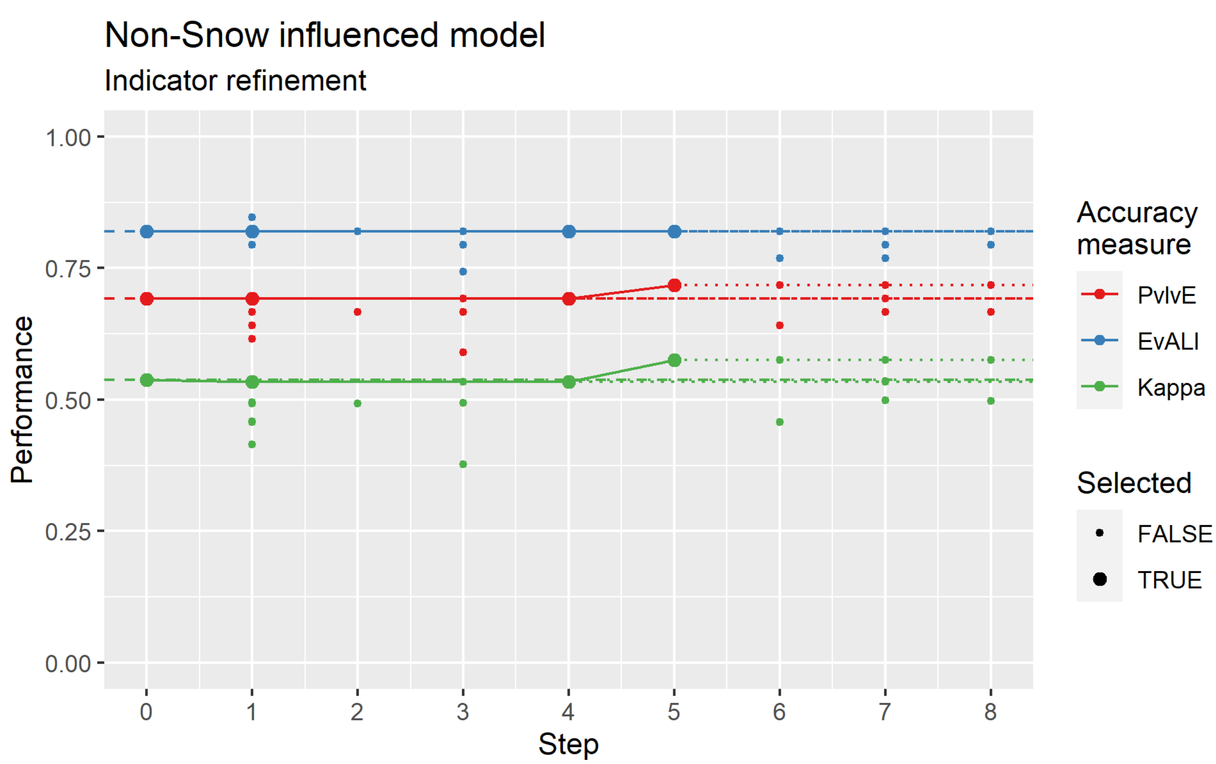

Figure 10. Impact of indicator refinement on the accuracy of the non-snow influenced model. Solid lines show the performance of the best model from each step. Dotted lines show the performance of model selected at each step. Dashed lines show performance of the original model.

Increased confidence required for classifications

Random forest models, when used in classification mode, traditionally make assignments based on the class that receives the highest number of votes by each "tree" in the forest. Thus, in a 3-way decision, the class with the most votes could receive much less than a majority of all votes—as low as 34%. The RSC believed such low-confidence classifications may not provide sufficient defensibility for some management decisions, instead the RSC recommended exploring approaches to distinguish between high- and low-confidence classifications.

Based on this input from the RSC, we explored increasing the minimum number of votes required to make a confident classification from 30% to 100% by increments of 1%. When the final model was applied to a novel test reach and a single class received a sufficient percent of votes, then the reach was classified accordingly. If none met the minimum, but the combined percent of votes for intermittent and perennial classes exceeded the minimum, then the reach was classified as *at least intermittent*. In all other cases, the reach was classified as *need more information*. This decision framework reflects the opinion of the RSC that distinguishing between *ephemeral* and *at least intermittent* reaches is a high priority use of the SDAM, more so than distinguishing between *perennial* and *nonperennial* (ephemeral and intermittent) reaches. The percent of reaches under each of the five possible classifications with increasing minimum vote agreement thresholds was calculated. The snow-influenced and non-snow influenced models were analyzed together to evaluate the overall impact of this modification to the entire WM.

At a minimum required proportion of votes of 0.5, only 5% of reaches were classified as *at least intermittent* and none were classified *need more information* (Figure 11). Classifications of *at least intermittent* first appear with a minimum proportion of 0.38 (0.45 in the testing data set), whereas classifications of *need more information* appear at 0.51 (in both the training and testing data sets). Although they cannot be ruled out, it appears unlikely that the beta SDAM WM will result in classifications of *need more information.* Based on these results, the RSC recommended a minimum proportion threshold of 0.5 for flow classification.

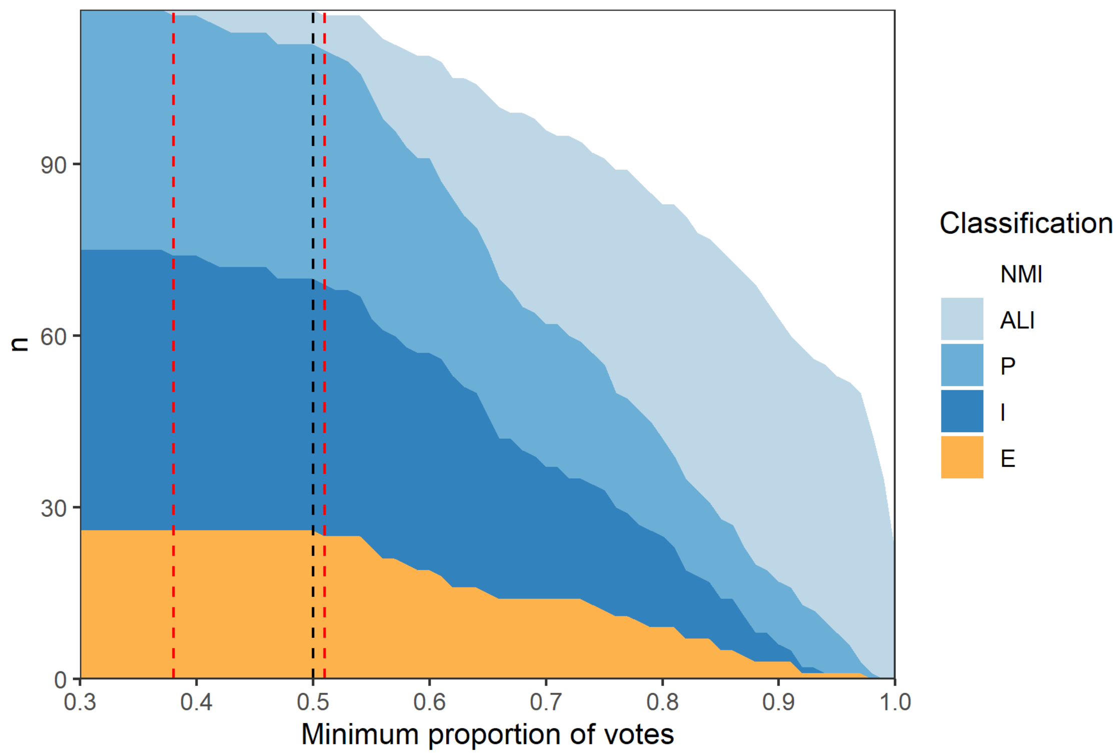

Figure 11. Influence of the minimum proportion of votes required to make a classification on n (the number of reaches in each class). NMI: Need more information. ALI: At least intermittent. P: Perennial. I: Intermittent. E: Ephemeral. The vertical black line represents a minimum proportion of required votes of 0.5, reflecting the final recommendation of the RSC. The two red lines represent the proportion of votes that first result in classification of ALI (the lower line) or NMI (the upper line). Only results from the training data set are shown.

Addition of single indicators of at least intermittent flow

Single indicators can supersede model classifications of *ephemeral* 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 wildlife and rapidity of determining a flow classification, which is why they are used in most other SDAMs (e.g., NMED 2011, Nadeau et al. 2015, Dorney and Russell 2018, Mazor et al. 2021a). The following potential single indicators, based on recommendations from the RSC were evaluated:

- Presence of aquatic invertebrates
- Presence of EPT individuals, or at least 5 EPT individuals
- Presence of hydrophytes, or at least 2 or 3 hydrophytic plant species
- Algal cover ≥ 10%
- Presence of fish

The number of instances where inclusion of the single indicator would correct a misclassification (i.e., the reach was truly intermittent or perennial) and the number of times it would introduce a misclassification (i.e., the reach was truly ephemeral) were quantified.

Several single indicators had minimal impact on performance or introduced more errors than they corrected (Figure 12). Based on these results, the RSC recommended using only the presence of fish (apart from mosquitofish) as single indicators in the beta SDAM WM.

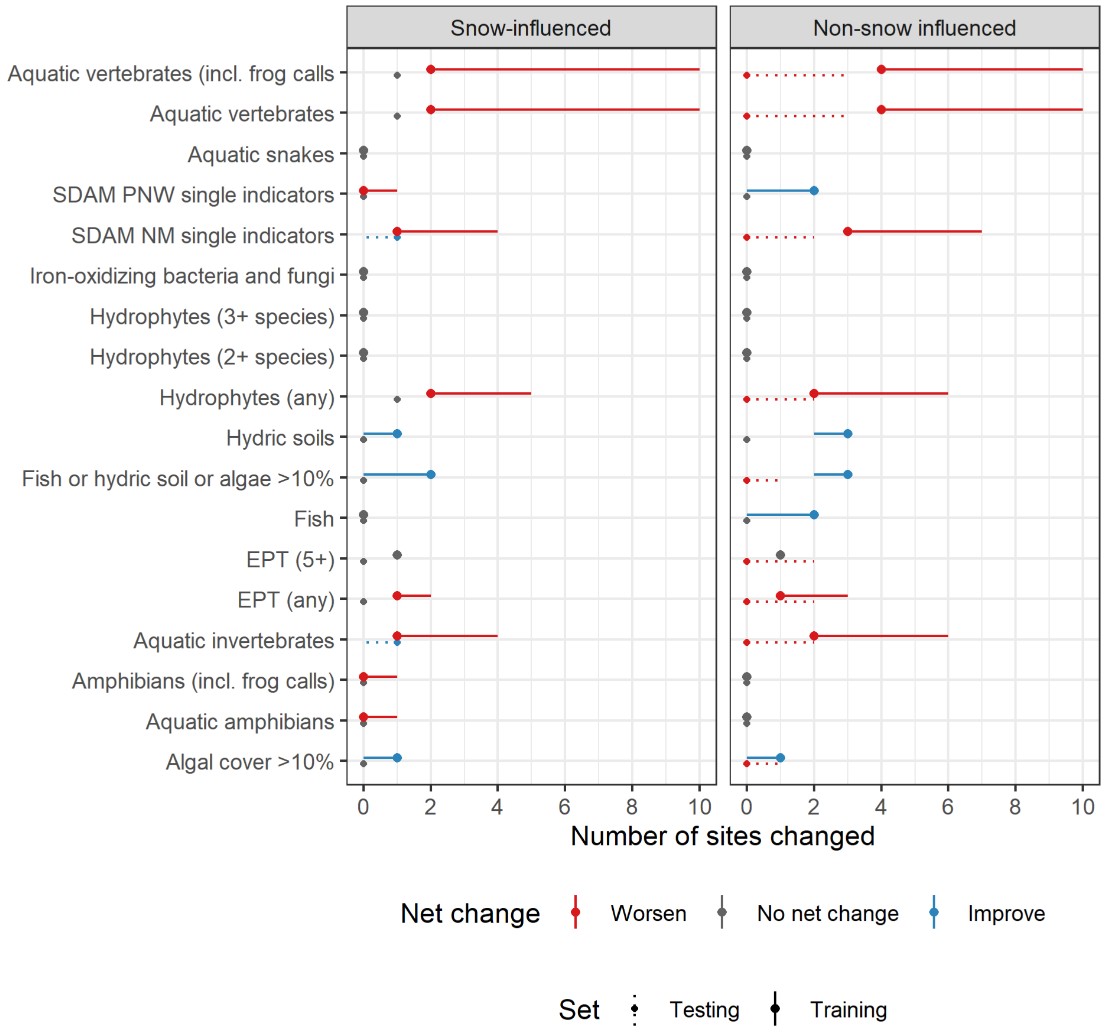

Figure 12. Influence of single indicators on performance of snow-influenced and non-snow influenced models

Performance of the beta SDAM WM

Performance of the final, simplified model for the beta SDAM WM is summarized in Table 7. The overall accuracy was 74% in the training dataset (and 53% in the testing dataset), but this accuracy increased to 93% in the training dataset (and 88% in the testing data set) when only *ephemera*l versus *at least intermittent* classifications were considered (i.e., both blue and green cells in Table 7 were treated as correct). Among 42 reaches marked as disturbed by human activity, accuracy among all classes was 79% and 95% when only *ephemeral* versus *at least intermittent* classifications were considered.

Table 7. Classifications of the final version of the beta SDAM WM on training and testing datasets. Blue cells indicate correct classifications of perennial, intermittent, at least intermittent, and ephemeral reaches, whereas green cells indicate correct classifications as ephemeral versus at least intermittent.

	True streamflow duration class							
			Intermittent					
Beta SDAM WM Classification	Ephemeral		Dry		Flowing		Perennial	
	Train	Test	Train	Test	Train	Test	Train	Test
Ephemeral	20	4	4	1	0	0	0	0
Intermittent	3	3	17	3	16	5	8	7
At least intermittent	1	0	2	0	2	1	3	0
Perennial	0	0	0	1	11	3	30	4

Data and code availability

All data used to develop the method and R code used in analysis are available here.

Next steps

Continued data collection within the WM is underway and will provide greater representation of the diversity of stream conditions found within the region. Data from this effort will be used to develop a final method (expected after 2023) to replace the beta method.

Acknowledgements

The development of this method and supporting materials was guided by a RSC consisting of representatives of federal regulatory agencies in the Western U.S.: James T. Robb (U.S. Army Corps of Engineers [USACE]—South Pacific Division, Sacramento District), Robert Leidy (U.S. Environmental Protection Agency [USEPA]—Region 9), Aaron Allen (USACE—South Pacific Division, Los Angeles District), Gabrielle C. L. David (USACE—Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory), Loribeth Tanner (USEPA—Region 6), Rachel Harrington (USEPA – Region 8), Joe Morgan (USEPA—Region 9), Matt Wilson (USACE—Headquarters), Tunis McElwain (USACE—Headquarters), Silvia Gazzera (USACE – Headquarters), Kevin Little, (USACE – Northwestern Division, Omaha District), Jess Jordan (USACE – Northwestern Division, Seattle District), and Rose Kwok (USEPA—Headquarters).

We thank Abel Santana, Robert Butler, Duy Nguyen, Kristine Gesulga, and Anne Holt for assistance with data management and Jeff Brown, Liesl Tiefenthaler, Mason London, John Olson, Matthew Robinson, Emma Haines, Jess Turner, Katharina Zimmerman, Kelsey Trammel, Marcus Beck, Savannah Peña, Abigail Rivera, and Andrew Caudillo for assistance with data collection. Rob Coulombe provided training.

Numerous researchers and land managers with local expertise assisted with the selection of study reaches to calibrate the method: Patricia Spindler, Eric Stein, Andrew C. Rehn, Peter R. Ode, Nathan Mack, Shawn McBride, Stephanie Kampf, Lindsey Reynolds, Kris Barrios, Marcia Radke, Keith Bouma-Gregson, Kira Puntenney-Desmond, Andy Brummond, Don Lee, Ed Schenk, Eric Hargett, Gabe Rossi, Mark Ockey, Sean Tevlin, Sean Lovill, Josh Smith, and Michael Bogan. We thank the California Department of Fish and Wildlife's Aquatic Bioassessment Lab and Daniel Pickard for use of imagery from the macroinvertebrate digital reference collection.

Cited literature

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.

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. Elsevier, San Diego, CA.

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.

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 online: https://www.hydroshare.org/resource/1c62269aa802467688d25540caf2467e/

Hart, E., and K. Bell. 2015. Prism: Access Data From The Oregon State Prism Climate Project.

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.

Kuhn, M. 2020. caret: Classification and Regression Training.

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.

Mazor, R. D., and K. S. McCune. 2021. Review of flow duration methods and indicators of flow duration in the scientific literature: Western Mountains. Pages 55. Southern California Coastal Water Research Project, Costa Mesa, CA.

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. Pages 83. Document No. EPA-800-K-21001, U.S. Environmental Protection Agency, Washington, D.C. Available online: 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, 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 Duratoin Assessment Method for the Western Mountains of the United States. Version 1.0. Pages 116. Document No. EPA 840-B-21008, U.S. Environmental Protection Agency, Washington, D.C. Available online: https://www.epa.gov/system/files/documents/2021-12/beta-sdam-for-the-wm-user-manual.pdf

McCune, K., and R. D. Mazor. 2019. Review of flow duration methods and indicators of flow duration in the scientific literature: Arid Southwest. Pages 90; Available online: https://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/1063_FlowMethodsReview.pdf. Southern California Coastal Water Research Project, Costa Mesa, CA.

Nadeau, T.-L. 2015. Streamflow Duration Assessment Method for the Pacific Northwest. Pages 36. Document No. EPA 910-K-14-001, U.S. Environmental Protection Agency, Region 10, Seattle, WA. Available online: 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.

New Mexico Environment Department (NMED). 2011. Hydrology protocol for the determination of uses supported by ephemeral, intermittent, and perennial waters. Page 35. Surface Water Quality Bureau, New Mexico Environment Department, Albuquerque, NM.

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.

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. EPA Report D-WQD-ECB-024-SOP-02. Environmental Protection Agency, Washington, D.C.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878–891.

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). Page 153. U.S. Army Engineer Research and Development Center, Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Links

Beta Streamflow Duration Assessment Method for the Western Mountains user manual: https://www.epa.gov/streamflow-duration-assessment/beta-streamflow-duration-assessment-method-western-mountains

Reginal Streamflow Duration Assessment Methods website: https://www.epa.gov/streamflow-duration-assessment

Web application for the beta SDAM for WM: https://sccwrp.shinyapps.io/beta_sdam_wm/

Western Mountain beta SDAM data and R code: https://doi.org/10.23719/1526066