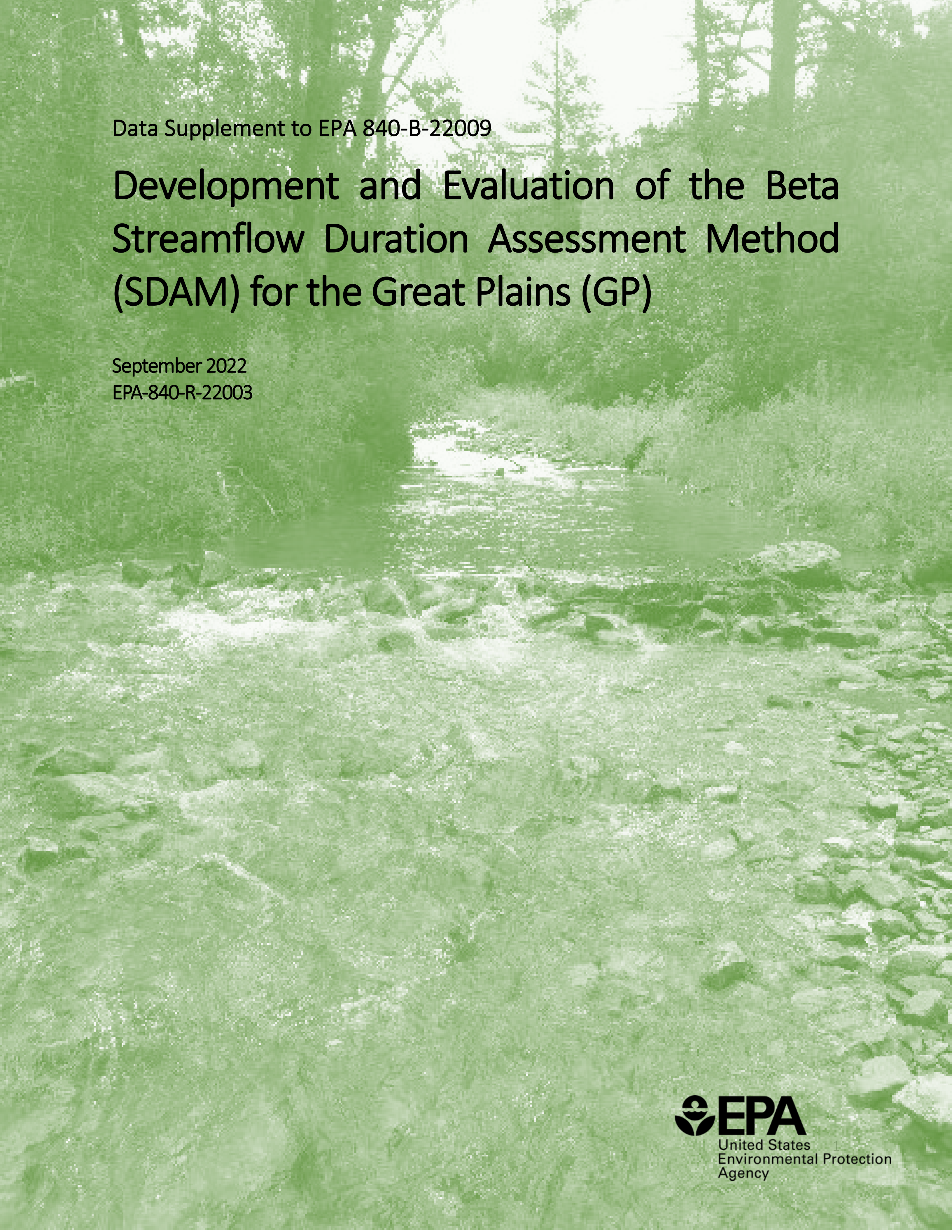

Data Supplement to EPA 840-B-22009

Development and Evaluation of the Beta Streamflow Duration Assessment Method (SDAM) for the Great Plains (GP)

September 2022
EPA-840-R-22003

United States
Environmental Protection
Agency

Development and Evaluation of the Beta Streamflow Duration Assessment Method for the Great Plains

Data analysis supplement

Michele Eddy
RTI International
Research Triangle Park, NC 27709

Shannon Gross
RTI International
Fort Collins, CO 80528

Ken Fritz
Office of Research and Development
Cincinnati, OH 45268

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

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

Rachel Fertik Edgerton
Office of Wetlands, Oceans, and Watersheds
Washington, DC 20004

Julie Kelso, ORISE Fellow (former)
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. This report fulfills EPA QA requirements. The research for the data was conducted under the Office of Water approved Quality Assurance Project Plan "Streamflow Duration Assessment Method (SDAM) development in the Great Plains (GP) and Western Mountains (WM)" which was given an ORD ID of J-WECD-0033408-QP-1-0. 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 contracts EP-C-17-001 and 68HERC21D0008 for data management and analysis, respectively, 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: Eddy M., Gross, S., Fritz, K.M., Nadeau, T.-L., Topping, B., Fertik Edgerton, R., and Kelso, J. 2022. Development and Evaluation of the Beta Streamflow Duration Assessment Method for the Great Plains. Document No. EPA-840-R-22003.

Introduction

Streamflow duration assessment methods (SDAMs) are rapid, field-based methods to determine flow duration class at the reach scale. The development of a beta SDAM for the Northern and Southern Great Plains regions (hereafter referred to as the GP) followed the conceptual framework and process steps presented by Fritz and others (2020) to integrate the three key components of an SDAM development study: hydrological data, indicators, and study reaches.

This supplemental document describes the data collection, data analysis, and evaluation steps that resulted in the beta SDAM for the GP. This document is available to inform public review and comment on the beta method, as well as serving as a companion to the beta SDAM GP for those that are interested in more background on the development of the method and the underlying data. For a complete description of the beta SDAM GP protocol, please see the User Manual (https://www.epa.gov/system/files/documents/2022-09/beta-sdam-for-the-gp-user-manual.pdf). The data used to develop the beta SDAM GP can be found here: (https://doi.org/10.23719/1527943). 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: https://www.epa.gov/streamflow-duration-assessment .

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 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 Great Plains

The beta SDAM GP uses a small number of indicators to predict the streamflow duration class of stream reaches. All indicators are measured during a single field visit. The beta SDAM GP results in one of four possible classifications: *ephemeral*, *intermittent*, *perennial*, or *at least intermittent*. The latter 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 (Figure 1). 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. This approach was previously used to develop regional SDAMs for the Pacific Northwest (PNW; Nadeau et al. 2015, Nadeau 2015), Arid West (AW; Mazor et al. 2021a, Mazor et al 2021b), and Western Mountains (WM; Mazor et al. 2021c; Mazor et al. 2022).

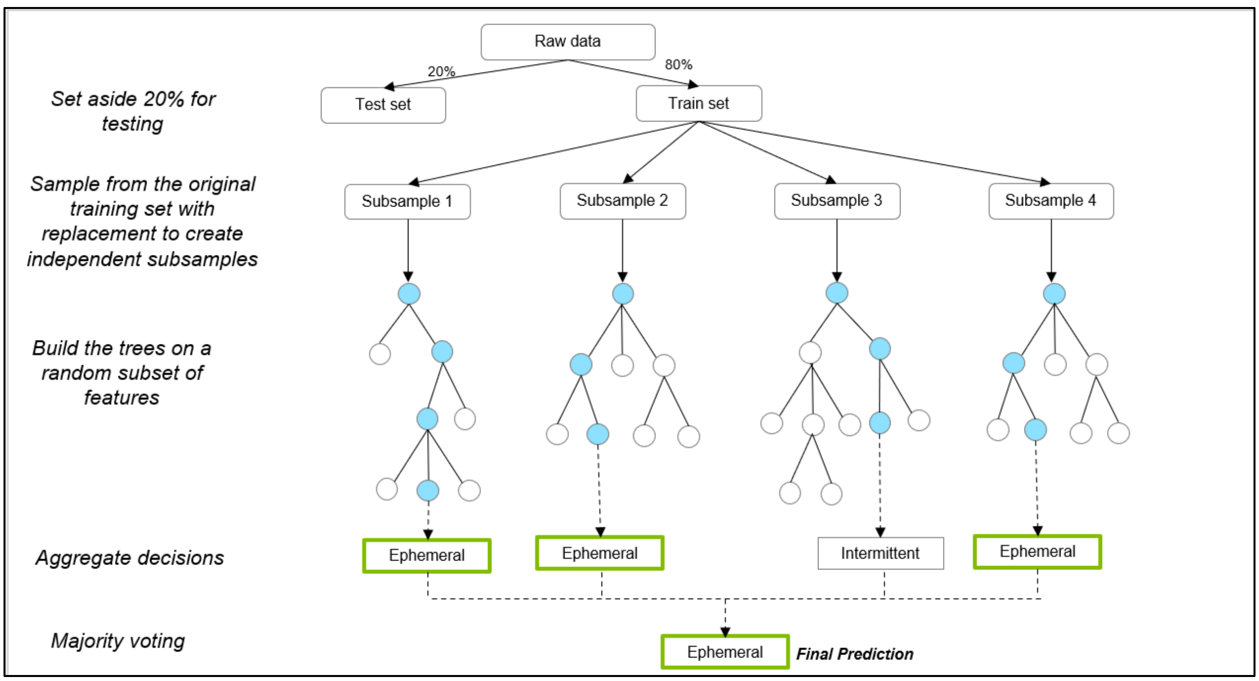

Figure 1. Random forest procedure used to determine a flow classification.

Development of the Beta Great Plains SDAM

The specific data analysis steps described in this document follow the approach used to develop and evaluate the beta SDAM WM (Mazor et al. 2022).

Study Area

The GP spans the central U.S. from Canada to Mexico and encompasses all or portions of 15 states (Figure 2). It includes areas largely dominated by native prairie-type vegetation (tall, short, and mixed grass) that generally receive less than 40 inches of precipitation a 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). The GP regions are divided into Northern and Southern GP regions based on the importance of snowmelt to river discharge; the boundary between the two approximately follows the line south of which mean annual snowfall is less than 0.7 m/y (<2 ft/y; Wohl et al. 2016). 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.

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, occur across the GP (Vengosh et al. 2014, Perkin et al. 2017). Within a portion of the Southern GP region, there is one SDAM currently in use, applicable to New Mexico (New Mexico Environment Department [NMED] 2011).

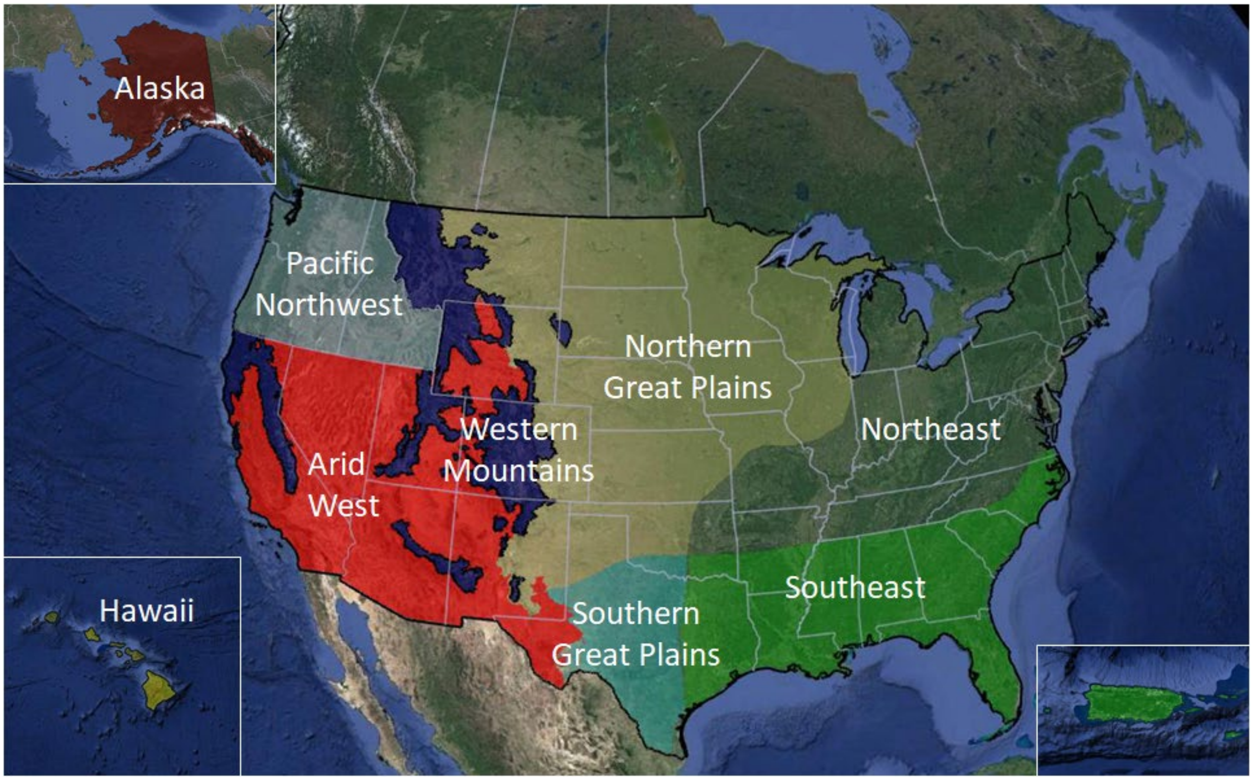

Figure 2. Map of SDAM study regions (based on Wohl et al. 2016). The beta SDAM GP applies to the Northern and Southern Great Plains as shown.

Preparation and Candidate Indicators

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 GP 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, identifying resources such as sources of hydrologic data, and providing input on the model selection.

We identified candidate indicators that were supported by the scientific literature (James et al. 2022) or used in the New Mexico SDAM (herein referred to as NM method; NMED 2011). In addition, we included candidate indicators from the SDAM PNW (Nadeau 2015). Following input from the RSC, these candidate indicators were then screened using the criteria described by Fritz and others (2020), including:

<u>Primary criteria</u>

- *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?

Secondary criteria

- *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: 1) met all the primary criteria; 2) at least one of the secondary criteria; or 3) were included in the NM method (Level 1 only) to facilitate comparison (because not all NM indicators met all primary criteria). Desktop geospatial indicators (derived using a geographic information system and applicable spatial datasets) that characterize mechanisms affecting flow duration and have been explored in other flow duration classification tools (e.g., Eng et al. 2016, Jaeger et al. 2019, Mazor et al. 2021c) were also included in the analysis.

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 (USEPA 2019) and derived from sources resulting from a literature review completed by James et al. (2022) 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
Slope	Valley slope measured with a handheld clinometer	PNW
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

Candidate indicator	Description	Origin
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 3 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
Hydrophytic plant species	Number of OBL or FACW-rated plants (as listed in Lichvar et al. 2016) growing within the channel or one 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

Candidate indicator		Description	Origin
	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			
	Elevation	Elevation above mean sea level	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
	Strata (location)	The four subregions or 'strata' into which the Northern and Southern Great Plains have been subdivided: Northern Prairie, Central Prairie, Upper Midwest, and Southern Plains	OTH
	Baseflow Index (BFI)	The ratio of baseflow to total flow, expressed as a percentage and provided as a 1-kilometer raster grid for the conterminous U.S. (Wolock, 2003)	OTH

Candidate Reach Identification and Data Collection

We had two objectives in selecting candidate reaches for 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 GP to support applicability of the method across anticipated conditions. To support our goal of geographic representativeness, we subdivided the Northern GP into 3 subregions or strata, based on EPA Level II Ecoregion boundaries (Omernik 1995). This resulted in 4 strata: Central Prairie, Northern Prairie, Upper Midwest, and Southern Great Plains. We aimed to select 290 stream-reaches (one assessed location per reach) with equal representation of perennial, intermittent, and ephemeral flow duration classes among and within the four GP strata (Figure 3).

To screen reaches for use in method development, we first compiled a list of 3566 candidate study reaches based on existing hydrologic data records (e.g., U.S. Geological Survey (USGS) stream gages, water presence loggers, wildlife cameras, field photos), published studies, and interviews with local experts familiar with the specific reach's hydrology. Most of these reaches (2945) were derived from the database of stream gages operated by the USGS and 2298 (78%) of them were perennial. (Actual streamflow duration class was determined by applying the flowchart in Figure 4, which was informed by existing definitions (Hedman and Osterkamp 1982, Hewlett 1982).) Consequently, other sources were required to identify candidate ephemeral and intermittent reaches. Another 621 candidate study reaches were identified from published studies or consultation with local experts. Whenever possible, multiple sources of hydrologic information were used to confirm actual streamflow classifications. In the resulting set of candidate reaches, 7.5% were determined to be ephemeral, 26.1% were intermittent, and 66.3% were perennial.

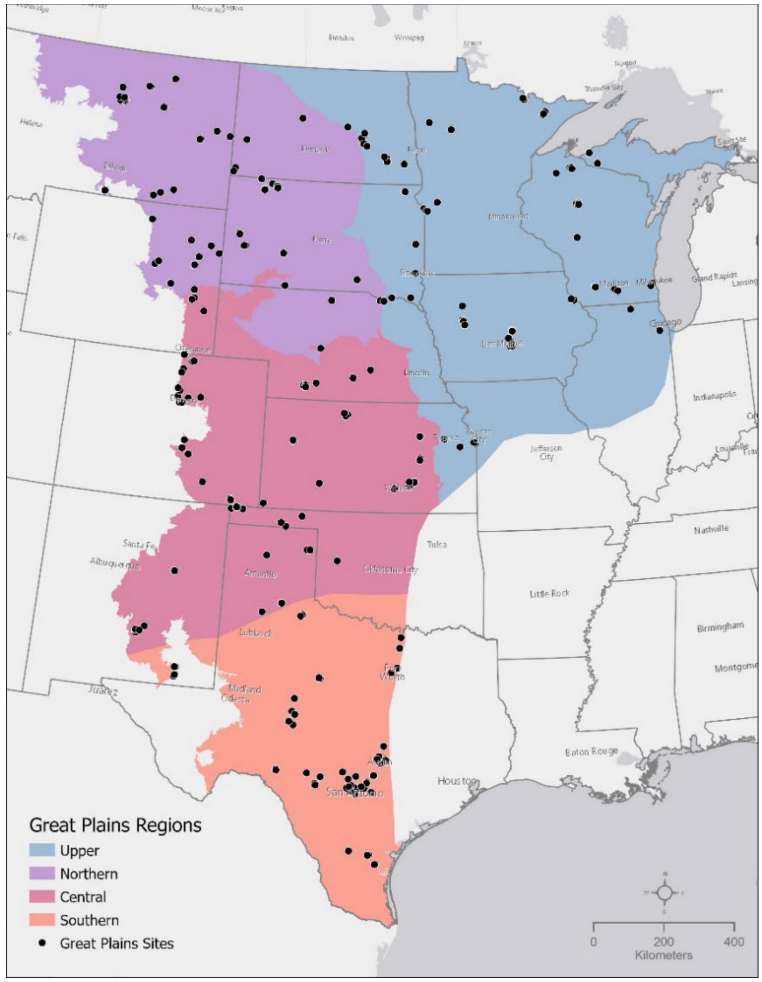

Figure 3: The four GP sub-strata; study reaches shown are those used to calibrate the beta SDAM GP.

Reaches were prioritized for study inclusion based on being accessible (e.g., on public property or with landowner permission), being wadable, and the number and type of data sources available to determine actual streamflow duration classification. Reaches where streamflow duration class 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 strata where an insufficient number of "preferred" and "USGS gage" reaches classified as intermittent or ephemeral could be identified.

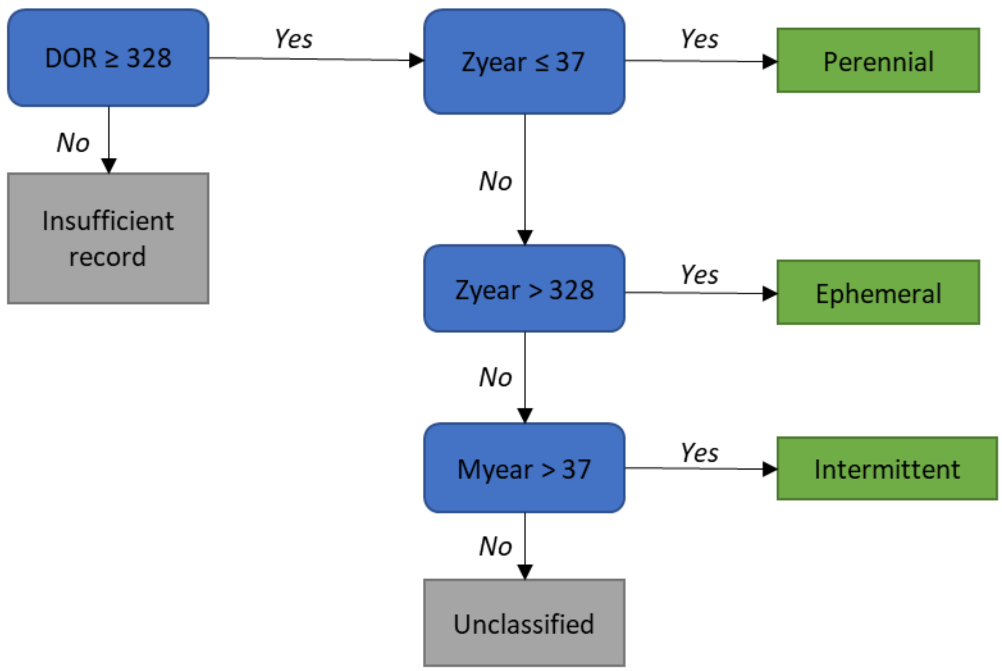

Figure 4. Flowchart used to determine actual streamflow duration class of reaches based on continuous measures of water presence (e.g., USGS streamgages). DOR: days of record. Zyear: Average number of dry days per year. Myear: Average length of longest continuous wet period per year, in days. For USGS gages, at least 20 years of data were analyzed whenever possible (Kelso and Fritz 2021).

Of the 3566 candidate reaches, 293 study reaches were sampled from November 2019 to June 2021. These study reaches were parsed into 'instrumented' and 'single-visit' reaches[1]. Instrumented reaches (183) were visited multiple times (up to four), and each had at least one Stream Temperature, Intermittence, and Conductance (STIC; Chapin et al. 2014) logger deployed, with 10% of instrumented reaches having duplicate data loggers. Instrumented reaches generally had fewer existing lines of evidence to determine actual streamflow duration classification before sampling; therefore, post-sampling reach classifications were reviewed in light of the STIC logger data and hydrology indicator data that were direct measures of water presence collected during each visit. For further details on STIC data loggers and their verification/calibration, deployment, and data retrieval, see Schumacher and Fritz (2019). Single-visit reaches (110) were visited once (with a 10% resample) and did not have loggers deployed. Because actual streamflow duration classification of most single-visit reaches was determined using existing data, these reaches generally had multiple direct flow duration data sources. Ultimately, due to data loss from STIC loggers and other factors, actual streamflow duration class at 42 reaches (35 instrumented and seven single-visit reaches) could not be

[1] These reaches were termed 'baseline' and 'validation', respectively, in prior beta SDAMs but have been renamed for clarity.

determined with confidence and were excluded from analysis used to develop the beta SDAM GP. Of the 251 study reaches used to develop the beta SDAM GP, 71 were ephemeral, 100 were intermittent, and 80 were perennial (Table 2).

Table 2. Distribution of reaches used to develop the beta SDAM GP. Instrumented reaches were visited up to four times and had Stream Temperature, Intermittence, and Conductance loggers installed and single-visit reaches were visited once (rarely, twice) and did not have loggers installed.

	Single-Visit		Instrumented			
Class	Gaged	Preferred	Gaged	Preferred	Acceptable	Total
Ephemeral	*14*	*14*	*6*	*7*	*30*	*71*
-Northern Prairie	3	6	0	0	9	10
-Upper Midwest	0	2	0	1	2	7
-Central Prairie	7	4	1	3	13	6
-Southern Plains	4	2	5	3	6	8
Intermittent	*13*	*26*	*10*	*15*	*36*	*100*
-Northern Prairie	4	3	4	2	6	19
-Upper Midwest	1	13	1	1	23	39
-Central Prairie	2	7	2	8	5	24
-Southern Plains	6	3	3	4	2	18
Perennial	*32*	*4*	*23*	*9*	*12*	*80*
-Northern Prairie	9	0	6	0	1	16
-Upper Midwest	8	1	5	6	9	29
-Central Prairie	8	1	7	1	1	18
-Southern Plains	7	2	5	2	1	17

During each field visit to a study reach the suite of candidate indicators (Table 1) were measured following the development protocol (USEPA 2019). This compilation of indicators from a single field visit constitutes one reach sample (or observation) in terms of the analyses described within this data analysis supplement. Surrounding land use may affect or disturb streamflow duration indicators without substantially shifting flow duration at reaches (e.g., changes in water quality). Up to two predominant land use categories within a 100-m radius of each study reach were noted on each field visit. If “urban” or “agriculture” were the identified land use category the sample was considered disturbed; otherwise, the sample was considered not disturbed for comparisons of beta SDAM GP performance.

Data analysis

Metric calculation

Candidate indicator data were used to create 95 candidate metrics, of which 52 were biological, 11 were geomorphological, ten were hydrologic (eight directly measured water presence, and two were indirect measurements), and 22 were geospatial (Table 3).

Table 3. Candidate metrics evaluated for the development of the beta SDAM GP. Please see Appendix A for full definitions of Candidate metrics. Asterisks (*) indicate hydrologic metrics that directly measure the presence of water. Abbreviations in Candidate metric names include – EPT: Ephemeroptera, Plecoptera, and Trichoptera insect orders. GOLD: Gastropoda, Oligochaeta, and Diptera invertebrate groups. OCH: Odonata, Coleoptera, and Heteroptera insect orders. For Type the following categories apply – Ord: Ordinal metrics. Cat: Categorical metrics. Bin: Binary metrics. Con: Continuous metrics. 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 in Table 5. NA = not applicable

Candidate metrics	Group	Type	PctDom	Min	Max	Range	PvIvE	EvNE	PvNP	PvIwet	EvIdry	rf_MDA	Screened
ai_present	Bio	Bin	64%	0	1	1	267.06	21.08	19.12	4.24	3.99	0.01	Pass
Algae_score	Bio	Ord	48%	0	3	3	126.58	15.63	12.90	4.95	3.27	0.01	Pass
algdead_cover_score	Bio	Ord	89%	0	3	3	11.17	4.83	3.49	2.35	1.29	0.00	Pass
algdead_noupstream_cover_score	Bio	Ord	89%	0	3	3	11.52	4.73	3.58	2.56	1.29	0.00	Pass
alglive_cover_score	Bio	Ord	52%	0	4	4	102.24	14.77	11.35	4.08	3.14	0.01	Pass
alglivedead_cover_score	Bio	Ord	50%	0	4	4	106.29	15.09	11.50	4.21	3.58	0.01	Pass
amphib_score	Bio	Bin	83%	0	1	1	17.82	8.11	2.43	0.62	2.11	0.00	Pass
BMI_score	Bio	Ord	43%	0	3	3	292.33	22.08	21.59	7.36	3.50	0.01	Pass
DifferencesInVegetation_score	Bio	Ord	27%	0	3	3	86.72	12.36	9.71	3.61	4.34	0.00	Pass
EPT_abundance	Bio	Con	57%	0	45	45	117.44	13.85	11.26	7.84	2.15	0.01	Pass
EPT_relabd	Bio	Con	57%	0	1	1	125.81	15.40	12.33	7.38	1.73	0.01	Pass
EPT_reltaxa	Bio	Con	57%	0	1	1	141.40	16.00	13.23	7.61	1.95	0.01	Pass
EPT_taxa	Bio	Con	57%	0	7	7	172.21	15.77	14.07	9.45	2.01	0.01	Pass
Fish_score	Bio	Ord	65%	0	3	3	116.37	14.52	12.18	6.49	0.03	0.00	Pass
fishabund_score2	Bio	Ord	67%	0	3	3	116.06	14.39	12.12	6.43	1.87	0.00	Pass
frogvoc_score	Bio	Bin	85%	0	1	1	10.63	5.90	1.80	1.01	2.02	0.00	Pass
GOLD_abundance	Bio	Con	51%	0	29	29	37.81	10.88	5.32	0.85	2.48	0.00	Pass
GOLD_relabd	Bio	Con	51%	0	1	1	30.34	9.27	2.77	2.63	1.29	0.00	Pass
GOLD_reltaxa	Bio	Con	51%	0	1	1	39.55	10.15	4.37	1.90	1.57	0.00	Pass
GOLD_taxa	Bio	Con	51%	0	5	5	75.47	13.95	8.56	2.18	2.54	0.00	Pass
GOLDOCH_relabd	Bio	Con	42%	0	1	1	54.00	11.44	3.20	2.99	3.64	0.01	Pass
GOLDOCH_reltaxa	Bio	Con	42%	0	1	1	76.36	13.53	5.28	2.25	3.98	0.01	Pass
hydrophytes_present	Bio	Ord	22%	0	8	8	116.68	15.60	10.60	3.77	5.15	0.00	Pass
hydrophytes_present_any	Bio	Bin	78%	0	1	1	116.21	11.87	10.45	2.02	5.71	0.00	Pass
hydrophytes_present_noflag	Bio	Ord	22%	0	8	8	117.79	15.88	10.53	3.59	5.26	0.01	Pass

Candidate metrics	Group	Type	PctDom	Min	Max	Range	PvIvE	EvNE	PvNP	PvIwet	EvIdry	rf_MDA	Screened
iofb_score	Bio	Bin	89%	0	1.5	1.5	4.75	3.65	0.81	1.52	0.91	0.00	Pass
liverwort_cover_score	Bio	Ord	97%	0	3	3	5.35	2.40	2.44	2.24	0.55	0.00	Fail
mayfly_abundance	Bio	Con	64%	0	30	30	83.84	12.38	9.67	6.48	1.98	0.00	Pass
mayfly_gt6	Bio	Bin	81%	0	1	1	63.01	11.22	8.54	5.70	2.08	0.00	Pass
moss_cover_score	Bio	Ord	92%	0	3	3	18.09	6.27	4.47	3.18	1.27	0.00	Pass
Noninsect_abundance	Bio	Con	61%	0	30	30	31.19	10.38	4.58	0.64	2.58	0.00	Pass
Noninsect_relabund	Bio	Con	61%	0	1	1	23.53	8.35	2.48	1.79	1.87	0.00	Pass
Noninsect_reltaxa	Bio	Con	61%	0	1	1	29.39	9.00	3.36	1.45	1.97	0.00	Pass
Noninsect_taxa	Bio	Con	61%	0	4	4	50.62	12.11	6.61	1.64	2.57	0.00	Pass
OCH_abundance	Bio	Con	58%	0	29	29	17.45	6.94	1.97	0.47	3.97	0.00	Pass
OCH_relabd	Bio	Con	58%	0	1	1	18.54	6.81	1.46	0.99	3.84	0.00	Pass
OCH_reltaxa	Bio	Con	58%	0	1	1	31.55	9.29	2.78	0.91	4.05	0.00	Pass
OCH_taxa	Bio	Con	58%	0	6	6	39.62	11.03	4.90	0.74	4.14	0.00	Pass
PctShading	Bio	Con	32%	0	1	1	5.32	2.42	1.04	2.73	1.12	0.00	Pass
peren_present	Bio	Bin	72%	0	1	1	133.73	12.37	13.62	8.83	0.49	0.00	Pass
perennial_abundance	Bio	Con	72%	0	32	32	53.94	8.29	7.82	5.86	0.09	0.00	Pass
perennial_live_abundance	Bio	Con	72%	0	32	32	52.84	8.23	7.75	5.79	0.09	0.00	Pass
perennial_taxa	Bio	Con	72%	0	5	5	98.70	10.06	10.80	8.02	0.27	0.00	Pass
Richness	Bio	Con	36%	0	18	18	189.94	19.03	15.32	7.04	3.81	0.02	Pass
ripariancorr_score	Bio	Bin	70%	0	1	1	39.98	8.12	4.75	0.53	3.06	0.00	Pass
snake_score	Bio	Bin	98%	0	1	1	3.26	2.30	1.96	1.88	1.09	0.00	Fail
TotalAbundance	Bio	Con	36%	0	86	86	121.21	16.60	11.51	5.16	3.89	0.02	Pass
turt_score	Bio	Bin	95%	0	1	1	7.06	4.30	2.71	1.34	0.55	0.00	Fail
UplandRootedPlants_score	Bio	Ord	57%	0	3	3	180.91	15.56	16.32	4.54	3.13	0.01	Pass
vert_score	Bio	Bin	71%	0	1	1	32.79	10.33	3.80	0.58	3.03	0.00	Pass
vert_sumscore	Bio	Ord	79%	0	3	3	22.05	8.68	3.68	0.69	2.27	0.00	Pass
vertvoc_sumscore	Bio	Bin	71%	0	4	4	27.02	9.57	3.67	0.03	2.74	0.00	Pass
BankWidthMean	Geomorph	Con	2%	0.4	68.3	67.9	24.76	4.16	6.63	4.36	0.79	0.02	Pass
ChannelDimensions_score	Geomorph	Ord	57%	0	3	3	3.25	1.40	1.37	1.13	2.40	0.00	Pass
erosion_score	Geomorph	Bin	89%	0	1	1	0.18	0.53	0.01	0.56	0.19	0.00	Fail
floodplain_score	Geomorph	Bin	66%	0	1	1	2.37	1.39	0.87	1.41	0.79	0.00	Pass
RifflePoolSeq_score	Geomorph	Ord	30%	0	3	3	40.49	7.71	7.69	3.27	0.69	0.00	Pass
SedimentOnPlantsDebris_score	Geomorph	Ord	29%	0	1.5	1.5	30.88	7.25	5.91	0.80	0.41	0.00	Pass

Candidate metrics	Group	Type	PctDom	Min	Max	Range	PvIvE	EvNE	PvNP	PvIwet	EvIdry	rf_MDA	Screened
Sinuosity_score	Geomorph	Ord	49%	0	3	3	15.96	6.01	1.40	1.05	4.16	0.00	Pass
Slope	Geomorph	Ord	40%	0	20	20	4.57	1.77	3.70	1.96	0.32	0.00	Pass
slope_gt10.5	Geomorph	Bin	98%	0	1	1	3.17	0.51	3.03	2.26	0.78	0.00	Fail
slope_gt16	Geomorph	Bin	100%	0	1	1	1.20	1.00	1.00	0.00	1.00	0.00	Fail
SubstrateSorting_score	Geomorph	Ord	33%	0	3	3	77.71	8.20	13.21	6.77	1.35	0.01	Pass
BFI	GIS	Con	5%	7	76	69	38.44	4.19	8.47	6.30	0.21	0.01	Pass
Elev_m	GIS	Con	3%	13	2643	2630	3.77	2.41	0.42	2.27	0.77	0.01	Pass
MeanSnowPersistence_01	GIS	Con	1%	0.000	52.789	52.789	34.02	8.76	5.23	2.18	4.83	0.01	Pass
MeanSnowPersistence_05	GIS	Con	1%	0.096	50.826	50.730	33.88	8.56	5.33	2.20	4.84	0.01	Pass
MeanSnowPersistence_10	GIS	Con	1%	0.074	51.522	51.448	34.02	8.57	5.34	2.23	4.71	0.01	Pass
ppt	GIS	Con	1%	287.21	1056.46	769.25	12.69	5.03	2.98	0.12	1.57	0.01	Pass
ppt.m01	GIS	Con	1%	6.33	70.31	63.97	6.84	1.75	3.63	2.30	1.35	0.01	Pass
ppt.m02	GIS	Con	1%	7.44	69.55	62.11	5.72	2.69	2.94	1.10	0.11	0.01	Pass
ppt.m03	GIS	Con	1%	9.38	90.06	80.68	3.52	1.73	2.61	0.97	0.89	0.01	Pass
ppt.m04	GIS	Con	1%	9.77	103.05	93.29	13.89	5.77	3.08	0.13	1.72	0.01	Pass
ppt.m05	GIS	Con	1%	25.45	152.61	127.15	7.95	4.07	1.83	1.00	1.22	0.01	Pass
ppt.m06	GIS	Con	1%	28.57	146.24	117.68	21.84	7.01	1.41	1.82	3.68	0.01	Pass
ppt.m07	GIS	Con	1%	25.22	123.55	98.33	19.77	6.77	1.49	1.08	4.24	0.01	Pass
ppt.m08	GIS	Con	1%	16.13	121.45	105.32	14.81	5.96	0.71	1.77	3.32	0.02	Pass
ppt.m09	GIS	Con	1%	16.68	130.63	113.95	11.64	4.66	3.01	0.48	1.56	0.01	Pass
ppt.m10	GIS	Con	1%	18.72	110.64	91.91	7.88	2.93	3.51	1.44	0.11	0.01	Pass
ppt.m11	GIS	Con	1%	9.53	76.10	66.57	8.14	2.74	3.74	1.90	0.43	0.01	Pass
ppt.m12	GIS	Con	1%	7.10	75.11	68.01	5.82	2.21	3.18	1.66	0.64	0.01	Pass
Strata	GIS	Cat	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Pass
tmax	GIS	Con	2%	9.13	28.66	19.53	28.69	7.71	3.70	0.80	5.51	0.01	Pass
tmean	GIS	Con	2%	3.09	22.68	19.59	20.48	6.32	3.06	0.73	5.11	0.01	Pass
tmin	GIS	Con	2%	-2.98	17.26	20.24	12.34	4.77	2.28	0.64	4.43	0.01	Pass
IsolatedPools_number *	H20 (Direct)	Ord	88%	0	20	20	8.20	1.17	6.00	2.81	2.50	0.00	Pass
SoilMoist_MaxScore *	H20 (Direct)	Ord	79%	0	2	2	180.34	13.46	12.08	0.00	6.45	0.01	Pass
SoilMoist_MeanScore *	H20 (Direct)	Ord	79%	0	2	2	200.30	14.19	12.46	0.00	6.88	0.01	Pass
springs_score *	H20 (Direct)	Bin	94%	0	3	3	4.29	3.63	0.44	1.45	0.19	0.00	Pass

Candidate metrics	Group	Type	PctDom	Min	Max	Range	PvIvE	EvNE	PvNP	PvIwet	EvIdry	rf_MDA	Screened
SurfaceFlow_pct *	H20 (Direct)	Ord	56%	0	100	100	465.35	32.48	23.62	3.89	3.00	0.04	Pass
SurfaceSubsurfaceFlow_pct *	H20 (Direct)	Ord	60%	0	100	100	456.57	32.26	22.38	2.19	3.05	0.03	Pass
WaterInChannel_score *	H20 (Direct)	Ord	46%	0	6	6	531.00	31.42	24.02	6.28	6.11	0.04	Pass
HydricSoils_score	H20 (Indirect)	Bin	78%	0	3	3	90.55	10.93	6.47	0.13	7.76	0.00	Pass
WoodyJams_number	H20 (Indirect)	Ord	85%	0	100	100	1.83	1.56	1.50	2.45	1.19	0.00	Pass

Metric Screening

As an initial data exploration step, we visualized the relationships between actual streamflow duration class (hereafter "flow class") and indicators by ordinating all 95 metrics for all samples in the dataset in a nonmetric multidimensional scaling using Gowers' distance (Gower 1971). Convex hulls were drawn around each flow class to help visualize their distributions in ordination space. The ordination of all candidate metrics for Northern and Southern GP samples showed intermittent reaches overlapped with ephemeral and perennial reaches and there was more separation between ephemeral and perennial reaches (Figure 5). Axis 1 tended to separate reaches with flowing and dry conditions at the time of sample collection.

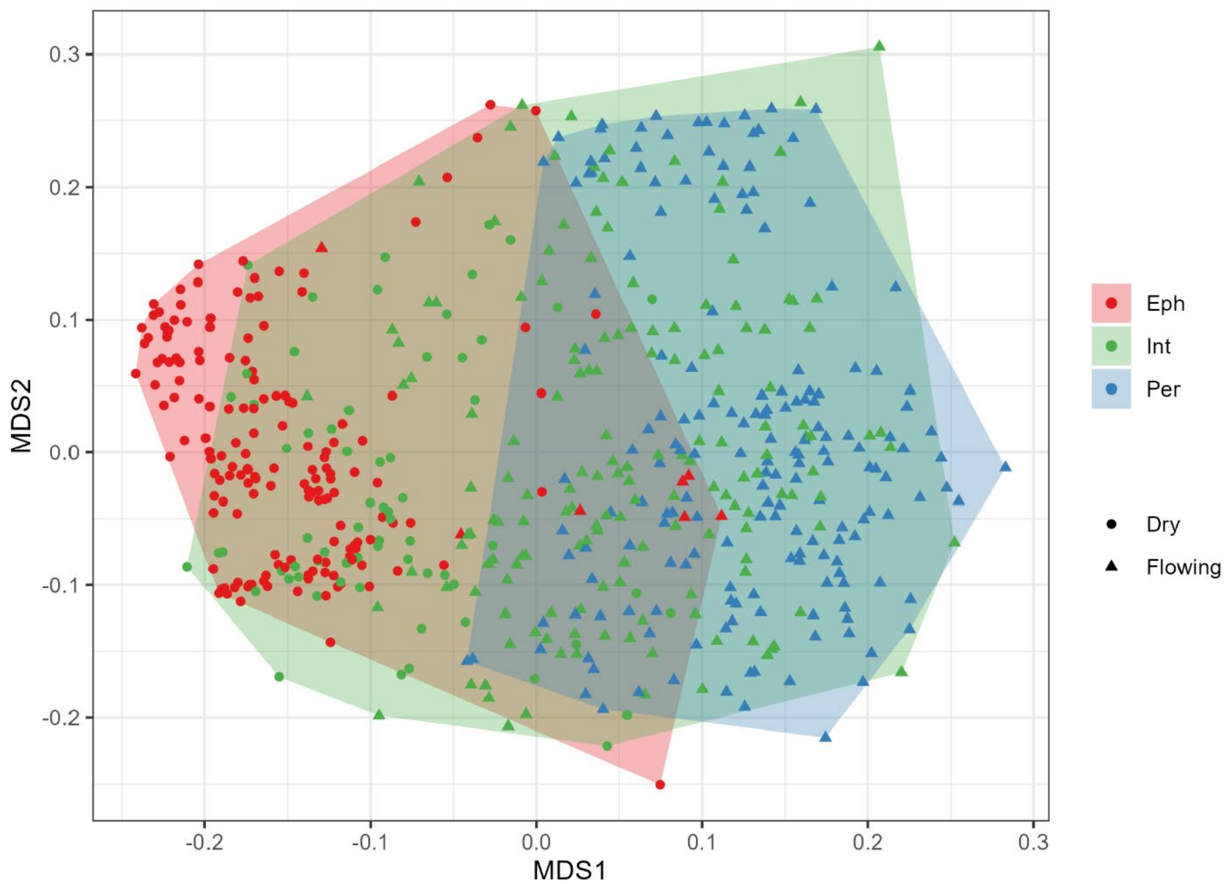

Figure 5. Beta SDAM GP candidate metric ordination.

Next, candidate metrics were evaluated using criteria for inclusion in the beta SDAM GP (Table 4):

- Distribution statistic criterion: calculated as percent dominance of the most common value (which was typically zero); all metrics had to meet this criterion.
- Criteria 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 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).

Candidate metrics had to meet at least one responsiveness criterion, in addition to the distribution criterion, to be considered in further analyses. An exception was Strata, which is the metric representing the four strata among which the study reaches were geographically distributed; therefore, it was included in further analyses. A total of 89 of the 95 candidate metrics were considered as screened metrics. Of the six metrics that failed, all but one (erosion_scored) failed due to Percent Dominance (PctDom) scores greater than 95%. Note that this evaluation was carried out using the testing dataset described in the next section.

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

As in the development of previous SDAMs, direct measures of water were excluded from further analysis. Metrics that directly measure water (e.g., soil moisture, number of isolated pools, water in channel) can greatly increase performance. However, such metrics introduce circularity (because water presence was used to confirm and update actual streamflow duration classes in the development data set) and may degrade the ability of the SDAM to perform well during atypical conditions, such as drought. See Mazor et al. (2021b) for a discussion of the implications of including direct measures of water presence as an indicator in SDAMs.

Data Preparation

Prior to method development, a portion of the data was withheld for use in final model testing. Samples from 20% of the study reaches, balanced by Class and Strata, were withheld into a "test" dataset. These samples were used to inform the final model selection and refinement, by evaluating the model on novel reaches. Samples from the remaining 80% of the reaches were used to develop (or "train") the model and are referred to hereafter as the training dataset.

Repeat reach visits

Of the 251 reaches included in the GP dataset, each was visited between one and four times, yielding a total of 692 samples. Figure 6 shows the distribution of repeat reach visits.

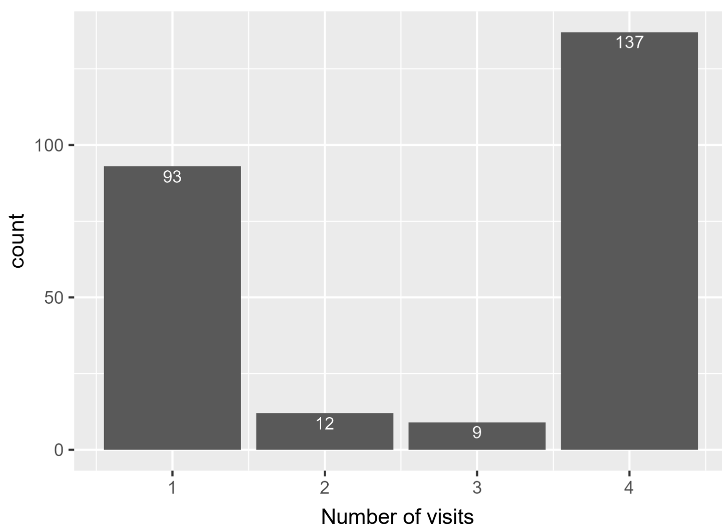

Figure 6. Distribution of number of visits across the 251 study reaches. Numbers inside of bars are the number of study sites with 1, 2, 3 or 4 visits.

To minimize bias, oversampling was performed on the training dataset (Figure 7). Oversampling is a common preprocessing step that serves to give under-represented classes more visibility in the data (Mohammed et al. 2020).

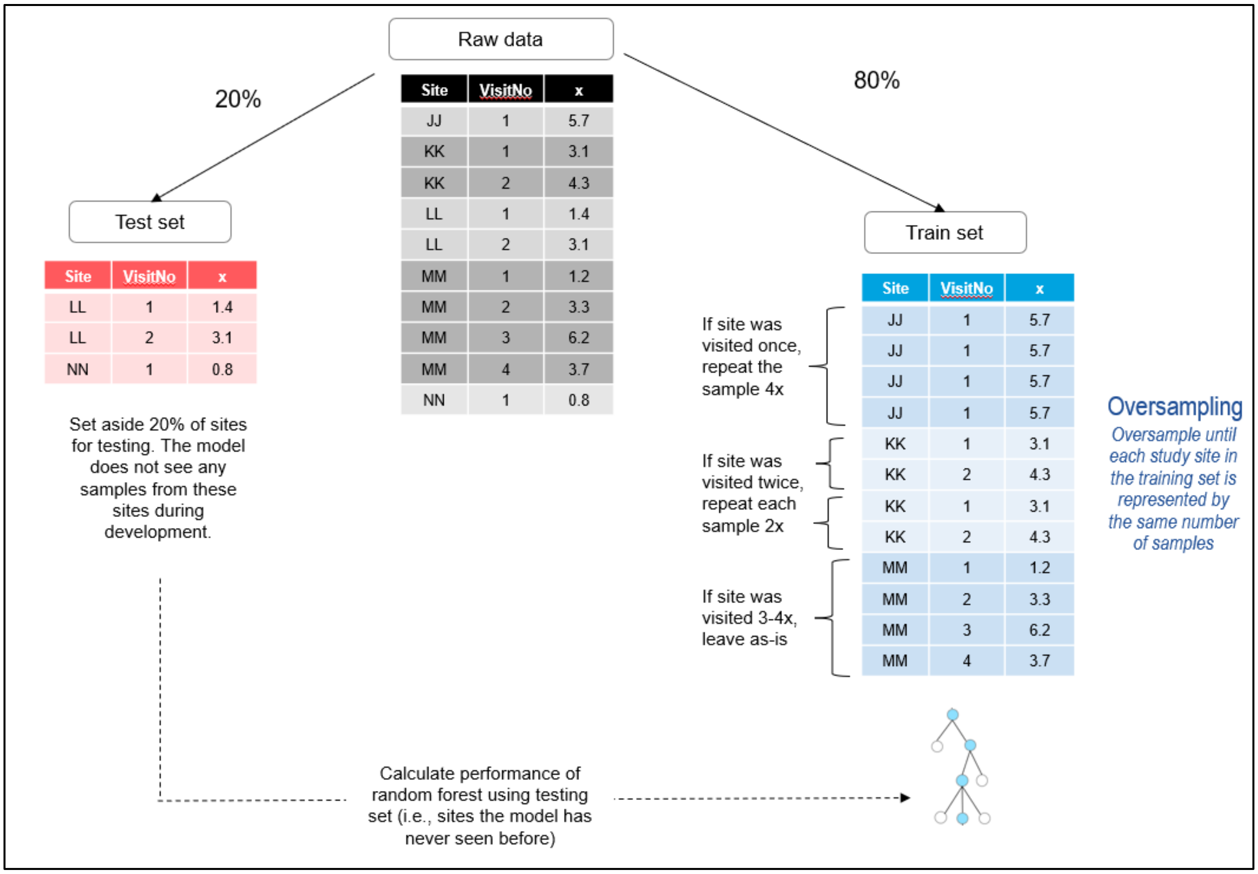

Figure 7.Oversampling process used for training dataset. x is a hypothetical candidate indicator

Oversampling was performed on the training dataset only (no manipulations were conducted on the test dataset) and was included the following steps:

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

The result of the oversampling process was that each study reach had three or four samples used in the analysis process for method development and the distribution of flow duration classes was preserved from the original training dataset to the oversampled training dataset, which also matched well to the distribution of flow duration classes within the testing dataset (Figure 8). Therefore, the augmented (oversampled) training data with 822 samples were used in the next step of the method development analysis process to select screened metrics.

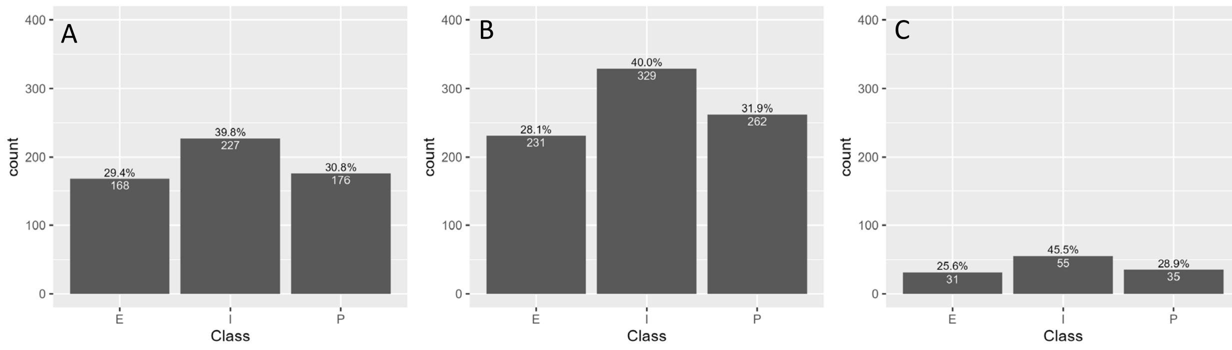

Figure 8: Distribution of ephemeral (E), intermittent (I), and perennial (P) classes in the (A) training dataset before oversampling, (B) training dataset after oversampling, and the (C) testing dataset. Shown for each bar is the number of samples for a streamflow duration class and the percent of samples within the datasets. A balanced distribution between classes is important to mitigate against bias and improve model accuracy.

Metric selection

The screened metrics were reduced to a final set of metrics for the beta SDAM GP 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 incrementally by eliminating the least important metrics. Here, the most complex model was first considered. Then, 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 variable was eliminated in each successive step. The best performing model (highest accuracy in predicting true streamflow duration class) 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 best-performing 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. 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 (James et al. 2013).

This modeling process (including RFE) was applied to the dataset to produce 10 models:

- The entire Great Plains (Northern and Southern Great Plains) dataset (unstratified model set)
- Datasets for each stratum (stratified model sets): Central Prairie, Northern Prairie, Southern Plains, and Upper Midwest (Figure 3)

There are advantages and disadvantages to including geospatial metrics in an SDAM. Geospatial metrics may improve SDAM performance but would require GIS analysis in the application of the resulting method. See Mazor et al. (2021b) for a discussion of the implications of including geospatial metrics in SDAMs.

The 10 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 5, and the selected metrics for each model are shown in Figure 9.

Table 5. Design characteristics of the 10 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**
Unstratified models						
Unstratified	Entire Great Plains	822	121	61	11	0.13
Unstratified GIS	Entire Great Plains	822	121	82	6	0.03
Models stratified by region						
Stratified	Northern Prairie	174	18	61	20	0.20
Stratified	Southern Plains	180	29	61	9	0.10
Stratified	Upper Midwest	237	38	61	7	0.17
Stratified	Central Prairie	231	36	61	20	0.10
Stratified GIS	Northern Prairie	174	18	82	11	0.02
Stratified GIS	Southern Plains	180	29	82	18	0.07
Stratified GIS	Upper Midwest	237	38	82	13	0.01
Stratified GIS	Central Prairie	231	36	82	20	0.02

Biological metrics, particularly those based on aquatic invertebrates, were among the most widely selected metrics across model sets (Figure 9). Among non-biological metrics, mean bankfull width was the only frequently selected geomorphological metric.

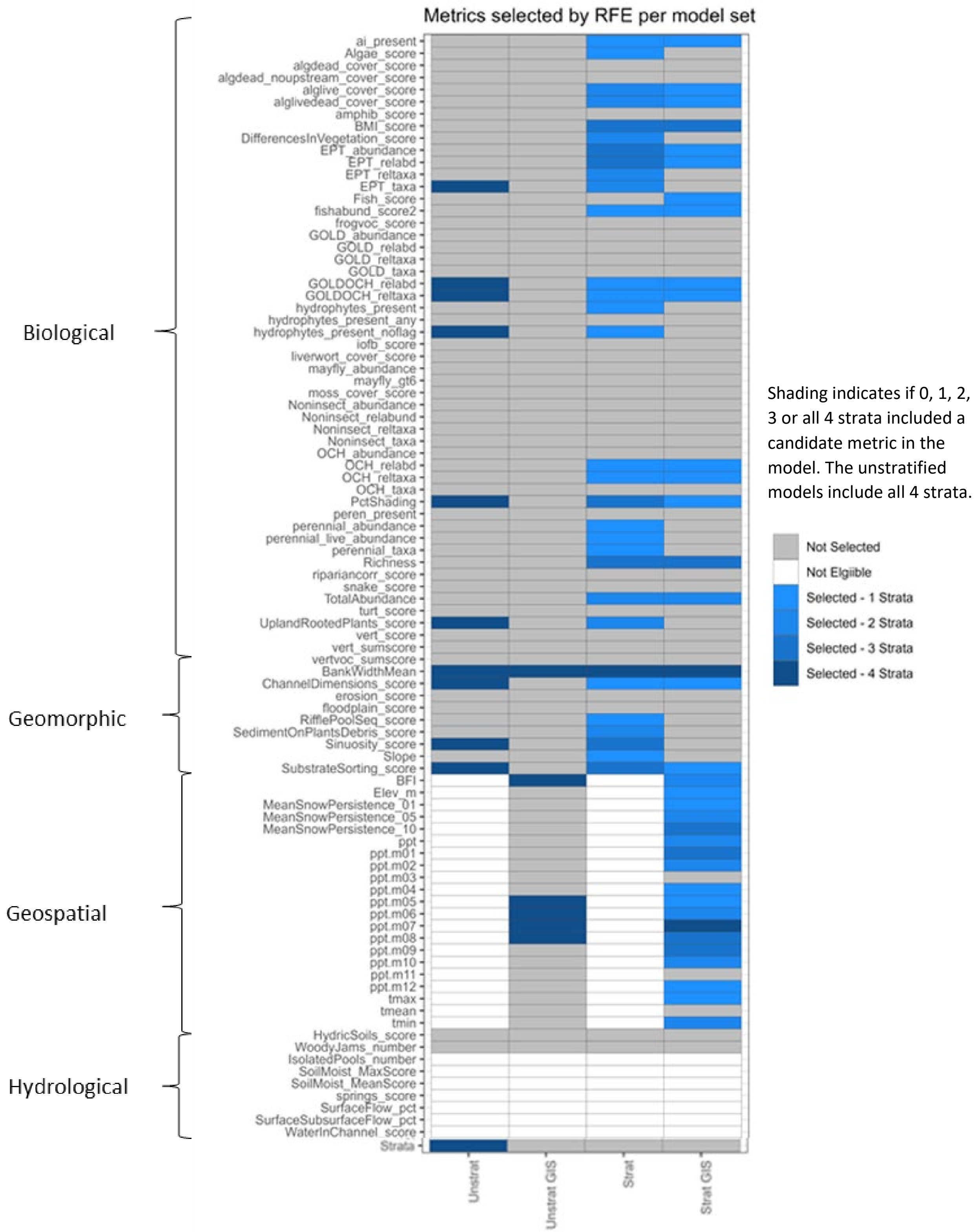

Figure 9. Screened metrics (left) selected by RFE for each model set (bottom). White tiles indicate that a screened metric was ineligible for selection in that model set (e.g., Elev_m was ineligible for models that did not allow GIS metrics). X-axis labels refer to model sets described in Table 6. Y-axis labels refer to screened metrics described in Table 4 and Appendix A.

Preliminary model calibration and performance assessment

Random forest models were fit for each of the 10 models 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.

Model performance evaluation focused on two aspects: accuracy and repeatability (Table 7 and Figure 8). 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 [EvALI], perennial versus wet intermittent reaches [PvIwet], etc.; Table 5). 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), suggesting 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.

Repeatability, or precision, was assessed using data from the 158 reaches that were resampled (Figure 6) and was calculated as the percent of reaches where model classifications from repeated samples at the same reach were consistent (regardless of classification accuracy). Due to the limited amount of data, repeatability was only assessed for the entire GP and not within each stratum.

Along with the 10 models, the classification accuracy of existing SDAMs (models) for the PNW (Nadeau 2015), NM (NMED 2011), and beta AW (Mazor et al. 2021a) as applied to the GP dataset was also compared (Table 6 and Figure 10).

Table 6. Performance evaluation of the 10 RF model options developed for the GP and 3 existing SDAMs. PvIvE: Percent of reach samples classified correctly as perennial, intermittent, or ephemeral. EvALI: Percent of reach samples classified correctly as ephemeral or at least intermittent. PvNP: Percent of reach samples classified correctly as perennial or non-perennial. PvIwet: Percent of flowing reach samples classified correctly as perennial or intermittent. IvEdry: Percent of dry reach samples correctly classified as intermittent or ephemeral. Train: Result for training data. Test: Result for testing data. Model sets are described in Table 6. AW: Results for the Beta SDAM AW. PNW: Results for the SDAM PNW. NM: Results for the SDAM NM.

	Accuracy										
	PvIvE		EvALI		PvNP		PvIwet		IvEdry		
Model set	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Precision
Unstrat	87	74	93	89	93	84	89	75	85	73	87
Unstrat GIS	97	50	98	73	99	71	98	39	96	66	94
Strat	86	72	93	89	92	83	86	73	87	72	83
Strat GIS	97	69	98	93	99	76	98	59	97	79	92
AW	43	48	78	85	46	52	42	48	39	43	71
PNW	47	49	84	87	62	62	40	40	63	57	78
NM	55	54	84	87	68	66	55	52	56	52	86

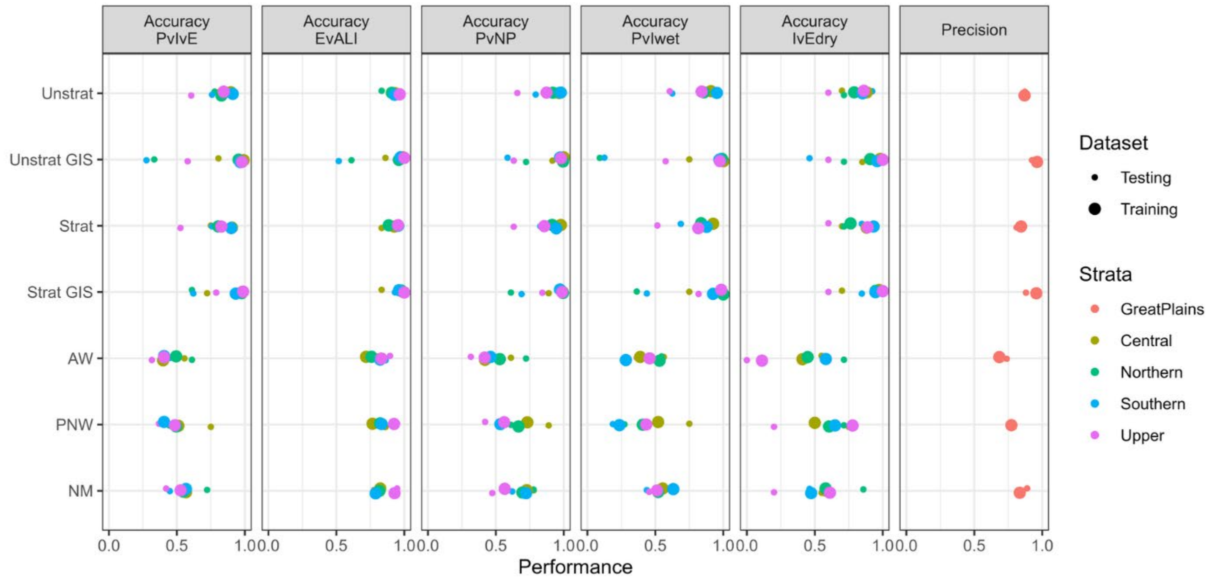

Figure 10. Performance of the various model sets evaluated within strata defined by sub-region. PvIvE: Proportion of reach samples classified correctly as perennial, intermittent, or ephemeral. The y-axis labels on the left indicate the stratifications used to develop the models (if any). EvALI: Proportion of reach samples classified correctly as ephemeral or at least intermittent. PvNP: Proportion of reach samples classified correctly as perennial or non-perennial. PvIwet: Proportion of flowing classified correctly as perennial or intermittent. IvEdry: Proportion of dry reach samples correctly classified as intermittent or ephemeral. Model sets are described in Table 5. AW: Results for the Beta SDAM AW; PNW: Results for the SDAM PNW; NM: Results for the SDAM NM.

Selection of the final model

SDAM models newly developed through the current effort using data from the GP had better performance than previously developed SDAMs, confirming higher classification accuracy is achieved through development of region-specific SDAMs.

Among the 10 models, performance was highest in the training dataset for the unstratified and stratified model versions that included GIS metrics (Figure 10; Table 6). However, performance of the models containing GIS data sharply decreased when evaluated against the testing dataset, indicating that the GIS models were overfitting to the training dataset (Figure 11).

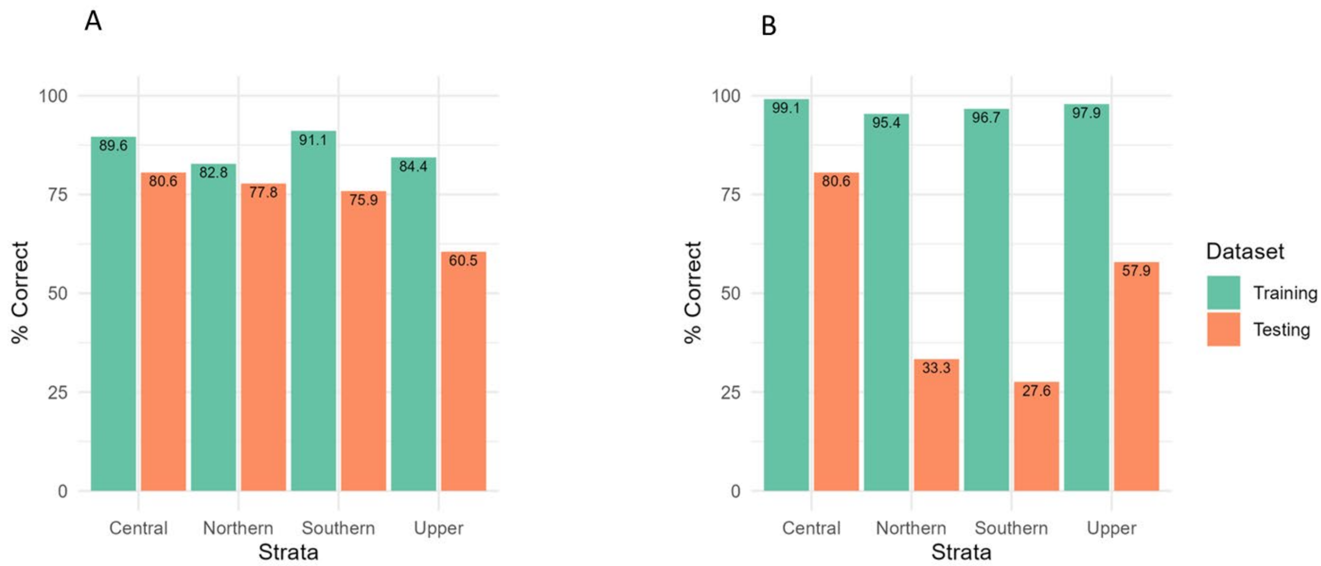

Figure 11: Accuracy of the (A) unstratified GP model without GIS metrics and (B) unstratified GP model with GIS metrics based on training and testing datasets by strata (943 total observations). Numbers shown in bars are the percent of correctly classified samples as perennial, intermittent, or ephemeral.

Between the stratified and unstratified models that did not include GIS metrics, performance was similar and there was no clear best model (Figure 10; Table 6). 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 sub-region 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; Figure 12) compared to distinguishing between all three classes (PvIvE; Figure 11).

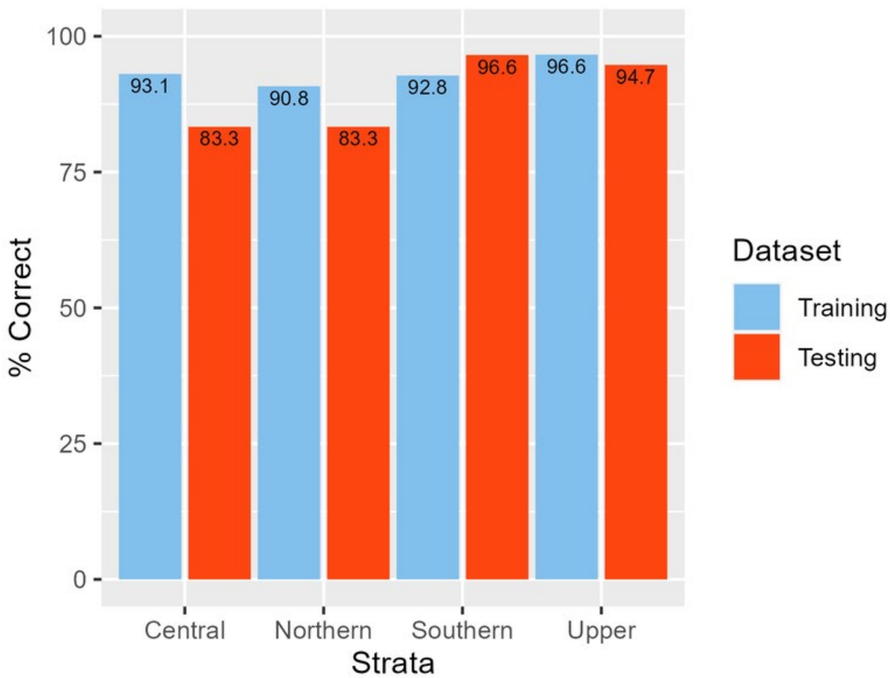

Figure 12: Accuracy of the unstratified Great Plains model (no GIS) in distinguishing between Ephemeral and At Least Intermittent for training and testing datasets by strata.

For these reasons, the unstratified model (no GIS) was selected as the beta SDAM GP to apply to the GP.

Unstratified (no GIS) model description

Eleven metrics were selected via RFE for the unstratified (no GIS) model. The metrics are shown in Figure 13 by their order of importance. Here, importance to the random forest model is considered in two ways: (1) through mean decrease in accuracy and (2) through mean decrease in Gini Index, which is a measure of node impurity, or how important the metric is in splitting between different flow duration classes.

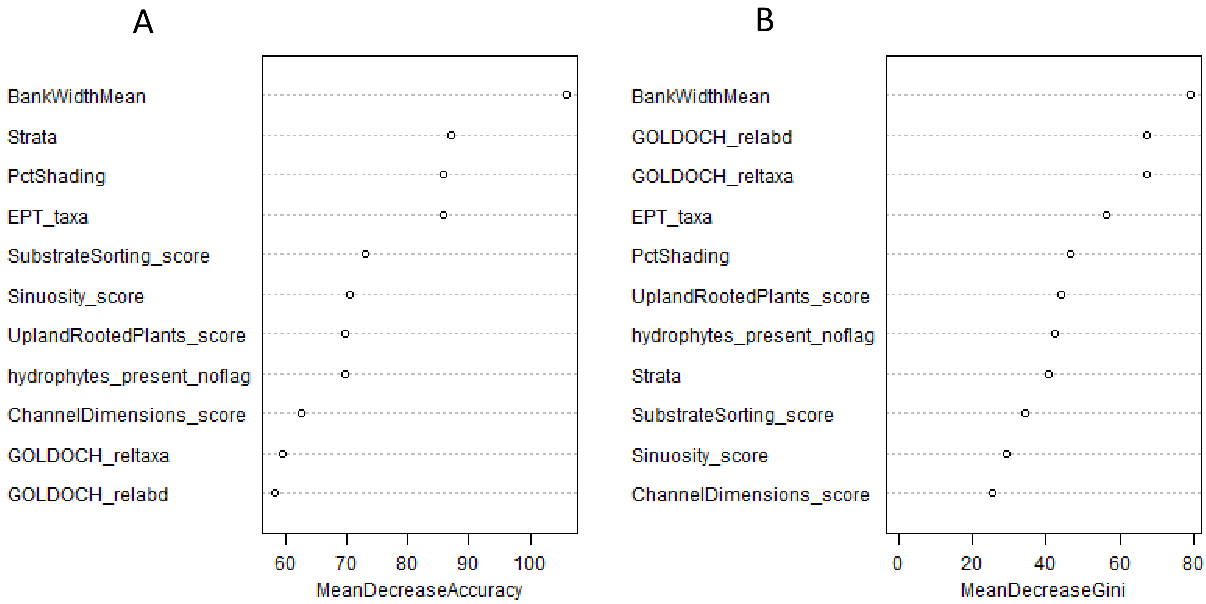

Figure 13: Metrics included in the unstratified (no GIS) model, by their order of importance. (A) Mean Decrease in Accuracy is the relative loss in predictive performance when the particular variable is omitted from the model. (B) Mean Decrease in Gini: Gini Index is a measure of node impurity, or how important the variable is in splitting between different streamflow duration classes.

To evaluate the overall performance of the unstratified (no GIS) model, confusion matrices were created for both training and testing datasets (Figure 14). Overall classification accuracy was higher for ephemeral reach samples (training 89.2%, testing 90.3%) than for perennial (training 86.6%, testing 74.3%) and intermittent reach samples (training 85.7%, testing 61.8%). No perennial reach samples were misclassified as ephemeral in either testing or training datasets; only two ephemeral reach samples were misclassified as perennial in the training dataset. The unstratified (no GIS) model had similar misclassification predictions of intermittent reach samples as ephemeral or perennial reaches in the testing and training datasets.

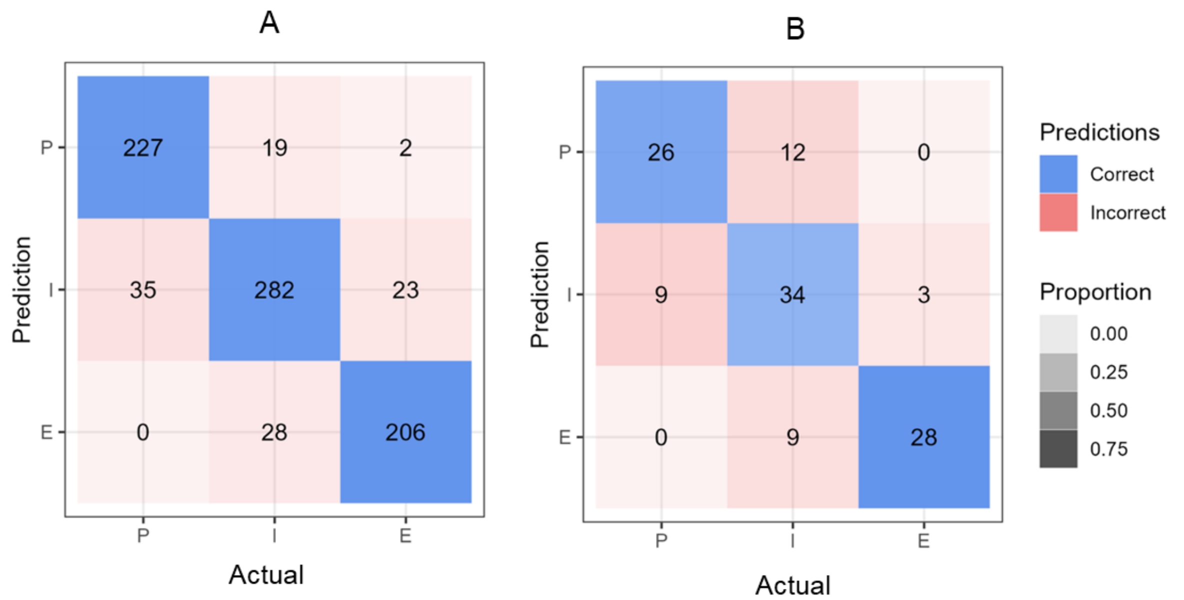

Figure 14. Confusion matrices of the (A) training vs (B) testing dataset on the unstratified (no GIS) model. The training dataset contained a total of 822 reach samples and the testing dataset contained 121 reach samples. X-axis shows actual flow duration class and Y-axis shows predicted flow duration class. Blue diagonal indicates correct predictions. P = perennial, I = intermittent, and E = ephemeral. Shading of boxes in matrices describe the proportion of reach samples in each dataset.

Simplification of the selected model

Upon selection of the unstratified (no GIS) model, the next step was to simplify the selected model. Simplification was intended 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 selected 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 more than 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 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 site visit.

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 upland rooted plants and hydrophyte 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 of three 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 Cohen's Kappa - a measure of accuracy. Note that the Kappa statistic varies from 0 to 1, where 0 equals agreement equivalent to chance and 1 equates to perfect agreement.

Ten refinements of the unstratified (no GIS) model were performed and are summarized in Table 7 and Figure 15. For example, a refinement made between Version 0 and Version 1 was the binning the mean bankfull width from continuous data to binary data (<20 m and ≥20 m).

*Table 7. Ten model refinement versions of the statistically determined unstratified model without GIS metrics. Includes refinement descriptions, metrics included and accuracy of refined models (PvIvE: Percent of reach samples classified correctly as perennial, intermittent, or ephemeral; EvALI: Percent of reach samples classified correctly as ephemeral or at least intermittent) as measured using the testing dataset. **Bold** metrics included in refined models identify the iterative metric refinements made to the previous model refinement version.*

Version 0	**Version 1**	**Version 2**	**Version 3**	**Version 4**	**Version 5**	**Version 6**	**Version 7**	**Version 8**	**Version 9**	**Version 10**
Unstratified, no GIS model (no refinements)	Bin continuous variables into discrete groups	GOLDOCH presence/ absence	GOLD presence/ absence	OCH presence/ absence	GOLD and OCH presence/ absence	GOLDOCH abundance binned	without GOLDOCH variables	Upper, Northern and Central strata combined	Southern, Northern and Central strata combined	Upper and Northern strata combined
Metrics Included										
BankWidthMean	**BankWidth binned**	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned	BankWidth binned
Strata	Strata	Strata	Strata	Strata	Strata	Strata	Strata	**Strata UNC**	**Strata SNC**	**Strata UN**
PctShading	**PctShading binned**	PctShading binned	PctShading binned	PctShading binned	PctShading binned	PctShading binned	PctShading binned	PctShading binned	PctShading binned	PctShading binned
EPT taxa	**EPT taxa binned**	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned	EPT taxa binned
Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score	Substrate Sorting score
Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score	Sinuosity score
hydrophytes present	**hydrophytes binned**	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned	hydrophytes binned
Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score	Upland Rooted Plants score
Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score	Channel Dimensions score
GOLDOCH reltaxa	**GOLDOCH reltaxa binned**	**GOLDOCH y/n**	**GOLD y/n**	**OCH y/n**	**GOLD y/n**	**GOLDOCH relabd binned**				
GOLDOCH relabd					**OCH y/n**					
Model Accuracy										
PvIvE: 72.7 EvALI: 90.1	PvIvE: 68.6 EvALI: 90.1	PvIvE: 67.8 EvALI: 88.4	PvIvE: 65.3 EvALI: 89.3	PvIvE: 61.2 EvALI: 83.5	PvIvE: 65.3 EvALI: 88.4	PvIvE: 66.1 EvALI: 88.4	PvIvE: 62.8 EvALI: 84.3	PvIvE: 68.6 EvALI: 87.6	PvIvE: 62.8 EvALI: 84.3	PvIvE: 62.8 EvALI: 86.8

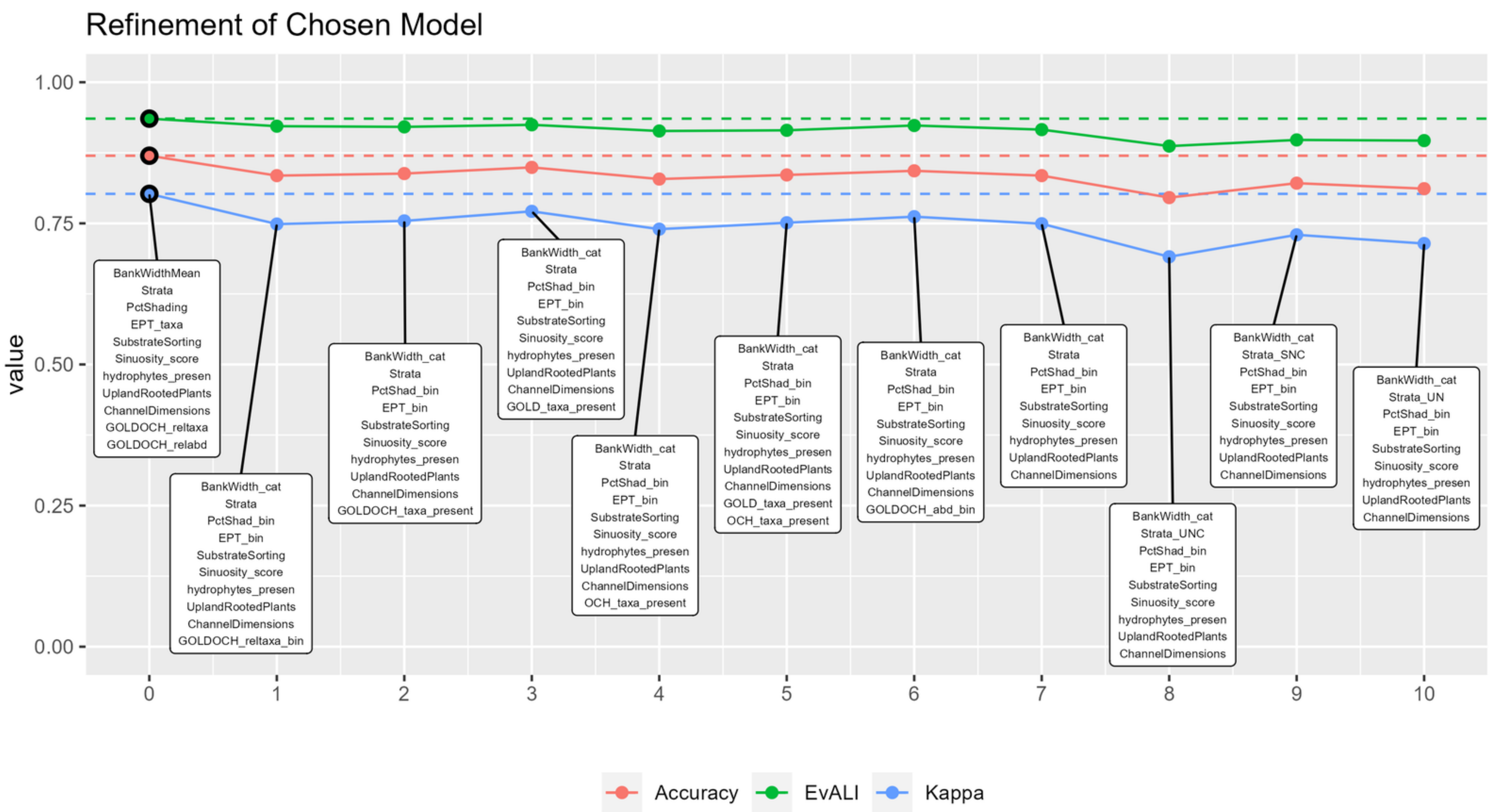

Figure 15. Impact of refinement of metric set on the model performance relative to the unstratified (no GIS) model using the training dataset. Each refinement description is relative the description at 0 (unstratified, no GIS model). Black circles indicate the highest Accuracy, EvALI, and Kappa scores. Dashed lines show performance of the unstratified (no GIS) model.

As shown by the decreasing performance lines in Figure 15, none of the attempted refinements improved the performance of the unstratified (no GIS) model in terms of PvIvE accuracy, EvALI accuracy, or Cohen's Kappa. However, the slight decrease in model predictive performance was weighed against the relative advantages of simplifying field data collection. For this reason, the two GOLDOCH metrics were removed due to the data collection effort required.

Final model selection

After consultation with the PDT and RSC, the final model selected was the Version 8 refinement of the unstratified (no GIS) model. The Version 8 refinement differs from the unstratified (no GIS) model as follows:

- BankWidthMean, originally a continuous metric on the scale of 0.4 – 68.3 meters, was binned into two discrete groups (less than 20m, greater than or equal to 20m) based on visual interpretation of the metric distributions across ephemeral, intermittent, and perennial classes, and through trial-and-error testing.
- Strata, originally containing four strata, was simplified into the two Great Plains Regions: the Southern Great Plains, and the Northern Great Plains (containing the Upper Midwest, Northern Prairie, and Central Prairie strata).
- Percent Shading, originally a continuous metric ranging from 0-100%, was binned into discrete groups (less than 10% and greater than or equal to 10%) based on visual interpretation of the metric distributions across ephemeral, intermittent, and perennial classes, and through trial-and-error testing.
- Number of EPT families ranged from zero to seven in the original dataset. This was simplified in the refined model into two discrete groups (zero to one family, two or more families). This metric binning was based on visual interpretation of the metric distributions across streamflow duration classes and through trial-and-error testing. However, the beta SDAM GP User Manual recommends enumerating up to five families, if present, to provide redundancy.
- Number of hydrophytic species recorded ranged from zero to eight species in the original dataset. This was simplified in the refined model into two discrete groups (fewer than two species, 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. However, the beta SDAM GP User Manual recommends enumerating up to five families, if present, to provide redundancy.
- GOLDOCH_reltaxa and GOLDOCH_relabd were removed from the model.

The performance of the Version 8 refined model is shown as confusion matrices (Figure 16). There was a decrease in performance based on the training dataset (Figure 15), relative to the unstratified (no GIS) model, but of similar performance based on the testing dataset (Table 7).

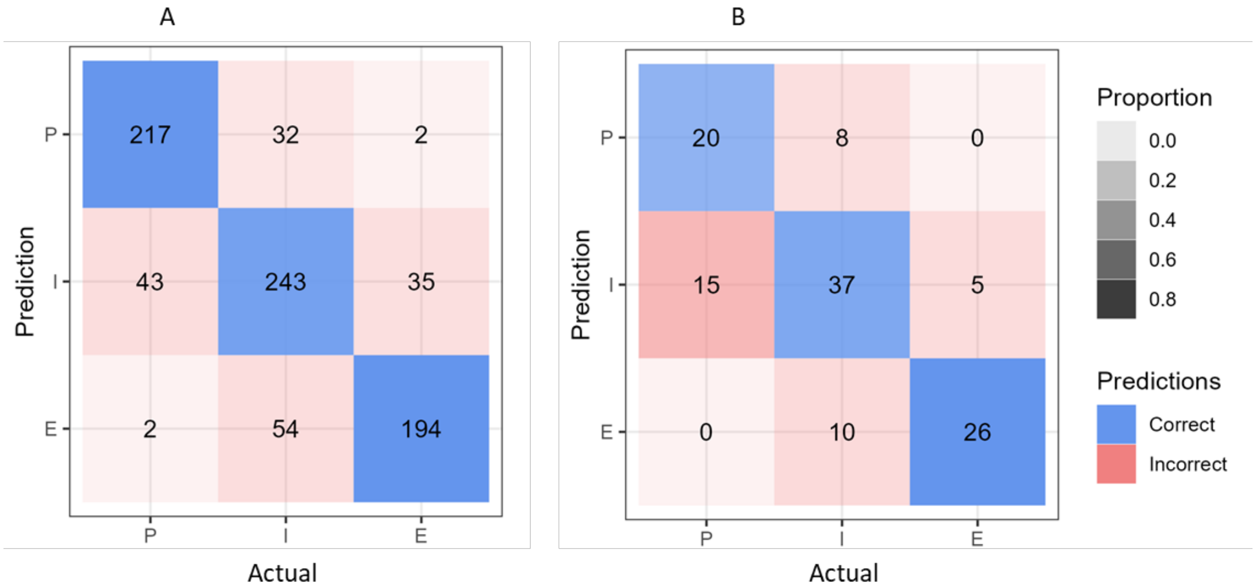

Figure 16: Performance of the selected refined model based on (A) training (822 reach samples) and (B) testing (121 reach samples) datasets. X-axis shows actual flow duration class and Y-axis shows predicted flow duration class. Blue diagonal indicates correct predictions. P = perennial, I = intermittent, and E = ephemeral. Shading of boxes in matrices describe the proportion of reach samples in each dataset.

Using the refined model, two reaches in the training dataset continued to incorrectly predict *ephemeral* when the correct classification was *perennial* during one of four visits to the sites. In addition, two reaches in the training dataset incorrectly predicted *perennial* when the correct classification was *ephemeral* during one of four visits to the sites. The four sites were the following:

Reach Code	State	Strata	Actual	Predicted
TXSB14	TX	Southern	P	E
WIUB20	WI	Upper	P	E
WIUB37	WI	Upper	E	P
WYNB1	WY	Northern	E	P

No incorrect predictions between *ephemera*l and *perennial* occurred using the refined model on the testing dataset.

Increased confidence required for classifications

Random forest models created for classification traditionally make assignments based on the class that receives the highest number of votes by each "tree" in the forest. Thus, in a three-way decision (ephemeral, intermittent, or perennial), the class with the most votes could receive much less than a majority of all votes—as low as 34%. Given concern that such low-confidence classifications may not provide sufficient defensibility for some management decisions, approaches to distinguish between high- and low-confidence classifications were explored.

We explored increasing the minimum number of votes required to make a confident classification from 30% to 100% by increments of 1% to understand the effect on classification. When the selected refined 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 that distinguishing between ephemeral and at least intermittent reaches is a high priority use of the beta SDAM GP. The percent of reaches under each of the five possible classifications with increasing minimum vote agreement thresholds were calculated.

At a minimum required proportion of votes of 0.5, only 3.5% of reach samples in the training dataset (5% of reach samples in the test dataset) were classified as *at least intermittent*, and none were classified *need more information* (Figure 16). Classifications of *at least intermittent* first appear with a minimum proportion of 0.37 in the training dataset (0.45 in the testing dataset), whereas classifications of *need more information* appear at 0.51 in both the training and testing datasets. Although it cannot be ruled out, it is unlikely that the beta SDAM GP will result in a classification of *need more information.* Based on these results the RSC recommended a minimum proportion threshold of 0.5 for flow classification.

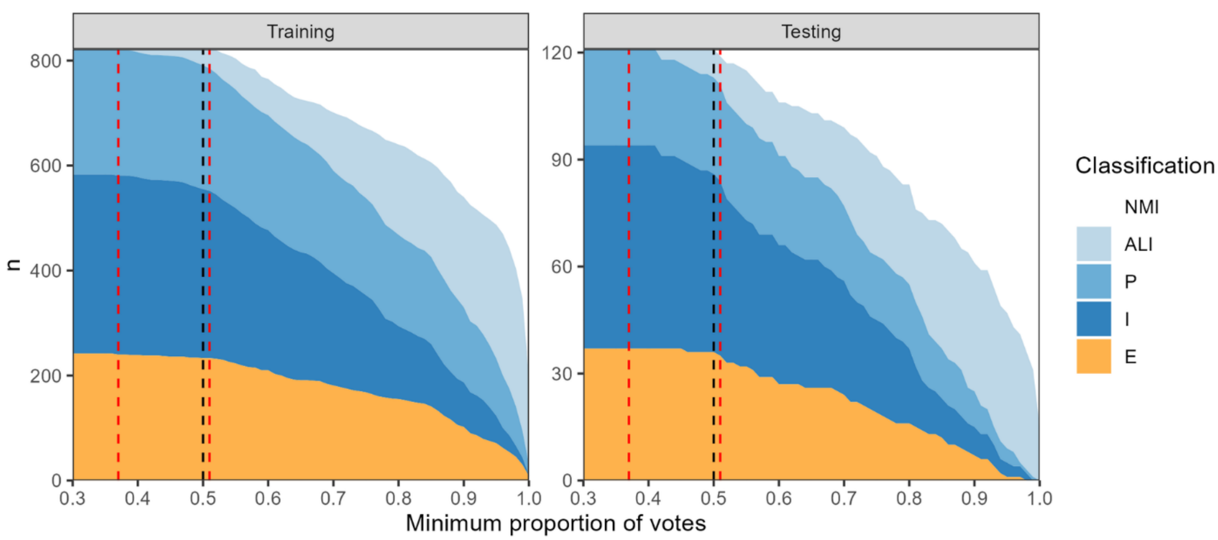

Figure 17. Influence of the minimum proportion of votes required to make a classification on n (the number of reaches in each class). NMI: Need more information. ALI: At least intermittent. P: Perennial. I: Intermittent. E: Ephemeral. The vertical black line represents a minimum proportion of required 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) for the dataset.

Evaluation of single indicators of at least intermittent flow

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.

We evaluated single indicators used in previous SDAMs. The number of instances where inclusion of a prior single indicator would correct a misclassification (i.e., the reach was truly intermittent or perennial) and would introduce a misclassification/mistake (i.e., the reach was truly ephemeral) was quantified. All single indicators investigated had minimal impact on performance or introduced more errors than were corrected (Figure 18). Based on these results, the RSC did not recommend including any of the evaluated single indicators in the beta SDAM GP.

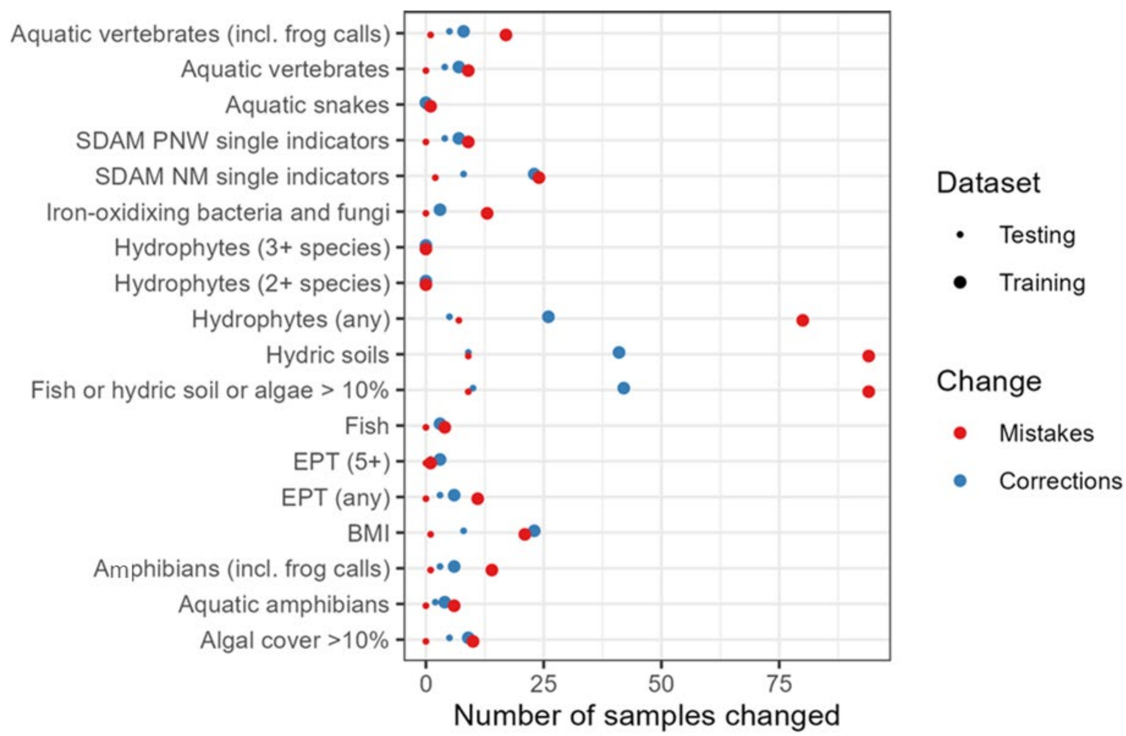

Figure 18. Influence of single indicators on performance of the refined model

Performance of the beta SDAM GP

Performance of the selected refined model (with a minimum proportion voting threshold of 0.5) for the beta SDAM GP is summarized in Table 8. The overall classification accuracy among the three classes (*perennial*, *intermittent*, *ephemeral*) was 81% in the training dataset (and 68% in the testing dataset), but this accuracy increased to 89% in the training dataset (and 87% in the testing dataset) when only *ephemeral* versus at *least intermittent* classifications were considered (i.e., both blue and green cells in Table 8 were treated as correct). Note, after applying the voting threshold one of the two instances in the training dataset that incorrectly predicted *perennial* when the correct classification was *ephemeral* changed to a prediction of *at least intermittent* (WYNB1).

Table 8. Classifications of the final version of the beta SDAM GP. Blue cells indicate correct classifications of perennial, intermittent, at least intermittent and ephemeral reaches, whereas green cells indicate correct classifications of ephemeral versus at least intermittent. Green numbers represent the reach visits with matching actual and predicted classes and red numbers are reach visits with non-matching actual and predicted classes.

	Actual streamflow duration class					
Predicted Class	Ephemeral (Training)	Ephemeral (Testing)	Intermittent (Training)	Intermittent (Testing)	Perennial (Training)	Perennial (Testing)
Ephemeral	193	24	47	9	2	0
Intermittent	30	5	236	31	40	14
ALI	7	2	17	7	5	1
Perennial	1	0	29	8	215	20

Using the LandUse indicator to identify reaches that were disturbed (LandUse = urban or agriculture, alone or in combination with any other land use category) and not disturbed (LandUse does not include urban or agriculture) at the time of the site visit, there were 133 individual reaches identified as disturbed during at least one site visit with a total of 229 disturbed samples (before augmentation). There were 192 (34%) and 37 (31%) disturbed samples included in the training and testing datasets, respectively. These tallies and the accuracy results provided below focus on the samples of the original dataset before augmentation (n = 692).

Among the samples identified as disturbed by human activity in the training dataset, accuracy among all classes was 76%, which improved to 86% when only *ephemeral* versus *at least intermittent* classifications were considered. For samples in the training dataset that were not disturbed, the accuracy values indicated similar performance to that of the disturbed sites (i.e., 73% PvIvE and 84% EvALI).

For the samples in the testing dataset, the accuracy among all classes for disturbed sites was 78%, which improved to 89% when only *ephemeral* versus *at least intermittent* classifications were considered. For samples in the testing dataset that were not disturbed, accuracy among all classes was 64%, which improved to 86% when only *ephemeral* versus *at least intermittent* classifications were considered.

Data and code availability

All data used to develop the method and R code used in analysis are available at the following git repository: https://doi.org/10.23719/1527943

Next steps

The beta SDAM GP is being made available for one year for public review and comment while additional data at the study sites are collected through 2022, after which a final method will be developed and released to replace the beta method.

Acknowledgements

The development of this method and supporting materials was guided by a regional steering committee (RSC) consisting of representatives of federal regulatory agencies in the Great Plains of the U.S.: Micah Bennett (U.S. Environmental Protection Agency [USEPA]—Region 5), Andrew Blackburn (U.S. Army Corps of Engineers [USACE]—Great Lakes and Ohio Valley Division, Chicago District), Kirsten Brown (USACE—Mississippi Valley Division, Rock Island District), Billy Bunch (USEPA—Region 8), Gabrielle C. L. David (USACE—Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory), Gabriel DuPree (USEPA—Region 7), Wayne Fitzpatrick (USACE—Southwestern Division, Galveston District), Jeremy Grauf (USACE—Northwestern Division, Omaha District), Ed Hammer (USEPA—Region 5), Rachel Harrington (USEPA—Region 8), Faye Healy (USACE—Mississippi Valley Division, St. Paul District), Shawn Henderson (USEPA—Region 7), Rob Hoffman, (USACE—Southwestern Division, Tulsa District), Rose Kwok (USEPA—Headquarters), April Marcangeli (USACE—Mississippi Valley Division, St. Paul District), Tunis McElwain (USACE—Headquarters), Elizabeth Shelton (USACE—Southwestern Division, Galveston District), Chelsey Sherwood (USEPA—Region 6), Loribeth Tanner (USEPA—Region 6), Kerryann Weaver (USEPA—Region 5), and Matt Wilson (USACE—Headquarters).

We thank Abel Santana, Robert Butler, Duy Nguyen, Kristine Gesulga, and Anne Holt for assistance with data management, and Abe Margo, Alex Martinez, Addison Ochs, Morgan Proko, Alec Lambert, Zak Erickson, Alex Berryman, Jack Poole, Joe Kiel, Joe Klein, Jackson Bates, Buck Meyer, Margaret O'Brien, Elliot Broder, Jason Glover, and James Treacy for assistance with data collection. Amy James provided document editorial and formatting assistance.

Numerous researchers and land managers with local expertise assisted with the selection of study reaches to calibrate the method: Tim Bonner, Jeffrey Brenkenridge, Taylor Dorn, Tim Fallon, John Genet, Linda Hansen, Garret Hecker, Stephanie Kampf, Kort Kirkeby, Ji Yeow Law, John Lyons, Kyle McLean, Miranda Meehan, Steve Robinson, Mateo Scoggins, Patrick Trier, Linda Vance, Ross Vander Vorste, and Jason Zhang.

References

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.

Eng, K., D. M. Wolock, and M. D. Dettinger. 2016. Sensitivity of intermittent streams to climate variations in the USA. River Research Applications 32: 885–895.

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.

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27: 857-874.

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.

Hedman, E. R., and Osterkamp, W.R. 1982. Stream Flow Characteristics Related to Channel Geometry of Streams in Western United States. USGS Water-Supply Paper 2193, Washington, DC. p. 17. DOI:10.3133/wsp2193.

Hewlett, J. D. 1982. Principles of Forest Hydrology; University of Georgia Press: Athens, GA, USA, p. 192.

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: 1000005.

James, A., K. McCune, and R. Mazor. 2022. Review of Flow Duration Methods and Indicators of Flow Duration in the Scientific Literature, Great Plains of the United States. Document No. EPA-840-B-22006. 56 pp. (Available from: https://www.epa.gov/system/files/documents/2022-09/FlowDurationLitReview-gp.pdf)

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning. Springer, NY. 440 pp.

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. Environmental Protection Agency, Cincinnati, OH. 25 pp.

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

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

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. Document No. EPA 800-K-21001. 83 pp. (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, 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. McCune, A. 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. Document No. EPA-840-B-21008. 116 pp. (Available from: https://www.epa.gov/system/files/documents/2021-12/beta-sdam-for-the-wm-user-manual.pdf)

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 – Data Supplement. Document No. EPA 840-R-22002. 38 pp. (Available from: https://www.epa.gov/system/files/documents/2022-05/WM%20Data%20supplement_5-4-22%20FINAL.pdf)

Mohammed, R., J. Rawashdeh, and M. Abdullah. 2020. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. Pages 243-248 in Proceedings of the 11th International Conference on Information and Communication Systems. Irbid, Jordan 7-9 April 2020.

Nadeau, T.-L. 2015. Streamflow Duration Assessment Method for the Pacific Northwest. EPA 910-K-14-001, U.S. Environmental Protection Agency. 36 pp. (Available from: https://www.epa.gov/sites/default/files/2016-01/documents/streamflow_duration_assessment_method_pacific_northwest_2015.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. Surface Water Quality Bureau, New Mexico Environment Department, Albuquerque, NM. 35 pp. (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)

Omernik, J.M. 1995. Ecoregions: a framework for managing ecosystems. The George Wright Forum 12: 35–50.

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 USA 114: 7373–7378.

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 J-WECD-ECB-SOP-1016-02. Environmental Protection Agency, Cincinnati, OH. 13 pp.

United States Environmental Protection Agency (USEPA). 2019. Flow duration protocol version 2.1. 38 pp.

Vengosh, A., R. B. Jackson, N. Warner, T. H. Darrah, and A. Kondash. 2014. A critical review of the risks to water resources from 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. Trier, and F. B. Vanderbilt. 2016. Synthesizing the Scientific Foundation for Ordinary High Water Mark Delineation in Fluvial Systems. Wetlands Regulatory Assistance Program ERDC/CCREL SR-16-5, U.S. Army Corps of Engineers Engineer Research and Development Center. 217 pp. (Available from: https://apps.dtic.mil/sti/pdfs/AD1025116.pdf)

Wolock, D. M. 2003. Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 03-263, digital dataset. (Available from: https://water.usgs.gov/lookup/getspatial?bfi48grd)

Appendix A: Glossary of Terms Used

Streamflow Class	Description
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.
At Least Intermittent (ALI)	Contain more than ephemeral flow but cannot be determined with high confidence if it is intermittent or perennial

Performance Measure	Description
PvIvE	Overall measure of accuracy. Ability of model to correctly classify between Perennial versus Intermittent versus Ephemeral. Calculated as the percent of reach-visits classified correctly (weighted by the number of visits per reach).
EvALI	Ability of model to correctly classify between Ephemeral and At Least Intermittent (I or P). Calculated as the percent of reach-visits classified correctly (weighted by the number of visits per reach).
Precision	For reaches that have multiple visits, are they consistently predicted correctly? Calculated as the proportion of visits within a reach with the most frequent classification, averaged across reaches.

Dataset	Description
Training	A subset of 80% of the total reaches that was used for model development. This subset was randomly selected, stratifying by strata (i.e., Southern, Central, Upper, and Northern), and actual streamflow duration class (i.e., perennial, intermittent, and ephemeral).
Testing	A subset of 20% of the total reaches that was used for model testing and is independent from the training reaches. This subset was randomly selected, stratifying by strata (i.e., Southern, Central, Upper, and Northern) and actual streamflow duration class (i.e., perennial, intermittent, and ephemeral).

Note: Data are divided by reach so that all visits at a single reach are included either in training or testing

Candidate Metric	Description	Type	Selected by RFE
Strata	SDAM subregions includes Central Prairie, Upper Midwest, Northern Prairie, Southern Plains. This is also used for the Northern Great Plains and Southern Great Plains.	GIS	No
Algae_score (NM)	Are Filamentous Algae and/or periphyton present at the reach? Higher scores indicate that algae were more prevalent and easier to find in the reach.	Bio (algae)	No
algdead_cover_score	Dead algal cover on the streambed within the study reach	Bio (algae)	No
algdead_noupstream_cover_score	Are algae on the streambed within the study reach likely from upstream source (i.e., dead mats deposited in downstream reach)?	Bio (algae)	No
alglive_cover_score	Live algal cover on the streambed within the study reach	Bio (algae)	No

Candidate Metric	Description	Type	Selected by RFE
alglivedead_cover_score	Visual estimate of the percent of streambed covered by live or dead algal growth	Bio (algae)	No
ai_present (PNW)	Presence/absence of aquatic invertebrate within the sample reach	Bio (aquatic inverts)	No
BMI_score (NM)	Benthic MacroInvertebrate (BMI) abundance. Higher scores indicate that BMI were more prevalent and easier to find in the reach.	Bio (aquatic inverts)	No
EPT_abundance	Abundance of mayflies, stoneflies, or caddisflies (i.e., Ephemeroptera, Plecoptera, Trichoptera, EPT)	Bio (aquatic inverts)	No
EPT_relabd	Relative abundance of EPT families	Bio (aquatic inverts)	No
EPT_reltaxa	Relative richness of EPT families	Bio (aquatic inverts)	No
EPT_taxa	Number of EPT families	Bio (aquatic inverts)	Yes
GOLD_abundance	Abundance of Gastropoda, Oligochaeta, and Diptera (GOLD)	Bio (aquatic inverts)	No
GOLD_relabd	Relative abundance of Gastropoda, Oligochaeta. and Diptera (GOLD) taxa	Bio (aquatic inverts)	No
GOLD_reltaxa	Relative richness of Gastropoda, Oligochaeta, and Diptera (GOLD) taxa	Bio (aquatic inverts)	No
GOLD_taxa	Number of Gastropoda, Oligochaeta, and Diptera (GOLD) families	Bio (aquatic inverts)	No
GOLDOCH_relabd	Relative abundance of GOLD and OCH taxa	Bio (aquatic inverts)	No
GOLDOCH_reltaxa	Relative richness of GOLD and OCH taxa	Bio (aquatic inverts)	No
mayfly_abundance	Abundance of mayflies	Bio (aquatic inverts)	No
mayfly_gt6 (PNW)	Mayfly abundance greater than six	Bio (aquatic inverts)	No
Noninsect_abundance	Abundance of non-insect invertebrate taxa	Bio (aquatic inverts)	No
Noninsect_relabund	Relative abundance of non-insect invertebrate taxa	Bio (aquatic inverts)	No
Noninsect_reltaxa	Relative richness of non-insect invertebrate taxa	Bio (aquatic inverts)	No

Candidate Metric	Description	Type	Selected by RFE
Noninsect_taxa	Richness of non-insect invertebrate taxa	Bio (aquatic inverts)	No
OCH_abundance	Abundance of Odonata, Coleoptera, and Heteroptera (OCH)	Bio (aquatic inverts)	No
OCH_relabd	Relative abundance of Odonata, Coleoptera, and Heteroptera (OCH) taxa	Bio (aquatic inverts)	No
OCH_reltaxa	Relative richness of Odonata, Coleoptera, and Heteroptera (OCH) taxa	Bio (aquatic inverts)	No
OCH_taxa	Number of Odonata, Coleoptera, and Heteroptera (OCH) families	Bio (aquatic inverts)	No
peren_present (PNW)	Presence/absence of perennial indicator invertebrate taxa within the study reach	Bio (aquatic inverts)	No
perennial_abundance	Abundance of perennial invertebrate indicator taxa	Bio (aquatic inverts)	No
perennial_live_abundance	Abundance of perennial invertebrate indicator taxa (living specimens only)	Bio (aquatic inverts)	No
perennial_taxa	Number of perennial invertebrate indicator taxa	Bio (aquatic inverts)	No
Richness	Total richness of aquatic invertebrate families	Bio (aquatic inverts)	No
TotalAbundance	Total abundance of aquatic invertebrates	Bio (aquatic inverts)	No
iofb_score (NM)	Presence/absence of iron-oxidizing bacteria and fungi.	Bio (other)	No
liverwort_cover_score	Liverwort cover on the streambed. Higher scores indicate higher liverwort cover on streambed.	Bio (other)	No
moss_cover_score	Moss cover on the streambed. Higher scores indicate higher moss cover on streambed.	Bio (other)	No
DifferencesInVegetation_score (NM)	Differences in vegetation between the riparian corridor and adjacent uplands score. Higher scores indicate a more distinct riparian corridor.	Bio (veg)	No
hydrophytes_present	Number of hydrophytic plant species (FACW or OBL) observed within the study reach channel and 1/2 channel width of the stream on either bank	Bio (veg)	No
hydrophytes_present_any (PNW)	Is the presence/absence of hydrophytes within the study reach channel and 1/2 channel width of the stream on either bank?	Bio (veg)	No
hydrophytes_present_noflag	Number of hydrophytic plant species (FACW or OBL) observed within the study reach channel and 1/2 channel width of the stream on either bank (excluding taxa with unusual distributions flagged by the field crew)	Bio (veg)	Yes
PctShading	Percent shading on the streambed.	Bio (veg)	Yes

Candidate Metric	Description	Type	Selected by RFE
ripariancorr_score (PNW)	With/without distinctive vegetation in the riparian corridor compared to surrounding upland vegetation.	Bio (veg)	No
UplandRootedPlants_score (NM)	Are upland rooted plants absent from the streambed score? Higher scores indicate fewer upland plants in the streambed.	Bio (veg)	Yes
amphib_score (PNW)	Detection of aquatic life stage(s) of amphibian(s) within the study reach.	Bio (verts)	No
Fish_score (NM)	Fish abundance score. Higher scores indicate that fish were more prevalent and easier to find in the reach.	Bio (verts)	No
fishabund_score2	When Mosquitofish are present, set to 0. Otherwise, use Fish_score (which is the abundance of fish).	Bio (verts)	No
frogvoc_score	Presence/absence of frog vocalizations	Bio (verts)	No
snake_score (PNW)	Presence/absence of aquatic snakes within the study reach	Bio (verts)	No
turt_score	Presence/absence of turtle(s) within the study reach	Bio (verts)	No
vert_score	Presence/absence of aquatic vertebrates. max(snake_score, amphib_score, turt_score, frogvoc_score)	Bio (verts)	No
vert_sumscore	Number of aquatic vertebrate types present. (Sum of snake_score, amphib_score, and turt_score)	Bio (verts)	No
vertvoc_sumscore	Sum of (snake_score, amphib_score, turt_score, frogvoc_score)	Bio (verts)	No
BankWidthMean	Mean of columns that start with 'Bankwidth'	Geom	Yes
ChannelDimensions_score (NM)	Scored channel entrenchment metric from the New Mexico protocol; higher scores indicate less entrenchment and more access to the floodplain. Higher scores indicate the channel was less confined (had higher entrenchment ratios).	Geom	Yes
erosion_score (PNW)	Presence/absence of evidence of fluvial erosion (e.g., undercut banks, scour marks, channel downcutting, channel incision) and/or deposition (e.g., bars, recent deposits) within the study reach channel?	Geom	No
floodplain_score (PNW)	Presence/absence of a true floodplain at the reach?	Geom	No
SedimentOnPlantsDebris_score (NM)	Visual estimate of the extent of evidence of sediment deposition on plants and on debris within the floodplain. Higher scores indicate that sediment deposition was more prevalent throughout the reach.	Geom	No
Sinuosity_score (NM)	Scored channel sinuosity. Higher scores indicate more sinuous channels.	Geom	Yes
Slope	Reach slope as measured with a handheld clinometer	Geom	No
slope_gt10.5 (PNW)	Straightline reach slope as measured with a handheld clinometer greater than or equal to 10.5%	Geom	No
slope_gt16 (PNW)	Straightline reach slope as measured with a handheld clinometer greater than or equal to 16%	Geom	No
SubstrateSorting_score (NM)	Visual estimate of the extent of evidence of substrate sorting within the channel. Higher scores indicate greater sorting of substrate within the channel relative to surrounding uplands.	Geom	Yes
RifflePoolSeq_score (NM)	Visual estimate of the diversity and distinctiveness of riffles, pools, and other flow-based microhabitats. Higher scores indicate more distinctive riffles, pools, and other flow habitats with clear transitions within the reach.	Geom	No
BFI	Base flow Index: estimated percentage of total flow that is attributed to groundwater discharge to streams by interpolating values from USGS stream gages	GIS	No

Candidate Metric	Description	Type	Selected by RFE
Elev_m	Watershed elevation retrieved from StreamCat database	GIS	No
MeanSnowPersistence_01	Mean snow persistence within a 1-km radius of the reach	GIS	No
MeanSnowPersistence_05	Mean snow persistence within a 5-km radius of the reach	GIS	No
MeanSnowPersistence_10	Mean snow persistence within a 10-km radius of the reach	GIS	No
ppt	Mean annual precipitation	GIS	No
ppt.m01	Mean January precipitation	GIS	No
ppt.m02	Mean February precipitation	GIS	No
ppt.m03	Mean March precipitation	GIS	No
ppt.m04	Mean April precipitation	GIS	No
ppt.m05	Mean May precipitation	GIS	No
ppt.m06	Mean June precipitation	GIS	No
ppt.m07	Mean July precipitation	GIS	No
ppt.m08	Mean August precipitation	GIS	No
ppt.m09	Mean September precipitation	GIS	No
ppt.m10	Mean October precipitation	GIS	No
ppt.m11	Mean November precipitation	GIS	No
ppt.m12	Mean December precipitation	GIS	No
tmax	Maximum annual temperature (PRISM 30-year normal)	GIS	No
tmean	Mean annual temperature (PRISM 30-year normal)	GIS	No
tmin	Minimum annual temperature (PRISM 30-year normal)	GIS	No
HydricSoils_score (NM)	Presence/absence of hydric soils within the study reach	Hydro	No
WoodyJams_number	Number of woody jams present within the study reach channel (or up to 10 m outside of the study reach). Woody jams much completely span the active channel and be in contact with the streambed. Contain at least 3 large pieces (>1 m long and >10 cm diameter). Cause sufficient blockage to disrupt flow of water or sediment under flowing conditions.	Hydro	No
IsolatedPools_number (PNW)*	Number of pools (must have surface water) with no evidence of surface water flow in or out	Hydro	No
SurfaceFlow_pct (PNW)*	Visual estimate of percentage of reach length that has flowing surface water.	Hydro	No
SurfaceSubsurfaceFlow_pct (PNW)*	Visual estimate of percentage of reach length that has flowing surface water or sub-surface (hyporheic) flow	Hydro	No
SoilMoist_MaxScore*	Soil is qualitatively assessed for moisture level (saturated, partly saturated, or dry) in three locations. This indicator uses the wettest score out of the three.	Hydro	No
SoilMoist_MeanScore*	Soil is qualitatively assessed for moisture level (saturated, partly saturated, or dry) in three locations. This indicator uses the mean moisture score observed over all three locations.	Hydro	No
springs_score (NM)*	Scored abundance of seeps and/or springs within the sample reach. Higher scores indicate larger numbers of seeps and/or springs.	Hydro	No
WaterInChannel_score (NM)*	Scored surface water flow/presence in the sample reach. Higher scores indicate channels with greater levels of surface water flow/presence.	Hydro	No

Asterisks (*) indicate hydrologic metrics that directly measure the presence of water.