Calibration of the Biological Condition Gradient (BCG) for Macroinvertebrate Assemblages in Freshwater Wadeable Streams in the Pacific Northwest Maritime Region of Oregon and Washington

***Prepared for*:**

US EPA Office of Water, Office of Science and Technology
Susan K. Jackson, Work Assignment Manager

US EPA Region 10
Lilian Herger
Lisa Kusnierz

***Prepared by*:**

Jen Stamp
Tetra Tech, Inc.
73 Main Street Room 38, Montpelier, VT

September 29, 2022

EXECUTIVE SUMMARY

The Biological Condition Gradient (BCG) model is a conceptual model describing how ecological attributes change in response to increasing levels of human-caused stress. The BCG is typically divided into six levels of biological condition along a generalized stressor-response curve, ranging from observable biological conditions found at no or low levels of stressors (level 1) to those found at high levels of stressors (level 6) (USEPA 2016). In this project, a panel of aquatic biologists from the Pacific Northwest used the conceptual BCG model as the template to develop numeric models using macroinvertebrate assemblage data from freshwater wadeable streams in the Pacific Northwest Maritime Region (PNMR) of Oregon and Washington, which encompassed five Omernik Level 3 ecoregions: Coast Range, Puget Lowland, Willamette Valley, Cascades, and North Cascades (Omernik 1987). The PNMR BCG models supersede the Puget Lowland/Willamette Valley (PL/WV) BCG model that had been calibrated during an earlier phase of work (Stamp and Gerritsen 2019). Many of the same panelists who calibrated the PNMR BCG participated in the earlier PL/WV exercise. The panel was comprised of representatives from federal agencies (USGS and USEPA ORD), state agencies (Oregon Department of Environmental Quality (ODEQ), Washington Department of Ecology (WA ECY), county agencies (King County, Snohomish County), taxonomists (Aquatic Biology Associates and Rhithron Associates Inc.) and a local expert/educator from Apolysis LLC.

Following an expert-knowledge-based approach to quantify the conceptual BCG (USEPA 2016), participants assigned sample data to BCG levels and documented their rationale for sample assignments, leading to development of narrative and then numeric decision rules. The rules developed by the panelists were compiled, tested, refined, and vetted with the panel through multiple meetings and webinars. The end products were quantitative models that predict the BCG levels of samples based on the rules developed by the panel with adjustments for gradient and elevation. The BCG model was calibrated for three classes of freshwater wadeable streams within the PNMR region: low gradient (<1%)/lower elevation (<750 m); higher gradient (≥1%)/lower elevation (<750 m); and higher gradient (≥1%)/higher elevation (≥750 m). There was good correspondence between the quantitative BCG model outputs and expert consensus scores. BCG level assignments from the models were within 0.5 units of the final expert consensus scores in 95.5% of the calibration samples and 95.4% of the confirmation samples. The BCG models can be used to supplement existing (independently developed) bioassessment tools, including the Puget Sound Lowlands Benthic Index of Biotic Integrity (B-IBI) (King County 2014; https://pugetsoundstreambenthos.org/), WA ECY B-IBI (Larson et al. 2019) and ODEQ's PREDictive Assessment Tool for Oregon (PREDATOR) (Hubler 2008), and to help inform setting protection and restoration goals.

ACKNOWLEDGEMENTS

The panelists in this effort (listed in Table 2) included representatives from federal and state agencies (Oregon Department of Environmental Quality (ODEQ), Washington Department of Ecology (WA ECY), USGS and USEPA ORD), county agencies (King County, Snohomish County), taxonomists (Aquatic Biology Associates and Rhithron Associates Inc.) and local experts and educators (Apolysis LLC). In addition, we would like to acknowledge John Pfeiffer (EcoAnalysts, Inc.), Mike Cole (Cole Ecological, Inc.), Ryan Hill (EPA ORD), Alan Herlihy (Oregon State University) and Patrick Edwards (Portland State University) for their contributions during the previous phase of work, when a BCG model was calibrated for the Puget Lowland and Willamette Valley ecoregions. We would also like to thank Gretchen Hayslip (retired), Christopher Zell, Lillian Herger and Lisa Kusnierz from US EPA Region 10 for their support throughout this project. Participants invested significant time and commitment in the process. We are grateful for their hard work and enthusiasm.

ACRONYMS

ALU	Aquatic Life Use
BCG	Biological Condition Gradient
BMP	Best Management Practices
CWA	Clean Water Act
B-IBI	Benthic Index of Biological Integrity
EPT	Ephemeroptera, Plecoptera and Trichoptera
ICI	Index of Catchment Integrity
IWI	Index of Watershed Integrity
ODEQ	Oregon Department of Environmental Quality
PL/WV	Puget Lowlands/Willamette Valley
PNMR	Pacific Northwest Maritime Region
PREDATOR	PREDictive Assessment Tool for Oregon
PSP	Puget Sound Partnership
PSSB	Puget Sound Stream Benthos
USEPA	U. S. Environmental Protection Agency
USGS	U. S. Geological Survey
WA ECY	Washington Department of Ecology
WQS	Water Quality Standards

Contents

Appendixes

Attachments

Supplemental materials

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The Biological Condition Gradient (BCG) is a conceptual model describing changes in aquatic systems with increasing levels of human disturbance (Davies and Jackson 2006, USEPA 2016, Hausmann et al. 2016, Gerritsen et al. 2017). It includes predicted changes in structural and functional characteristics of stream systems as they degrade in response to human disturbance (Figure 1). These measurable characteristics are defined as "attributes" of the biological communities and the physical habitat that reflect the condition of an aquatic ecosystem (USEPA 2016). The attributes (Table 1) include properties of the system and communities (e.g., taxa richness, structure, abundance, system functions) and organisms (e.g., tolerance, rarity, native range, physical condition). The model describes changes in the biota that have been commonly observed by aquatic scientists in different regions across the country (Davies and Jackson 2006). Ideally, assessment results can be communicated consistently regardless of location, assemblage measured and methodology. It is a useful construct, therefore, for interpreting biological indices and for comparing and reconciling regional differences in reference condition, types of indices, or even indices for different assemblages.

Since completion of the conceptual BCG framework (Davies and Jackson 2006), states have further developed and refined quantitative BCG models. In conjunction with other water quality management technical tools, state programs that have developed and applied the BCG have done so to help:

- Set scientifically defensible, ecologically-based aquatic life goals or uses (ALUs) based on existing conditions and potential for improvement;
- Determine baseline conditions and measure impacts of multiple stressors or system altering conditions (e.g., climate change) on aquatic life;
- Further the use of monitoring data for the assessment of water quality standards (WQS) and tracking changes in biological condition (e.g., documentation of incremental improvements due to controls and best management practices (BMPs));
- Identify high quality waters for protection (e.g., Tier III antidegradation);
- Communicate to stakeholders the likely impact of decisions on protection and management of aquatic resources.

This document describes the development of quantitative BCG models for three classes of freshwater wadeable streams in the Pacific Northwest Maritime Region (PNMR) of Oregon and Washington, which encompassed the following five Omernik Level 3 ecoregions: Coast Range, Puget Lowland, Willamette Valley, Cascades, and North Cascades. The PNMR BCG models supersede the Puget Lowland/Willamette Valley (PL/WV) BCG model that had been calibrated during an earlier phase of work (Stamp and Gerritsen 2019). As the panelists completed the initial PL/WV BCG model, they advocated for expanding the coverage to include the Coast Range, Cascades, and North Cascades ecoregions. The panelists strongly felt that the larger model would be more robust than the PL/WV model because of the expanded and more comprehensive dataset that would be available. Moreover, they wanted a BCG model that would cover a larger number of sites in Oregon and Washington, which would make it more useful for conducting assessments particularly for the State water quality monitoring programs. The decision was made then to expand the geographic coverage of the model, test and refine the

PL/WV model with data that covers the entire region, and revisit decision rules and classification as needed.

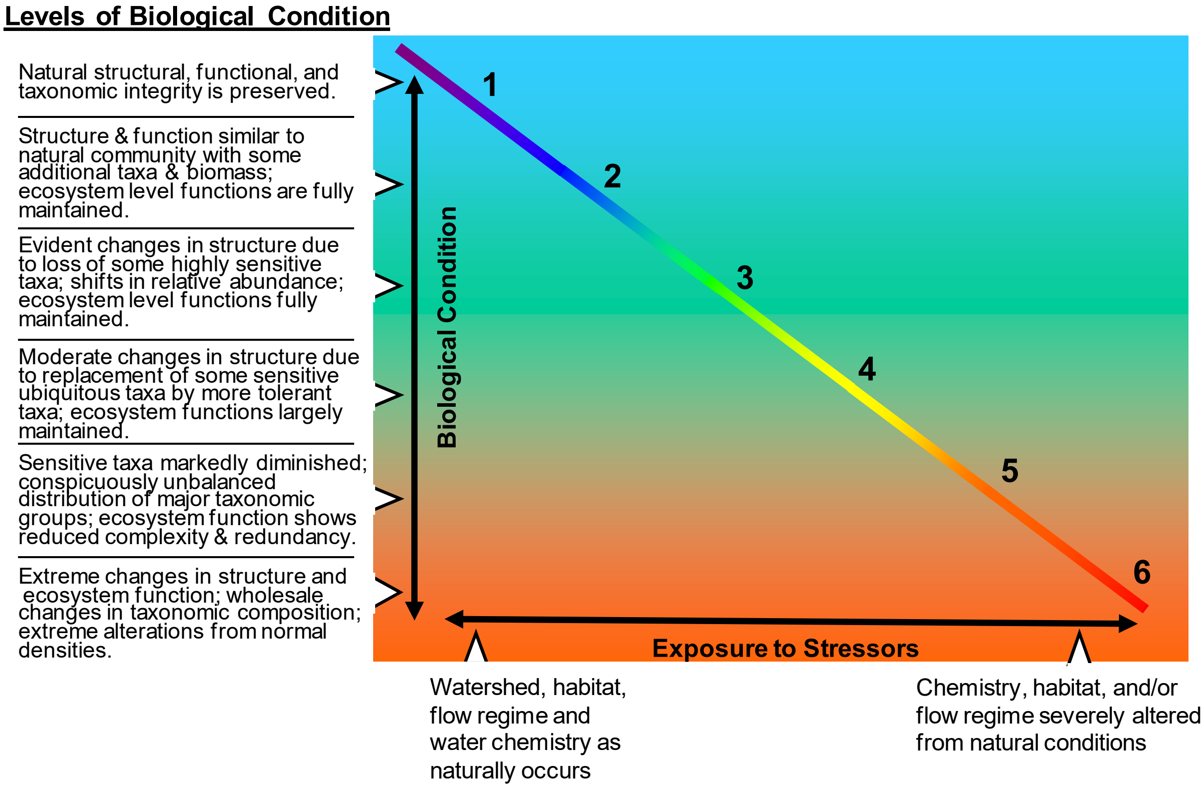


Figure 1. The Biological Condition Gradient (BCG), modified from Davies and Jackson 2006. The BCG was developed to serve as a scientific framework to synthesize expert knowledge with empirical observations and develop testable hypotheses on the response of aquatic biota to increasing levels of stress. It is intended to help support more consistent interpretations of the response of aquatic biota to anthropogenic stressors and to clearly communicate this information to the public, and it is being evaluated and piloted in several regions and states.

Table 1. Biological and other ecological attributes used to characterize the BCG. Attributes VII-X are in gray text because they were not explicitly considered for this project.

Attribute	Description
I. Historically documented, sensitive, long-lived, or regionally endemic taxa	Taxa known to have been supported according to historical, museum, or archeological records, or taxa with restricted distribution (occurring only in a locale as opposed to a region), often due to unique life history requirements (e.g., Pupfish, many Unionid mussel species).
II. Highly sensitive (typically uncommon) taxa	Taxa that are highly sensitive to pollution or anthropogenic stressors. Tend to occur in low numbers, and many taxa are specialists for habitats and food type. These are the first to disappear with disturbance or pollution (e.g., most stoneflies, Chinook salmon).
III. Intermediate sensitive and common taxa	Common taxa that are ubiquitous and abundant in relatively undisturbed conditions but are sensitive to anthropogenic stressors. They have a broader range of tolerance than Attribute II taxa and can be found at reduced density and richness in moderately disturbed sites (e.g., many mayflies, many darter fish species).
IV. Taxa of intermediate tolerance	Ubiquitous and common taxa that can be found under almost any conditions, from undisturbed to highly stressed sites. They are broadly tolerant but often decline under extreme conditions (e.g., filter-feeding caddisflies, many midges, many minnow species).
V. Highly tolerant taxa	Taxa that typically are uncommon and of low abundance in undisturbed conditions but that increase in abundance in disturbed sites. Opportunistic taxa able to exploit resources in disturbed sites. These are the last survivors (e.g., tubificid worms, Black Bullhead).
VI. Nonnative or intentionally introduced species	Any species not native to the ecosystem (e.g., Asiatic clam, zebra mussel, carp, European Brown Trout). Additionally, there are many fish native to one part of North America that have been introduced elsewhere.
VII. Organism condition	Anomalies of the organisms; indicators of individual health (e.g., deformities, lesions, tumors).
VIII. Ecosystem function	Processes performed by ecosystems, including primary and secondary production; respiration; nutrient cycling; decomposition; their proportion/dominance; and what components of the system carry the dominant functions. For example, shift of lakes and estuaries to phytoplankton production and microbial decomposition under anthropogenic eutrophication.
IX. Spatial and temporal extent of detrimental effects	The spatial and temporal extent of cumulative adverse effects of anthropogenic stressors; for example, groundwater pumping in Kansas resulting in change in fish composition from fluvial dependent to sunfish.
X. Ecosystem connectivity	Access or linkage (in space/time) to materials, locations, and conditions required for maintenance of interacting populations of aquatic life; the opposite of fragmentation. For example, levees restrict connections between flowing water and floodplain nutrient sinks (disrupt function); dams impede fish migration, spawning. Extensive burial of headwater streams leads to cumulative downstream impacts to biota through energy input disruption, habitat modification, and loss of refugia and dispersing colonists. Some taxa are considered to be indicative of connectivity, especially migratory fish such as Sturgeon, American Eel, Skipjack Herring (their presence indicates unbroken connectivity).

Source: Modified from Davies and Jackson 2006.

2 METHODS

Calibration of the BCG for a region is a collective exercise among regional biologists to develop consensus assessments of sites, and then to elicit the rules that the biologists use to assess the sites (Davies and Jackson 2006). A multistep process was followed to calibrate the BCG to freshwater wadeable streams in the PNMR region (Figure 2). The major steps (data preparation, taxon attribute assignments, scoring samples and developing narrative and quantitative rules) are described in greater detail in the ensuing sections. Panelists were already familiar with the steps in the process from the existing PL/WV model (Stamp and Gerritsen 2019). The process included assembling data, making taxon attribute assignments, assigning samples to BCG levels based on the framework shown in Figure 1, and, through a series of webinars, developing narrative and quantitative BCG rules. In addition, the BCGcalc R package (https://github.com/leppott/BCGcalc; Leppo 2018) was developed to calculate BCG model outputs. This calibration process was similar to those used in BCG development in other regions (Gerritsen et al. 2017, USEPA 2016). Assessing condition of biological communities, including all common biotic indexes, involves professional judgment, even though such judgment may be hidden in apparently objective, quantitative approaches (e.g., Steedman 1994, Borja et al. 2004, Weisberg et al. 2008). The BCG calibration process uses an explicit reliance on professional judgment and development of consensus, supported with targeted analyses.

The calibration dataset used in this second phase of the project was comprised of monitoring data from the Oregon Department of Environmental Quality (ODEQ), Washington Department of Ecology (WA ECY), King County and Snohomish County. Using an expert-knowledge based approach (USEPA 2016), a panel of aquatic scientists (Table 2) assigned sample data from the PNMR to levels of condition along a gradient of stress. The expert logic for sample assignments was documented and used to develop numeric decision rules. Through meetings and conference calls, rules were compiled, tested, and refined by the panel. The end products were quantitative BCG models predicting the biological condition level of three classes of freshwater wadeable streams within the PNMR: low gradient (<1%)/lower elevation (<750 m); higher gradient (≥1%)/lower elevation (<750 m); and higher gradient (≥1%)/higher elevation (≥750 m)[1].

[1]Source of gradient data: NHDPlusV2 flowline slope (McKay et a. 2016); source of elevation data: EPA StreamCat elevation, local catchment scale (Hill et al. 2016)

Steps in the Pacific Northwest Maritime Region BCG calibration

Calibration of the Puget Lowlands/Willamette Valley (PL/WV) BCG model (2016-2018)

↓

Assemble additional data (2019)
- Expand to include the Coast Range, Cascades, and North Cascades

↓

Reevaluate/update taxa attributes (2019)
- Run taxa tolerance analyses on expanded dataset
- Reevaluate/update BCG taxa attribute assignments
- Review/update traits (FFG, habit, voltinism, habitat)

↓

Perform classification analysis (2020)
- Revealed patterns related to gradient/elevation

↓

Assess sample worksheets via homework & webinars (2020)
- ORDEQ, WA ECY, King County & Snohomish County samples
- Calibrated by starting with best and worst samples

↓

Develop & confirm quantitative BCG models via homework & webinars (2021-2022)
- High gradient/lower elevation
- High gradient/higher elevation
- Low gradient/lower elevation

↓

Add Maritime BCG models to the BCGcalc R package

↓

Write report

Figure 2. Process diagram for calibration of the BCG in the Pacific Northwest Maritime Region.

Table 2. Pacific Northwest Maritime Region (PNMR) panelists, their affiliations, and short descriptions of their background.

Organization	Name	Background
Oregon Department of Environmental Quality	Shannon Hubler	Stream ecologist; leads ODEQ's biological monitoring and assessment group; very familiar with ODEQ data; developed a macroinvertebrate predictive model for assessing the biological condition of Oregon streams (PREDATOR)
Washington Department of Ecology	Chad Larson	Stream biologist; leads WA ECY's biological monitoring and assessment group; very familiar with WA ECY data; has developed data analysis tools for stream macroinvertebrates in Washington
King County	Kate Macneale	Stream biologist; very familiar with the Puget Sound Stream Benthos database; works on projects analyzing the effects of land use on the water quality and biotic health of streams and rivers
Snohomish County	Robert Plotnikoff	Senior Habitat Specialist; serves as technical lead for biological monitoring, quality assurance documentation and reporting; has developed numerous biological monitoring programs and assessment tools focused on streams of the Pacific Northwest
Rhithron Associates, Inc.	Sean Sullivan	Invertebrate taxonomist; holds 4 current Society for Freshwater Science (SFS) Level II certifications in Eastern and Western EPT and General Arthropods; has extensive experience in the analysis and interpretation of aquatic invertebrate data
Aquatic Biology Associates	Bob Wisseman	Invertebrate ecologist and taxonomist; certified by the SFS in Western EPT taxa; developed a widely used list of tolerance values for Pacific Northwest taxa; is a recognized expert in western North American caddisflies
Apolysis, LLC	Rick Hafele	Aquatic biologist; worked for ODEQ for over 20 years, and was manager of Oregon's biomonitoring program for much of that time; now an author and lecturer on fly fishing and stream ecology
US Geological Survey	Ian Waite	Aquatic ecologist; developed models for invertebrate metrics across nutrient/agricultural gradients from streams at regional and national scales
US Environmental Protection Agency Office of Research and Development (ORD)	Dave Peck	Aquatic ecologist; has worked on the EPA Environmental Monitoring and Assessment Program (EMAP) and the EPA National Aquatic Resource Surveys (NARS) to develop field and laboratory methods and biological indicators for rivers and streams

2.1 Data Compilation & Preparation

2.1.1 BCG Calibration Dataset

Calibration of the PNMR BCG model began with obtaining data from the Coast Range, Cascades, and North Cascades Omernik Level 3 ecoregions, which were then combined with the data from the Puget Lowland and Willamette Valley ecoregions from the previous phase of work. Data for the Klamath Mountains ecoregion in southern Oregon were also gathered but were not used for BCG model calibration because panelists were uncertain whether streams in that region were too unique to be pooled together with the other regions[2]. Sites were selected to represent a full gradient of natural and stressor conditions, based on existing classification schemes, stressor information, and biological indices. Data were then organized for supporting analyses and review by the experts.

The BCG calibration dataset consisted of 1889 routine biomonitoring samples from 930 unique freshwater wadeable stream sites (Table 3) with drainage areas ranging from 5 to 260 square kilometers (2 to 100 square miles). They ranged from valley bottoms to mountain streams, spanning wide elevation and slope gradients. Panelists assessed 173 of the 1889 samples. Table 3 shows distributions of sites and samples across Level 3 ecoregions and the four organizations (ODEQ, WA ECY, King and Snohomish Counties). Macroinvertebrate samples were collected using D-Frame kick-nets or Surber nets with 500-micrometer mesh net. WA ECY used a multi-habitat, reach-wide sampling method in which samples were collected from randomly selected locations along eight transects spanning the sampling reach (Hayslip 2007, WA ECY 2010). The other entities used a targeted-riffle method (Hayslip 2007). Sampling areas were at least 0.74 square meters (8 square feet) and had a minimum subsample size of 500 total organisms. Samples were collected between late June and mid-October, with most collections occurring during August and September. To harmonize the multiple data sets, regional taxonomists updated names of taxa that were affected by recent changes in taxonomic nomenclature over time, corrected misspellings, standardized naming schemes (e.g., spelled out 'group') and checked final taxa names for concordance with current standard taxonomic effort in the Pacific Northwest (Wisseman et al. 2015).

---

[2]The Klamath Mountains ecoregion has unique characteristics, in particular its geology, diverse topography, and endemic plant communities. The BCG models may need to be adapted for this ecoregion to account for these differences. More testing needs to be done before that determination can be made.

Table 3. Distribution of sites and samples across ecoregions and organizations in the Pacific Northwest Maritime Region BCG calibration dataset. Number of assessed samples are also shown.

Level 3 ecoregion	Organization	Full dataset		# Assessed Samples
		# Stations	# Samples	
Cascades	King County (DNRP)	8	42	2
	ODEQ	80	103	20
	WA ECY	46	83	15
Coast Range	ODEQ	337	502	34
	WA ECY	55	112	10
North Cascades	King County (DNRP)	8	26	2
	Snohomish County	9	9	3
	WA ECY	48	99	17
Puget Lowland	King County (DNRP)	142	666	32
	Snohomish County	47	55	6
	WA ECY	94	130	15
Willamette Valley	ODEQ	56	62	17
	Totals	**930**	**1889**	**173**

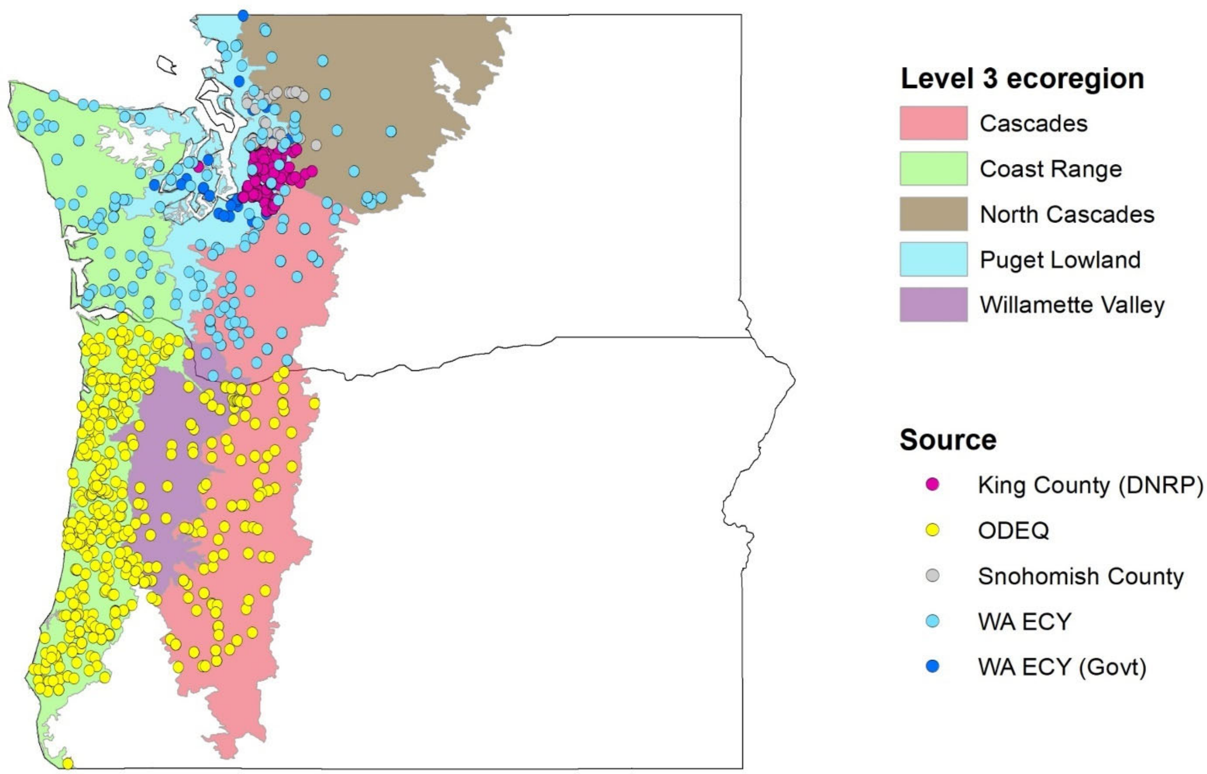


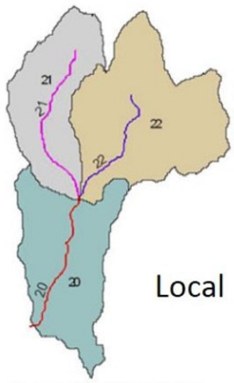
Figure 3. Sites in the Pacific Northwest Maritime Region BCG calibration dataset, color-coded by organization with an Omernik Level 3 ecoregion backdrop.

2.1.2 Taxon Attribute Categories

Panelists assigned taxa in the BCG calibration dataset to six BCG attribute categories (I – VI) that describe taxa frequency of occurrence, native range, and sensitivity or tolerance to stressors (Table 1). To help inform the attribute assignments, taxon tolerance analyses were performed to evaluate relationships between occurrence of individual taxa and variables that capture prevalent disturbance in the study area. Tolerance analyses allow for visualization of the shape of the taxon-stressor relationship across a continuous numerical scale (Yuan 2006). Results from tolerance analyses can be used to identify taxa that are likely to disappear with increased stress (i.e., sensitive) and those that can persist in altered environments (i.e., tolerant).

We characterized the taxon-stressor relationship in several ways. One was with the weighted average optima, which is commonly used for estimating the central tendency of a taxon along an environmental gradient (ter Braak and Looman 1986, Yuan 2006). It is calculated by multiplying taxon relative abundance (=the weighting factor) by the value of the disturbance variable for each sample, summing the resulting products, then dividing that by the sum of all the relative abundances (weights). In addition, we used customized R code (R Core Team 2021) to generate distribution maps, histograms, scatterplots of relative abundance and cumulative distribution functions (CDFs) for each taxon. The taxa tolerance analysis dataset differed from the PNMR BCG calibration dataset in that it included data from additional organizations to maximize sample sizes for as many taxa as possible (for the list of organizations, see Appendix A).

We ran the analyses on four anthropogenic disturbance variables that have been shown to be associated with degraded key watershed functions and have corresponding geospatial datasets that can be mapped: version 2.1 of the Indices of Watershed and Catchment Integrity (IWI & ICI, respectively) (Thornbrugh et al. 2018, Johnson et al. 2019), percent urban and percent agricultural land cover (Table 4). The disturbance variables came from the USEPA StreamCat dataset (Hill et al. 2016), which is an extensive database with hundreds of natural and anthropogenic landscape metrics for the conterminous US. The IWI and ICI provide overall measures of watershed condition and are scored on a scale of 0 (worst) to 1 (best). They are comprised of six components: hydrologic regulation, regulation of water chemistry, sediment regulation, hydrologic connectivity, temperature regulation, and habitat provision. The IWI is based on total watershed scale (which includes the local catchment plus the accumulated area of all upstream catchments) and the ICI is based on local catchment scale (which is defined as the landscape area draining to a single stream segment, excluding upstream contributions) (Figure 4). The percent urban and agricultural variables were based on the local catchment scale.



A. Local Catchments for Reaches 20, 21, and 22

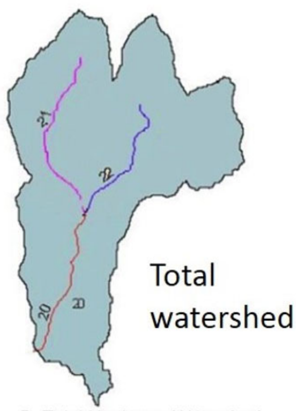
Local catchment

Definition: the landscape area draining to a single stream segment, excluding upstream contributions.

In this example, there are three local catchments (associated with unique flowline segments) –

- # 20 (green)
- # 21 (gray)
- # 22 (brown)

Each local catchment has a unique identifier (COMID or FEATUREID).



B. Total Upstream Watershed for Reach 20

Watershed-level

Definition: the local catchment plus the accumulated area of all upstream catchments

In this example there is one total watershed, comprised of the three local catchments (#20 + #21 + #22).

Figure 4. USEPA's StreamCat metrics (Hill et al. 2016) cover two spatial scales: local catchment and total watershed.

Table 4. List of disturbance variables that were used in the PNMR taxa tolerance analyses.

Metric (Abbrev)	Description	Source
Index of Watershed Integrity (IWI) version 2.1	Overall watershed condition at the total watershed scale (Figure 4), scored on a scale of 0 (worst) to 1 (best)	EPA StreamCat (Thornbrugh et al. 2018, Johnson et al. 2019)
Index of Catchment Integrity (ICI) version 2.1	Overall watershed condition at the local catchment scale (Figure 4) scored on a scale of 0 (worst) to 1 (best)	EPA StreamCat (Thornbrugh et al. 2018, Johnson et al. 2019)
% Urban land use (PctUrbCat) - local catchment scale, based on NLCD 2011	% of catchment area classified as developed, low-intensity land use (NLCD 2011 class 22) + medium-intensity land use (NLCD 2011 class 23) + high-intensity land use (NLCD 2011 class 24)	EPA StreamCat (NLCD 2011 - Homer et al. 2015)
% Agricultural land use (PctAgCat) - local catchment scale, based on NLCD 2011	% of catchment area classified as hay land use (NLCD 2011 class 81) + crop land use (NLCD 2011 class 82)	EPA StreamCat (NLCD 2011 - Homer et al. 2015)

The tolerance analysis dataset consisted of 5746 sites and captured a wide range of disturbance, with IWI scores ranging from 0.12 to 0.93 (with 0 being worst and 1 being best) (Table 5). Because of the prevalence of forested sites in the Cascade and Coastal Range ecoregions, disturbance metric values were skewed towards the better (less degraded) end of the disturbance scale, with median ICI and IWI scores of 0.84. The most disturbed streams were mostly in the Willamette River and Puget Lowland ecoregions, where there agricultural and urban areas were concentrated in lower elevation areas (Figure 5).

Table 5. Summary statistics for the disturbance variables in the PNMR taxa tolerance analysis.

Variable	# Sites with data	Min	Max	Mean	Std Dev	Percentiles						
						5th	10th	25th	50th	75th	90th	95th
IWI version 2.1	5746	0.12	0.93	0.75	0.17	0.39	0.48	0.66	0.84	0.87	0.89	0.90
ICI version 2.1	5746	0.14	0.94	0.73	0.19	0.34	0.44	0.59	0.84	0.87	0.89	0.90
% Urban land use (local catchment)	5746	0.00	99.25	13.38	24.82	0.00	0.00	0.00	0.20	11.55	58.86	77.37
% Agricultural land use (local catchment)	5746	0.00	92.52	3.99	12.67	0.00	0.00	0.00	0.00	0.25	10.06	27.90

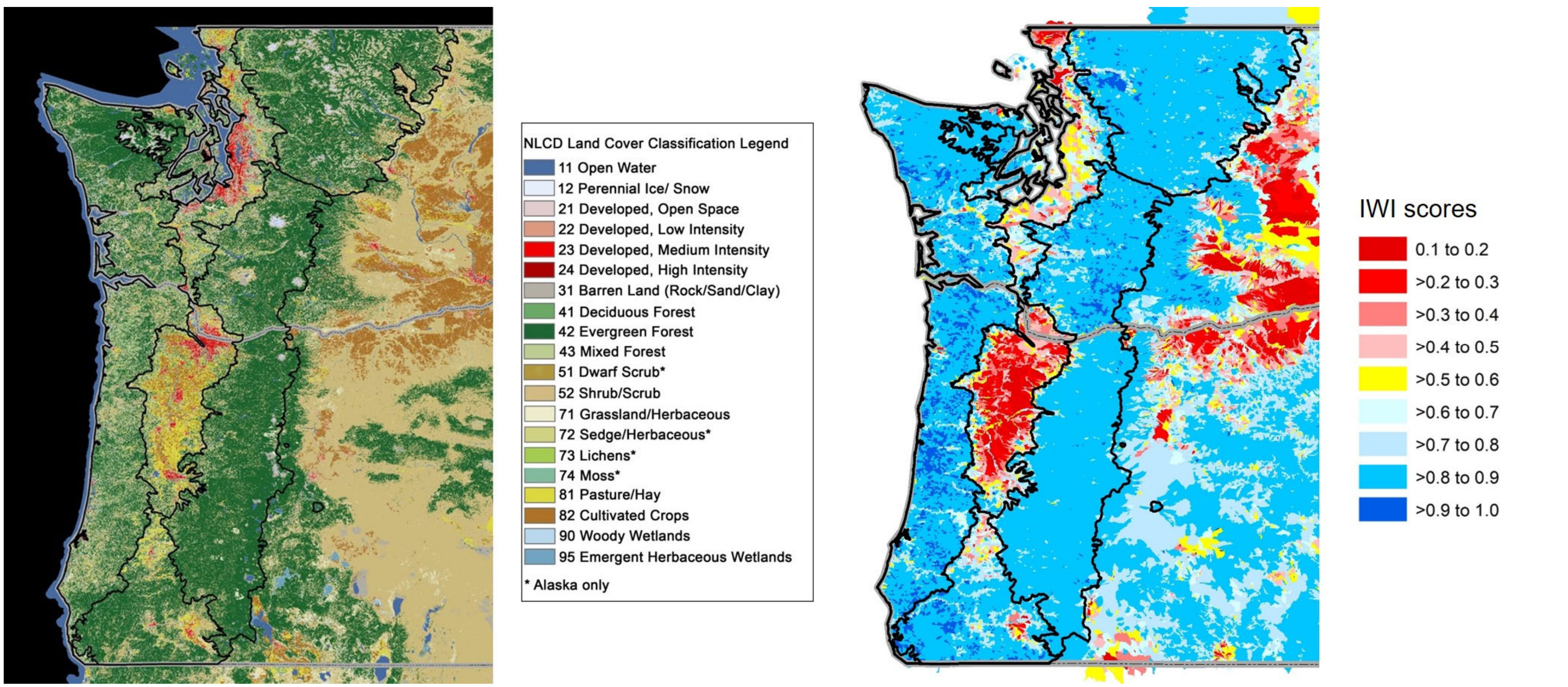


Figure 5. Land cover (NLCD 2011) (left) and IWI (version 2.1) scores in the Pacific Northwest Maritime Region Level 3 ecoregions (delineated by black lines). IWI scores range from 0 (worst, shown in red colors) to 1 (best, shown in blue colors).

Panelists assigned taxa to the following categories:

- **I –SPECIALISTS**. Historically documented, long-lived, or regionally endemic taxa
  - **Ii –** Sensitive
  - **Im –** Moderately tolerant
- **II - HIGHLY SENSITIVE**. Highly sensitive (typically uncommon) taxa
- **III – SENSITIVE.** Intermediate sensitive taxa
- **IV – INDISCRIMINATE**. Taxa of intermediate tolerance
- **V – TOLERANT**. Tolerant taxa
- **VI – EXOTIC**. Nonnative or intentionally introduced taxa
- **NA – NO ASSIGNMENT**. Insufficient information

Panelists evaluated 615 taxa. When making assignments, panelists considered multiple lines of evidence (in this order): 1) distribution across the stressor gradients as shown by the scatterplots, cumulative distribution functions and histograms (for examples, see Figure 6; for the full set of results, see the supplemental materials); 2) personal experience and best professional judgment (BPJ); 3) for taxa that had low sample size in the study area but were more common in California, taxon-response plots from an independent BCG project in California (Paul et al. 2020); and 4) weighted average optima and tolerance values. If they felt there was insufficient information, they did not make an assignment (NA). NA taxa typically had coarse-levels of taxonomic resolution (e.g., Order-level), occurred in very low numbers or were typically not used in freshwater assessments (e.g., terrestrial, estuarine). Panelists did not consider tolerance analysis results for taxa that occurred in fewer than 10 samples due to difficulties with interpreting patterns for taxa with so few data points. Panelists also did a thorough review of the functional feeding group (FFG), habit and voltinism (number of broods per year) trait assignments and assigned a 'noteworthy' taxa designation to taxa they felt were indicators of exemplary biodiversity or habitat.

Box # 1 - Clarification of Biological Condition Gradient (BCG) Terminology

Sometimes people mix up taxon attributes and BCG levels, since both share similar scales (typically 1 to 6), and there is often strong correspondence between taxon attributes and BCG levels (e.g., typically there are high proportions of Attribute II+III taxa in BCG level 2 & 3 samples). The following bullets are intended to help clarify the difference:

BCG Taxon Attributes: Six categories (I – VI) that describe taxa frequency of occurrence, native range, and sensitivity or tolerance to stressors (Table 1). Experts assign attributes to taxa before assessing samples.

BCG Levels: Six categories (1 – 6) that describe biological conditions as in Figure1. Experts rate, score, or vote to assign BCG levels to samples.

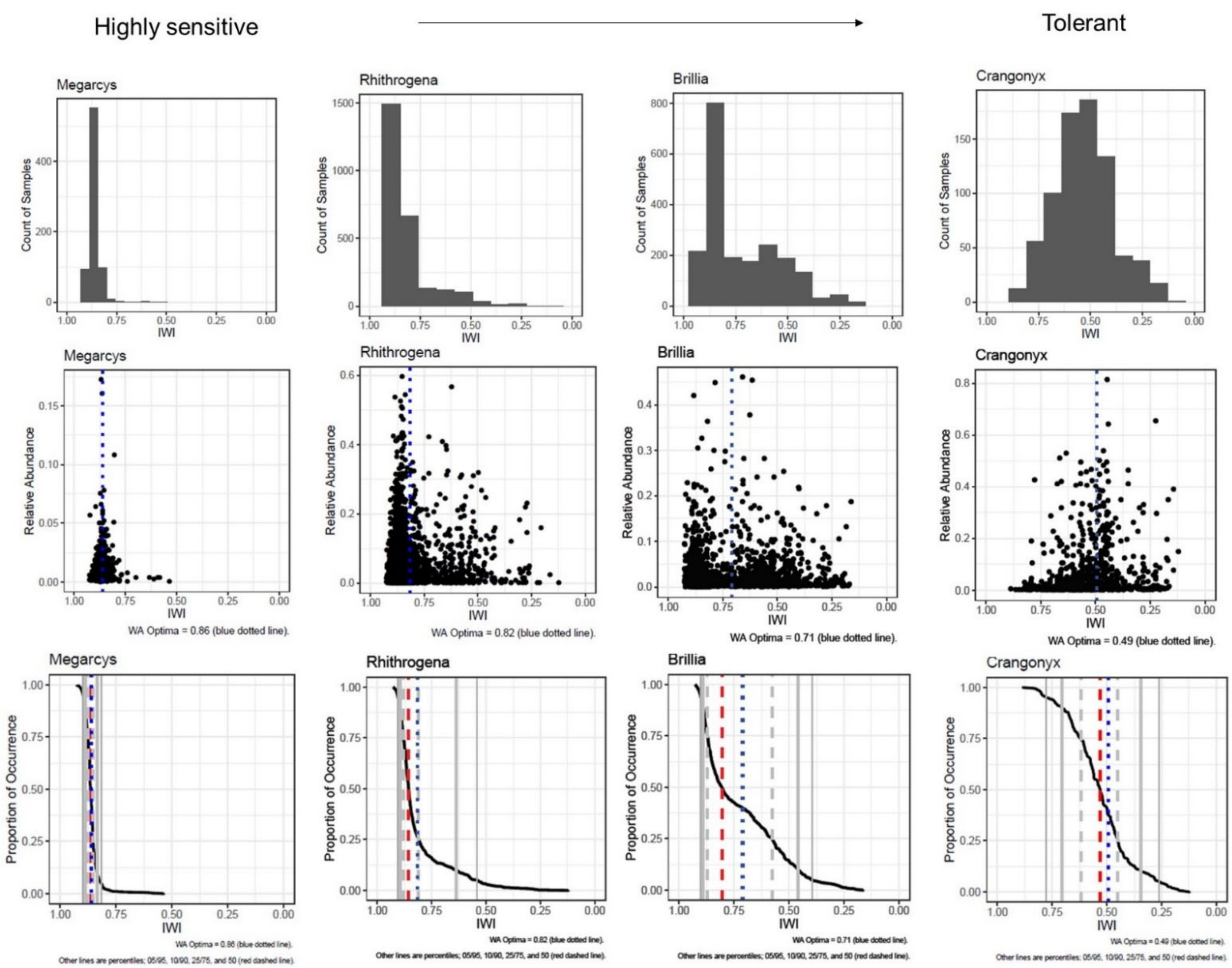


Figure 6. Patterns seen in BCG attribute II (highly sensitive) through V (most tolerant) taxa in the IWI histograms (top), relative abundance scatterplots (middle) and cumulative distribution function plots (bottom). The IWI scoring scale ranges from 0 (worst) to 1 (best). The x-axes have fixed scales based on the minimum and maximum values represented in the dataset. The histograms have 10 equal interval 'bins' (based on min/max in the dataset). The vertical blue dashed line represents the weighted average optima.

2.1.3 Metric calculations

Biological metrics are the basis of the BCG rules. Over a hundred candidate metrics were calculated and tested for replication of expert-based narrative decision rules. Prior to calculating the metrics, taxonomic nomenclature issues were resolved, and samples were randomly subsampled to 600 total individuals where needed. Six hundred total individuals was chosen as the upper limit because a target of ± 20% is in keeping with data quality objectives for most subsampling routines (Barbour et. al. 1999). The subsampling made the richness metrics more comparable across the 500-count samples, since differences in target numbers of organisms (e.g., 500-count vs. 300-count) affects estimates of taxonomic richness (the larger the subsample size, the higher the richness) (Gotelli and Graves 1996).

To account for differences in levels of taxonomic resolution across data sources, water mites were collapsed to the Order-level and worms were taken to subclass (Oligochaeta). In addition, 'non-target' taxa that are not included in bioassessment indices in Washington and Oregon (e.g., Hemiptera; Diptera: Culicidae; etc.) were excluded from all metric calculations. For richness metrics, potentially redundant taxa (also referred to as non-distinct or ambiguous taxa) were not counted. The redundant taxa designations, random subsampling and metric calculations were done with the BioMonTools R package (Leppo 2020; https://github.com/leppott/BioMonTools). Appendix B contains the list of candidate metrics that were considered during calibration of the PNMR BCG model.

2.1.4 Classification

Development of the three BCG classes (low gradient (<1%)/lower elevation (<750 m); higher gradient (≥1%)/lower elevation (<750 m); and higher gradient (≥1%)/higher elevation (≥750 m)) was an iterative process. We started with the low (<1%) and high gradient (≥1%) classification scheme that was used for the PL/WV BCG model (based on NHDPlusv2 flowline slope; McKay et al. 2016). Gradient was important because high gradient, hard-bottom streams naturally support slightly more sensitive taxa and slightly fewer tolerant taxa than the low gradient, soft-bottom streams (Stamp and Gerritsen 2019). In the expanded PNMR dataset, panelists felt it was important to consider elevation as well. In particular, they wanted to explore potential differences in macroinvertebrate communities in forested, highland streams that have cold or very cold water and are subject to more frequent disturbance than lower elevation sites.

After compiling the larger PNMR dataset, we performed two types of exploratory analyses to evaluate whether a new classification scheme was needed for the expanded study area. First, we generated box plots and evaluated distributions of metrics values across potential classes. Next, we created ordination plots using Non-metric Multidimensional Scaling (NMS) to look for patterns in taxonomic composition. The analyses were performed on 340 'reference' samples from the BCG calibration dataset, which were identified based on the following criteria: IWI and ICI scores ≥ 0.85; and BCG level 2 or 3 scores based on outputs from the existing PL/WV BCG model. Results from the exploratory analyses can be found in Appendix C.

Patterns related to gradient and elevation were evident in the sensitive/highly sensitive taxa metrics (BCG attribute Ii, II, III), with the lowest median metric values occurring in low gradient samples and the highest values occurring in high gradient/high elevation samples (Appendix C).

The high gradient/high elevation samples also had slightly lower total taxa richness than lower elevation sites. Gradient/elevation also captured patterns related to temperature, latitude and longitude. Based on these results, we first tried a gradient/elevation classification scheme that used the same gradient threshold (1%) as the PL/WV BCG model and an elevation threshold of 600-meters. During the sample assessment process, expert panelists decided to increase the elevation threshold to 750-meters because they felt the higher threshold more clearly differentiated between macroinvertebrate communities in transitional foothill streams versus forested highland streams. This threshold will be further evaluated as practitioners gain more experience with the BCG model in coming years. There were also some low gradient streams above 750-meters that could have potentially comprised a fourth class but there were too few sites in this category to calibrate a high elevation-low gradient BCG model. This can be revisited later if more sites in that class become available. Table 6 shows the distribution of samples across BCG classes, Level 3 ecoregions and organizations and Figure 7 shows the spatial distribution of the PNMR BCG classes across the study area.

While assessing high gradient-high elevation samples, we initially included streams in xeric areas on the east side of the Cascades in Washington, which include the Okanogan Pine/Fir Hills Level 4 ecoregion. This area has low mean annual precipitation (<650 mm based on PRISM 1981-2010 local catchment scale data). Other areas with low precipitation include the Olympic Rainshadow and southeastern Oregon (Figure 7). Macroinvertebrate communities in the xeric areas on the east side of the Cascades tended to receive worse BCG scores, likely due to natural differences such as warmer stream temperatures. There were too few xeric sites in our dataset to explore this further, so we restricted samples in the confirmation dataset to those with ≥ 650 mm mean annual precipitation. If the BCG model is applied to xeric streams, results should be interpreted with caution.

As part of the classification analysis, we also examined differences in targeted-riffle versus reachwide, multihabitat samples. This had also been done previously during calibration of the PL/WV BCG model. Overall, we did not find clear-cut, consistent differences between targeted-riffle versus reachwide, multihabitat samples. This was in keeping with the methods comparability studies by Gerth and Herlihy (2006) and Rehn et al. (2007), who found that targeted riffle and reachwide samples were generally interchangeable when used for ambient biomonitoring assessments. For purposes of the BCG, panelists felt comfortable moving ahead with calibrating a BCG model that could be applied to both types of samples.

Table 6. Distribution of samples in the Pacific Northwest Maritime Region dataset across BCG classes, ecoregions and organizations. Number of unique stations are also shown (some sites were sampled multiple times).

Level 3 ecoregion	Organization	Number of samples (unique sites)		
		LoGrad-LoElev	HiGrad-LoElev	HiGrad-HiElev
Cascades	ODEQ	3 (3)	36 (25)	64 (52)
	WA ECY	8 (5)	27 (20)	48 (21)
	King County (DNRP)	4 (1)	38 (7)	--
	Snohomish County	--	--	--
Coast Range	ODEQ	146 (99)	343 (227)	13 (11)
	WA ECY	43 (23)	66 (31)	3 (1)
	King County (DNRP)	--	--	--
	Snohomish County	--	--	--
North Cascades	ODEQ	--	--	--
	WA ECY	--	8 (6)	91 (42)
	King County (DNRP)	7 (2)	11 (3)	8 (3)
	Snohomish County	2 (2)	5 (5)	2 (2)
Puget Lowland	ODEQ	--	--	--
	WA ECY	52 (45)	76 (48)	2 (1)
	King County (DNRP)	290 (65)	376 (77)	--
	Snohomish County	41 (34)	14 (13)	--
Willamette Valley	ODEQ	27 (25)	35 (31)	--
	WA ECY	2 (2)	--	--
	King County (DNRP)	--	--	--
	Snohomish County	--	--	--
	Totals	**625 (306)**	**1035 (493)**	**231 (133)**

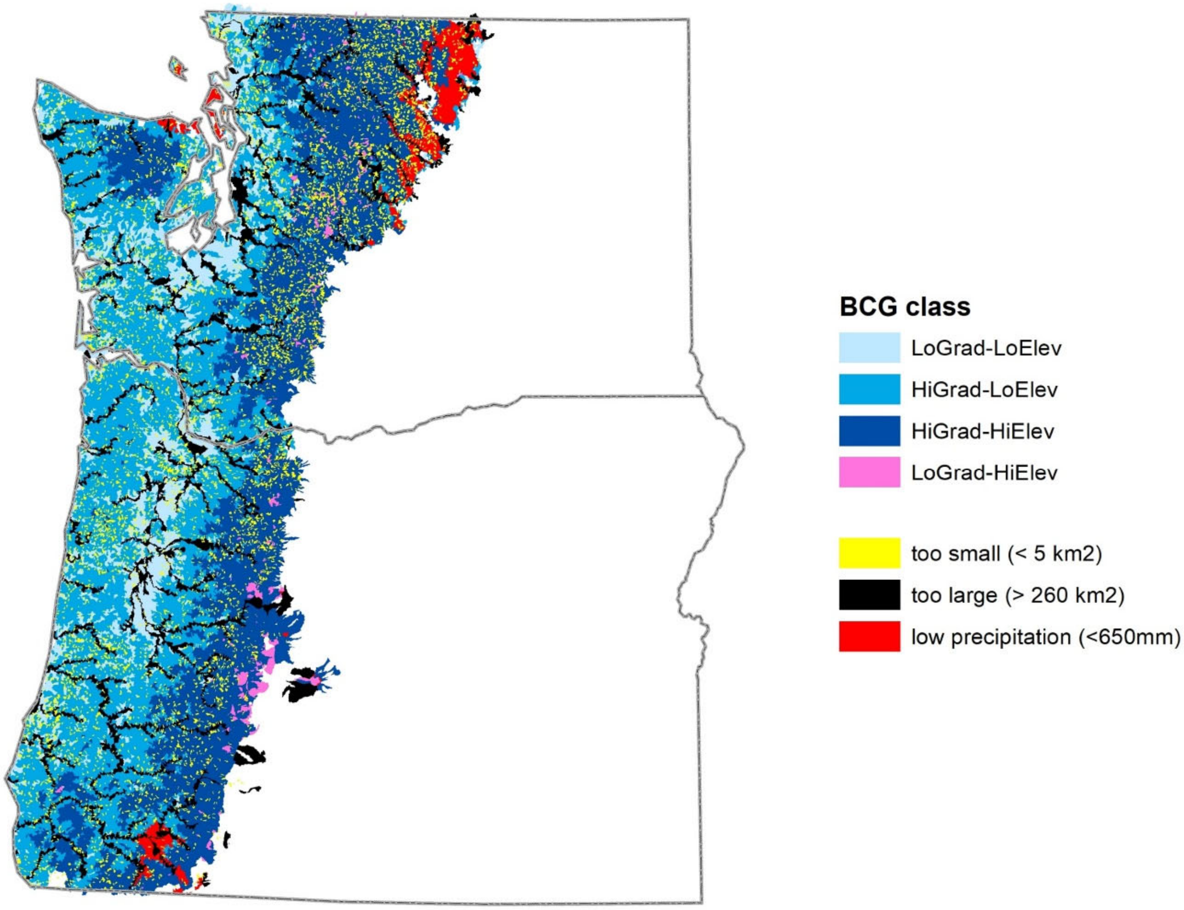


Figure 7. The Pacific Northwest Maritime Region BCG model was calibrated for three classes of freshwater wadeable streams: low gradient (<1%)/lower elevation (<750 m) (LoGrad-LoElev); higher gradient (≥1%)/lower elevation (<750 m) (HiGrad-LoElev); and higher gradient (≥1%)/higher elevation (≥750 m) (HiGrad-HiElev). NHDPlusV2 local catchments are color-coded by BCG class. Catchments with drainage areas and annual precipitation amounts outside the bounds of the of BCG calibration dataset are color-coded in yellow (<5 km2), black (>260 km2) and red (<650 mm).

2.1.5 Sample worksheets

We selected a subset of samples from the calibration dataset for panelists to assess, based primarily on the following considerations:

- **Stressor gradient**: samples spanned the full gradient of stressor conditions
- **Gradient/elevation class**: at least 30 samples from each of the three BCG classes were assessed during calibration
- **Organization**: adequate representation across the four organizations (ODEQ, WA ECY, King County, Snohomish County)
- **Unusual characteristics**: attempts were made to avoid samples with the following characteristics:
    - o Raw counts < 450 total individuals in 500-count samples

- Collected outside the normal June-October sampling period
- Collected with lower-than-normal sampling effort (sampling area should be ≥ 8 ft^2)
- Level of taxonomic resolution of Chironomidae coarser than genus-level
- Very small (< 5 km^2 (2 mi^2) drainage area) or larger than mid-order (> 260 km^2 (100 mi^2) drainage area)
- Low precipitation (mean local catchment-scale precipitation (PRISM 1981-2010) < 650 mm)
- Brackish water organisms present
- Dominated by one or two taxa

Data for the selected samples were organized into worksheets that contained lists of taxa, taxa abundances, taxa BCG attribute categories (Section 2.1.2), biological metrics, and limited site information, such as watershed area, gradient and ecoregion. Information on station identifiers, waterbody names or stressor information were hidden, so as not to bias the panelists' assessments (see example worksheet in Attachment A).

2.2 BCG Scoring

The BCG provides a common language to interpret and communicate current ecological conditions relative to baseline conditions that are anchored in BCG level 1 ("as naturally occurs"). In this way, the scoring scale is intended to be universal (USEPA 2016, Davies and Jackson 2006), but descriptions of communities, taxa, and their responses to the anthropogenic stress gradient are calibrated to be specific to the conditions and communities found in the study area (in this case, the PNMR).

To score samples, panelists went through sample worksheets like the one shown in Attachment A. The panelists examined biological data from individual samples and assigned them to BCG levels 1 through 6 (Figure 1). In some cases, panelists assigned pluses and minuses to the BCG levels to indicate somewhat better or worse condition, respectively (much as school letter grades can be modified, e.g., B+). Therefore, each BCG level had three possible scores (e.g., 3-, 3, or 3+). As an example, 3+ is level 3 tending towards level 2, whereas 3- is level 3 tending towards level 4. After each panelist assigned a BCG level to a given sample, the panelist consensus was determined by calculating the median of all the panelists' BCG level assignments. As the panel assigned samples to BCG levels, they also reported the critical information and criteria they used to make their decisions (see example in Table 7). These form preliminary, narrative rules that explain how panel members make decisions. For example, "For BCG level 2, sensitive (Attribute Ii+II+III) taxa must make up at least half of all taxa in a sample." The intent is to achieve consensus and to identify rationale that participants are using to make their assignments. Over the course of the BCG calibration and confirmation process, participants scored 173 total samples. Most assessments were done independently as 'homework.' If they were done together as a group, the facilitator asked each panelist for his or her BCG level assignment in a random or rotating sequence. After the whole group provided their scores, the facilitator asked each panelist for their rationale.

Table 7. Example of panelist reasons, which are used to develop quantitative decision rules.

Participant	Score	Reasons
Panelist 1	3	Solid 3 because of diversity; EPT could be greater but nice mix; Hydropsyche not overwhelming
Panelist 2	4	Expect more EPT taxa
Panelist 3	4	Sensitive EPT could be greater, but function appears to be maintained (diverse FFG)
Panelist 4	4	High filter feeders, some nutrient issues
Panelist 5	3-	Don't like the dominance of *Hydropsyche*
Panelist 6	3-	Would like to see more EPT taxa; don't like the dominance of *Hydropsyche*
Panelist 7	3-	Like the Attribute II taxa, but many are singletons
Panelist 8	3-	High percentage of Attribute IV taxa; ok but not great taxa richness
Panelist 9	4+	Simuliidae, *Baetis tricaudatus* and *Hydropsyche* bump this to a 4 vs. 3

2.2.1 BCG level 1

One of the steps in the BCG calibration process is to develop a current state-of-knowledge description of the biological assemblage of water bodies under pre-development, undisturbed condition to serve as a fixed, historic baseline (the level 1 prototype) (USEPA 2016). The problem that exists in nearly all regions of the country, however, is that completely natural undisturbed stream habitat rarely exists due to the level of human use and development over the past two to three hundred years. This is the case in the PNMR, where watersheds have undergone widespread and significant alterations since early European settlement times (circa1850), shifting from a hunting and trapping-based economy to a modern economy with urban and agricultural land use, as well as logging.

With the BCG, even if natural undisturbed conditions (BCG level 1) no longer exist and are no longer achievable, it is important to acknowledge and describe them to the best of our ability so that we can clearly understand and communicate how significantly biological communities have been altered (USEPA 2016). This understanding provides context for interpreting current conditions that may be significantly altered by human disturbance. This understanding, especially of any hydrological, physical or chemical processes that may persist, can also inform selection of sustainable restoration targets for degraded waters. Though narrative, the BCG level 1 description completes the response-stress gradient by providing context for understanding current conditions relative to departure, or alteration from, naturally occurring conditions. If there are no BCG level 1 sites available, then this description may be based on historical records and documents, ecology and autecology, data from sites close to natural, and whatever other sources can be found.

BCG level 1 – Puget Lowlands/Willamette Valley

During calibration of the PL/WV BCG model, panelists concluded that BCG level 1 conditions do not exist in the PL/WV (Stamp and Gerritsen 2019). Consequently, a subset of panelists researched and described, in as much detail as possible, narrative biological expectations for BCG level 1 and changes to these conditions over time due to alterations in the landscape and levels of stressors (e.g., increase in nutrients, sediments and temperature in streams) (Hafele et al. 2022). Their write-up, which is included in Appendix D, includes information for each ecoregion on historical landscape condition (pre-European settlement from the early 1800s onward) and how the physical template/stressor backdrop has changed over time with respect to land use and land cover, hydrology and water quality. Regarding the biology, they found long-term data on anadromous fish runs, but historic data were not available for macroinvertebrates. Sources of information included historical records and documents, ecology and autecology, and data from sites closest to natural (BCG level 2).

Through an iterative process (refined with research and dialogue among experts), the panelists developed the following BCG level 1 conceptual definition for perennial, medium to high gradient, small to mid-order (5 to 260 km^2) freshwater wadeable streams:

> *Streams with high habitat complexity; natural disturbance regimes to refresh micro-habitats; year-round flow without anthropogenic impacts to hydrology, temperature, or water quality; water often dominated by cool-cold flow from springs, groundwater accretion, and/or natural runoff; high resilience to disturbance including drought and flood extremes; intact riparian and aquatic habitat that provide longitudinal and lateral connectivity; exemplary biological diversity with high taxa richness of rheophilic[3], lotic-depositional, and micro-habitat specialist macroinvertebrates; non-native invasive species absent; biotic community supports all ecosystem functions*.

Table 8 contains more detailed descriptions of the fundamental characteristics mentioned in this definition based on a literature search and synthesis (Hafele et al. 2022). During development of the PL/WV BCG model, one of the panelists also developed a draft Biodiversity Index to identify sites or watersheds with exemplary biodiversity using current benthic macroinvertebrate data (with the term "exemplary" meaning "serving as a desirable model; representing the best of its kind") (Wisseman 2020). Application is limited to moderate to high gradient, low to mid-elevation streams in the PNMR with coarse mineral substrates and extensive erosional habitat. The Biodiversity Index can help identify sites with high habitat complexity and resilience to disturbance that support macroinvertebrate communities with high and/or unique taxa richness representative of a BCG level 1 stream. During preliminary evaluations, there was good correspondence between Biodiversity Index scores and BCG model outputs (index scores followed the expected pattern, with highest median index scores occurring in BCG level 2 samples and lowest in BCG level 5-6 samples). Practitioners can use both tools together to help to identify sites that harbor exemplary biodiversity and help inform prioritization of sites for protection.

---

[3] Rheophilic taxa prefer to live in fast moving water

Table 8. Fundamental characteristics of BCG level 1 perennial, medium to high gradient, small to mid-order (5 to 260 km2) freshwater wadeable streams in the PL/WV ecoregions (Hafele et al. 2022; Appendix D).

Fundamental Characteristics	Description
Stream channel	Channel connected to hyporheos and flood plain including wetlands, beaver ponds, etc.; diverse habitats present (e.g. braided channels, side channels, debris jams, mixture of steps and pools consistent with stream gradient); wood debris typically present and may be abundant; quality habitat and refugia persists during periods of both low and high stream-flows.
Riparian & watershed	Riparian zone supports intact community of overstory, understory and groundcover plants (including a mixture of mature conifer and hardwood trees with a diverse age structure in forested watersheds); upper watershed vegetation intact, supporting delivery of water of high chemical and thermal quality to lower reaches.
Hydrologic regime	Hydrologic regime natural, without alteration from dams and/or irrigation withdrawals or return flow; cool-cold water common from springs, groundwater accretion, and/or natural runoff; perennial surface or subsurface flow. Re-charge in the watershed sustains flow, especially during years of extreme drought. Perennial surface water in some portion of watersheds maintain endemic taxa that serve as recolonization sources sustaining high biodiversity at select locations. These locations promote resiliency in stream reaches that are periodically de-watered.
Disturbance regime and resilience	Natural seasonal range of high and low stream-flows present, which enhances and maintains channel and habitat complexity. Natural sediment transport based on local geology, soils and stream gradient. High resilience (ability to recover from disturbance) to natural and anthropogenic watershed stressors (Flotemersch et al. 2016). Watershed integrity maintains disturbance levels within ranges tolerable by endemic taxa and promotes connectivity for purpose of recolonization.
Biodiversity	Benthic macroinvertebrate community typically with high taxa richness, including many micro-habitat specialist taxa and taxa sensitive to human disturbance. Habitat complexity results in diversity of both rheophilic and lotic-depositional taxa. Non-native, invasive taxa not present.
Ecosystem function	Watershed supports full range of ecological processes and functions essential to maintaining high biodiversity provided by a minimally disturbed ecosystem. Food web, nutrient and energy flow linkages between aquatic and terrestrial environments fully supported.

BCG level 1 – Pacific Maritime Northwest Region

During calibration of the PNMR BCG model, the historic literature review was expanded to include the Coast Range and Cascades Level 3 ecoregions (Hafele et al. 2022). The write-up is included in Appendix D. Even though the Coast Range and Cascades have less disturbance than the PL/WV ecoregions, panelists concluded that BCG level 1 was still not represented in the PNMR BCG calibration dataset. Compared to early European settlement times (circa1850), streams in the Coast Range and Cascades have still had notable impacts, in particular from logging (which can contribute to increased sediment inputs and stream warming). In addition, impacts from climate change are becoming more noticeable, even in the most pristine high elevation streams, with warming, more extreme high and low flow events, and increased risk of wildfires (May et al. 2018).

For the PNMR, we used the same narrative description of BCG level 1 that had been developed for the PL/WV BCG model, which panelists felt still works well for the high gradient/lower elevation class. As practitioners gain more experience with the PNMR BCG models, they may decide to make some adjustments to the BCG level 1 descriptions for low gradient/low elevation and high gradient/higher elevation streams. For example, through targeted sampling, they may be able to find some better examples of high-quality low gradient/low elevation streams, which are difficult to find since low elevation areas have more urban and agricultural development (Hafele et al. 2022). Having better information on substrate (soft- vs. hard-bottom) and biodiversity in edge habitats (which are not captured well by the targeted riffle and reachwide sampling methods) would also be helpful. The narrative description for high gradient/higher elevation samples may need slight adjustments as well. Information from pristine (or nearly so) streams in National Parks and wilderness areas suggest forested, highland streams have only moderate taxa richness (e.g., glacial meltwater streams often support few species). Macroinvertebrate communities are subjected to high levels of natural disturbance, as well as cold or very cold thermal regimes, which restricts occurrence of certain taxa, as does past glaciation. Only about 12,000 years have lapsed since glaciers retreated and streams formed in most of this high elevation zone, leaving little time for an endemic fauna to develop.

2.3 BCG Rule Development

2.3.1 Background on Fuzzy Set Theory

BCG level descriptions in the conceptual model are intentionally general (e.g., “reduced richness”) to reflect shared patterns of biological change to human disturbance across waterbody types and ecological regions. To allow for consistent assignments of sites to levels, it is necessary to formalize the expert knowledge by codifying level descriptions into a set of rules that replicate the decision criteria of the expert panel (e.g., Droesen 1996). People tend to use strength of evidence in defining decision rules, and in allowing some deviation from their ideal for any individual attributes, as long as most attributes are in or near the desired range. For example, the definitions of “high,” “moderate,” “low,” etc., are qualitative (but ordinal) and can be interpreted and measured to mean different things. An important step in the BCG process is development of expert consensus defining these, or other, general terms and documenting the expert logic that is the basis for the decisions. The decision rules preserve the collective professional judgment of the expert group and set the stage for the development of models that can reliably assign sites to levels without having to reconvene the same group. In essence, the rules and the models capture the panel’s collective decision criteria.

Quantification of rules allows users to consistently assess sites according to the same rules used by the expert panel, and allows a computer algorithm, or other persons, to obtain the same level assignments as the panel. BCG quantitative models have been constructed for over ten different regions based on modern mathematical set theory and logic (called “fuzzy set theory”). Fuzzy set theory is directly applicable to environmental assessment and has been used extensively in engineering applications worldwide (e.g., Demicco and Klir 2004) and environmental applications have been explored in Europe and Asia (e.g., Castella and Speight 1996, Ibelings et al. 2003).

Mathematical fuzzy set theory allows degrees of membership in sets, and degrees of truth in logic, compared to all-or-nothing in classical set theory and logic. Membership of an object in a set is defined by its membership function, a function that varies between 0 and 1. To illustrate, we compare how classical set theory and fuzzy set theory treat the common classification of sediment, where sand is defined as particles less than or equal to 2.0 mm diameter, and gravel is greater than 2.0 mm (Demicco and Klir 2004). In classical “crisp” set theory, a particle with diameter of 1.999 mm is classified as “sand”, and one with 2.001 mm diameter is classified as “gravel.” In fuzzy set theory, both particles have nearly equal membership (approximately 0.5) in both classes (Demicco and Klir 2004). Very small measurement error in particle diameter greatly increases the uncertainty of classification in classical set theory, but not in fuzzy set theory (Demicco and Klir 2004). Demicco and Klir (2004) proposed four reasons why fuzzy sets and fuzzy logic enhance scientific methodology:

- Fuzzy set theory has capability to deal with “irreducible measurement uncertainty,” as in the sand/gravel example above.
- Fuzzy set theory captures vagueness of linguistic terms, such as “many,” “large” or “few.”
- Fuzzy set theory and logic can be used to manage complexity and computational costs of control and decision systems.
- Fuzzy set theory attempts to model human reasoning and decision-making, which is critically important for defining thresholds and decision levels for environmental management.

The BCG models use mathematical fuzzy logic to replicate human reasoning. Each linguistic variable (e.g., “high taxa richness”) is defined quantitatively as a fuzzy set (e.g., Klir 2004). Lower and upper (“fuzzy set”) bounds are set for each metric based on distributions of biological metrics across BCG levels. Each metric receives a membership value ranging from 0 to 1, depending on where the value falls in relation to the bounds. The rule threshold falls in the middle of these bounds. Metric values that are less than or equal to the lower bound receive a membership value of 0, while metric values that are greater than or equal to the upper bound receive a membership value of 1. In the example shown in Figure 8, the example rule for total taxa richness is ≥ 20 (15-25) (the lower bound is 15 and the upper bound is 25), which means –

- If there are 15 or fewer total taxa in the sample, the metric membership value is 0.
- If there are 25 or more total taxa in the sample, the metric membership value is 1.
- If the number of total taxa falls within the lower and upper bounds, the metric membership value will range from 0 to 1 (e.g., if there are 20 total taxa, the membership value will be 0.5; if there are 17 total taxa, the membership value will be 0.2; if there are 23 total taxa, the membership value will be 0.8).

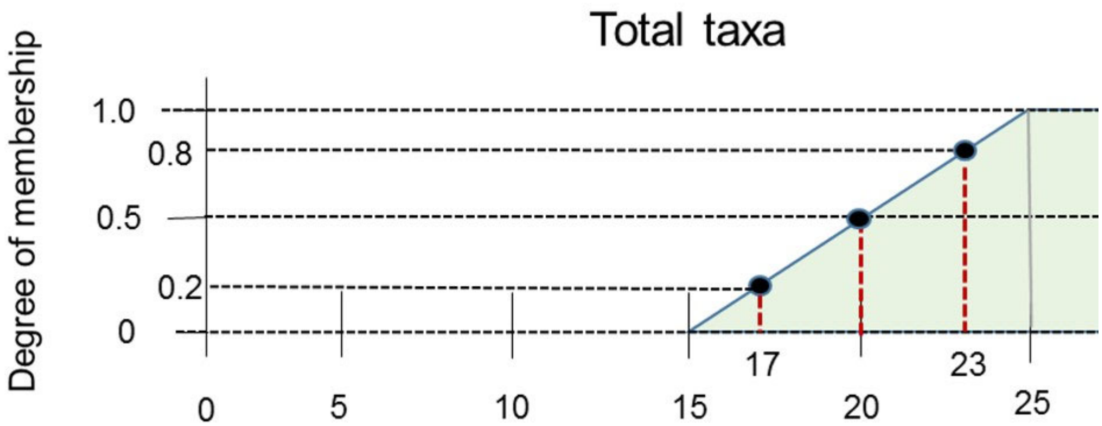


Figure 8. Illustration of the lower and upper ("fuzzy set") bounds for an example metric, total taxa richness. Each metric receives a membership value ranging from 0 to 1, depending on where the value falls in relation to the bounds. In this example, the BCG rule for total taxa richness is ≥ 20 (15-25) (the lower bound is 15 and the upper bound is 25). The black dots show examples of metric membership values assigned to different metric values (e.g., if there are 20 total taxa, the metric membership value will be 0.5; if there are 17 total taxa, the membership value will be 0.2; if there are 23 total taxa, the membership value will be 0.8).

BCG rules for a given level are typically comprised of multiple metrics (which are considered in combination). Together the rules for each BCG level work as a cascade from BCG level 1 to level 6, such that a sample is first tested against the level 1 rules; if the combined rule fails, then the level fails, and the assessment moves down to level 2, and so on (Figure 9). The BCG model evaluates metric membership values for all the metrics included in the rules for a given BCG level and considers the combination rules to derive the membership level for the sample. There are several different types of combination rules. If rules for two metrics are combined with an "AND" operator, then both metrics must meet the thresholds for a given BCG level (as a hypothetical example, let's say there are two rules for BCG level 3: total taxa richness ≥ 20 AND percent sensitive taxa ≥ 10%; both conditions must be met in order for the sample to be assigned to BCG level 2). If the two rules are combined with an "OR" operator (referred to as an 'alternate' rule), then *either* can be true for a sample to meet the requirements (both conditions are not necessary). Another option is having a 'best xx of xx' rule, where not all of the metrics in a group need to be met. For example, the PNMR BCG models have 'best two of three' rules, where rules for only two of the three metrics need to be met in order to meet the requirements for a given BCG level. Individual metrics that comprise the BCG rules are combined into an output that shows probability of membership in a BCG level. A sample can have full membership in a single BCG level, a tie between two levels or varying memberships among two or more levels (in which case, the level with the highest membership value is taken as the nominal level). Appendix E contains example BCG model outputs, as well as an interpretation guide.

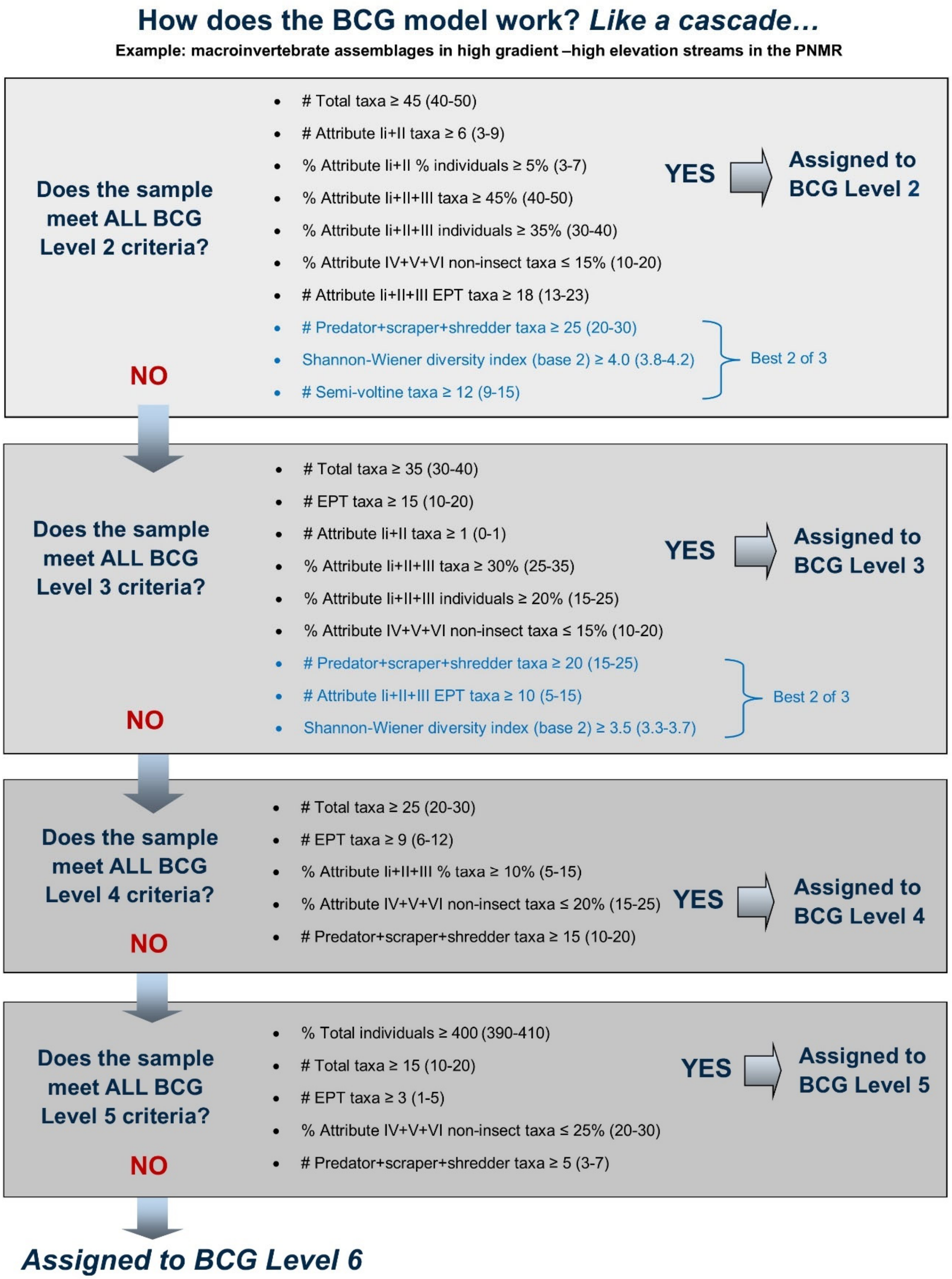


Figure 9. Example flow chart depicting how rules work as a logical cascade in the BCG model. This example is for macroinvertebrate assemblages in high gradient/high elevation freshwater wadeable PNMR streams. The flow chart starts with BCG level 2 because panelists did not assign any samples in this region to BCG level 1. Attribute Ii = sensitive, historically documented, regionally endemic; II = highly sensitive; III = intermediate sensitive; IV = intermediate tolerance; V = tolerant; VI = nonnative or intentionally introduced taxa.

2.3.2 Rule Development for the Pacific Northwest Maritime Region

During the sample scoring process, experts provided their rationale for assigning each sample to a BCG level. The rationale included general qualitative comparisons, qualitative and quantitative expectations based on attribute and taxonomic trait metrics, and expectations for indicator taxa. Different experts used different approaches to arrive at their respective ratings (for example, the taxonomists tended to focus on the specific taxa they saw (good or bad), while others focused more on metrics (e.g., EPT, trophic balance)). As the evidence built for assignments at each level, the group came to an agreement regarding general rules for each BCG level (and were able to achieve consensus despite coming at it from different angles). This agreement was captured in narrative statements compiled through expert contribution and review.

The development of the narrative descriptions was followed by development of quantitative rules, which was also an iterative process. A variety of factors were considered when selecting which biological metrics to include in the rules, including panelist inputs (which metrics best capture their narrative descriptions/rationale), discriminatory power (how well do the metrics discriminate between BCG levels) and redundancy (if a metric is removed, does it change the results?). To evaluate discriminatory power, we generated box plots for a large suite of candidate metrics and examined distributions of metric values across BCG levels (see example in Figure 10). Monotonic patterns are expected for sensitive and tolerant taxa metrics, where total and sensitive taxa metrics decrease and the tolerant and non-native metrics increase as the assigned BCG level (and level of anthropogenic disturbance) increases. Results for the full set of metrics can be found in Appendix F.

To start, we drew upon our experience with the PL/WV BCG model but after some initial exploratory application of this model to PNMR data, rapidly realized the necessity to refine taxa attribute category assignments and develop a more robust classification scheme (low gradient/lower elevation, higher gradient/lower elevation, and higher gradient/higher elevation). We started with the same narrative descriptions and metrics and then refined them for each class. We began by calibrating the high gradient/low elevation BCG model because it represented the greatest number of sites in the dataset, and the experts had the most knowledge and experience with this stream class. To help calibrate the panelists to the lower and upper bounds of the condition gradient, we selected ‘best’ and ‘worst’ samples based on an assortment of criteria (biological metrics, IWI and ICI scores, Biodiversity Index scores (Wisseman 2020)). Most sample assessments were done independently as ‘homework.’ Panelists provided BCG scores and rationale they used when assigning the scores. Refined narrative and draft numeric rules were derived from this information, along with information from the box plots (like Figure 10). The numeric rules were then automated in an Excel spreadsheet and BCG level assignments were calculated for each sample and compared with panelist consensus (median) BCG scores.

Following the initial round of sample assessments (calibration samples), the draft rules were tested with new ‘confirmation’ samples. The new samples spanned the range of anthropogenic stress. As with the calibration round, panelists assessed most of the new samples independently as ‘homework.’ After each round of homework, webinars were held to discuss samples where there were discrepancies between BCG model outputs and panelist consensus scores, and/or where there was high panelist variability (e.g., scores differed by 2 BCG levels). Over the course of this process, we held 10 webinars. Sometimes the BCG rules were revised to better replicate the panelist consensus. During the BCG process, it is not unusual for panelists to update some of their earlier scores because their thinking evolves as they see more samples and become more familiar with the BCG process and/or consider insight and knowledge provided by fellow panelists.

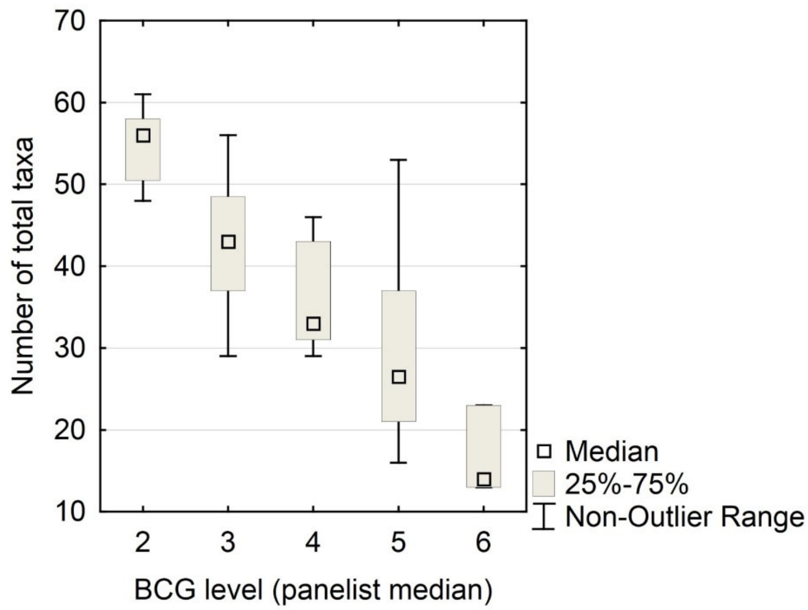


Figure 10. Example of the type of box plot (metric value vs. BCG panelist median) that was used to evaluate discriminatory power of candidate metrics. Plots are based on assessed samples. The whiskers (non-outlier range) cover 1 x the interquartile range (IQR) above the 75th percentile or below the 25th percentile. Box plots for all metrics that were evaluated can be found in Appendix F.

After each round of sample assessments, model performance was assessed. To evaluate the performance of the BCG model, the number of samples were assessed where the BCG decision model's nominal level exactly matched the panel's median ("exact match") and the number of samples where the model predicted a BCG level that differed from the median expert opinion ("mismatch" samples). Then, for the mismatched samples, the differences between the BCG level assignments were examined, and also whether there was a bias in the model (e.g., did the BCG model consistently rate samples better or worse than the panelists).

The quantitative BCG model output is in terms of relative membership of a site among BCG levels, from 0 to 100%, where memberships of all levels must sum to 100%. Most often, model output was 100% assigned to a single BCG level, but it could also yield ties between adjacent levels, or a majority assigned to one level over one or more others. Panelists could split among BCG levels as well. Tetra Tech used the following criteria when assigning ties:

- **BCG model ties**, where there is nearly equal membership in two BCG levels (the difference between the primary and secondary memberships is less than 0.2; e.g., membership of 0.54 in BCG level 2 and membership of 0.46 in BCG level 3).

- **Panelist ties**, where a single vote could have flipped the decision (e.g., 4-4, or 5-4 decisions).

Figure 11 provides examples of the two types of ties and also shows the BCG level rating scale that was used to assess model performance. Table 9 has a (hypothetical) example of a model performance worksheet that we used to tally the performance results that are reported in Section 3.4. It shows examples of: an exact match (where both the panel and model assigned a sample to the same BCG level or the same tie); differences of 0.5 BCG level (e.g., if the BCG model assignment was a BCG level 2/3 tie and panelist consensus was a BCG level 2, the model was considered to be 'off by a 0.5 BCG level'; or more specifically, the model rating was -0.5 BCG level (a half level "worse" than the panelists' consensus); and a difference of 1 BCG level.

After each round of sample assessments, the performance of the BCG model was assessed. For this project, satisfactory performance was rated as ≥ 90% of the BCG model outputs were within ±0.5 BCG level of the panelist consensus score (which is consistent with standard used in previous BCG projects). When this target was achieved with calibration samples, new samples, independent of previous samples, were evaluated by the panel to confirm that model replicates expert panel decision rules. These confirmation samples covered the range of anthropogenic stress. Typically, the confirmation dataset consisted of 10-15 samples (about 10% of the calibration dataset) but sample size was driven largely by the availability of the panelists to perform sample assessments. If the ≥ 90% target was not achieved, the samples were moved into the calibration dataset and BCG rules were refined as needed. This process was repeated until the confirmation target was achieved.

Is this the best way to assess model performance?
This approach is based on the distribution of errors (mismatches) combined with the panel consensus on the degree of error that is biologically meaningful. It is the best statistical estimate of goodness-of-fit that we know for the BCG at this time. Other measures do not yield interpretable results, for example, Cohen's Kappa (e.g., Agresti 2013) estimates *p*-values compared to random independence and the resultant p-values are so small (10^{-100} and smaller) as to be meaningless.

Scale used for evaluating model performance

BCG level assignments
2
2/3 tie
3
3/4 tie
4
4/5 tie
5
5/6 tie
6

0.5-level difference

1-level difference

Two types of ties

BCG model ties, difference between the primary and secondary memberships < 0.2. Examples:

Primary BCG level	Primary membership	Secondary BCG level	Secondary membership
3	0.50	4	0.50
3	0.55	2	0.45

Panelist ties, single votes flip the decision (e.g., 4-4, or 5-4 decisions). Examples:

Panelist consensus	Panelist assignments
3/4 tie	three 4s, three 3s, one 3/4 tie
2/3 tie	five 3s, four 2s

Figure 11. Model performance was evaluated based on 0.5-level increments (e.g., level 2 vs. level 2/3 tie = a 0.5-level difference; level 3 vs. level 4 = a 1-level difference). Model and panelist ties were defined as shown above.

Table 9. Example of the type of model performance worksheet that we used to tally the model performance results (Section 3.4).

Panelists		BCG model					
Panelist consensus	**Panelist scores**	**Primary BCG level**	**Primary membership**	**Secondary BCG level**	**Secondary membership**	**Tie**	**Performance**
5	seven 5s, one 4, one 6	5	1	-	0	-	Exact match
3/4 tie	three 4s, three 3s, one 3/4 tie	4	0.67	3	0.33	-	Model rating is 0.5 level worse
2	seven 2s, two 3s, one 2/3 tie	3	0.80	2	0.20	-	Model rating is 1 level worse
3	six 3s, two 2s	2	0.50	3	0.50	2/3 tie	Model rating is 0.5 level better

2.3.3 Panelist Variability

In addition to evaluating agreement between BCG model outputs and panelist median scores, we also evaluated the level of agreement among individual panelists. We expected higher levels of agreement in the confirmation versus calibration dataset. During the early rounds of BCG model calibration, as panelists were gaining experience and becoming familiar with samples in the calibration dataset, there tended to be more divergence (especially since panelists were mostly doing assessments independently as 'homework'). But during the confirmation round, we expected less variability (at that point, if panelist scores consistently differed from the group median by more than one BCG level, it was cause for concern). In addition, we wanted to evaluate differences across BCG levels (e.g., was there better agreement/less variability between panelist scores in samples on the extreme ends of the stressor gradient (BCG levels 2 and 6) versus in the middle (BCG level 3 vs. 4 and 4 vs. 5).

As described in Section 2.3, panelists were allowed to rate samples as a single BCG number, and sometimes applied a plus (+) or minus (-) to the level indicating somewhat better or worse condition, respectively (much as school letter grades can be modified, e.g., B+). We first compiled all the individual panelist BCG level scores. In situations where samples were assessed more than once, we only used the last assessment. The number of panelist calls per sample ranged from 4 to 9. Next, we converted the plus/minus scores to numeric by 0.25 increments. For example, 2+ = 1.75, 2 = 2, 2- = 2.25, 2/3 tie = 2.5, 3+ = 2.75, 3 = 3, 3- = 3.25, 3/4 tie = 3.5, 4+ = 3.75, 4 = 4, and so on. Then, we calculated differences between the individual scores and the panelist median scores for each sample and placed the individual calls into five 'bins' based on the magnitude of the difference: 0, ± 0.1 to 0.5, ± 0.6 to 1.0, ± 1.1 to 1.5 and ± 1.6 to 2. We calculated the percentage of individual panelist BCG scores in each 'bin' (for example, 50% of the panelists scores matched exactly with the group median score (difference of 0), 20% of the scores differed from the median by < 0.5 (and were placed into the '± 0.1 to 0.5' bin), and so on). Results were grouped by dataset (calibration versus confirmation) and presented in a table. In addition, we generated column graphs to make it easier to visualize the distribution of individual panelist scores compared to the group median.

3 RESULTS

3.1 Taxon Attribute Categories

Panelists evaluated 615 taxa and had sufficient information to assign 491 to BCG attribute categories I-VI. Table 10 shows the distribution of taxa across BCG attribute categories and includes examples from each group. Approximately 60% of the taxa were assigned to Attributes III and IV (~30% in each). Attachment B has the full list of taxa with BCG attribute and trait assignments that were used to calculate metrics for BCG model calibration. If running the PNMR BCG model, it is important to use these same attributes for metric calculations. The PNMR BCG attribute and trait assignments supersede those made during the calibration of the PL/WV BCG model.

Table 10. Summary of the number of taxa that were assigned to each BCG attribute category, with examples of taxa in each category. 491 of the 615 taxa in the PNMR BCG calibration dataset were assigned to BCG attribute categories.

BCG Attribute	Description	Number	Percent	Example Taxa
Ii (sensitive)	Historically documented, sensitive, long-lived or regionally endemic taxa	35	5.7%	Mayflies: *Drunella pelosa, Caudatella cascadia* Stoneflies: *Diura, Frisonia picticeps* Caddisflies: *Chyranda, Himalopsyche phryganea,* Beetles: *Amphizoa, Bryelmis,* Flies (Diptera): *Metacnephia*, *Lipsothrix*
Im (moderately tolerant)		1	0.2%	Mayfly: *Serratella levis*
II	Highly sensitive taxa, often occur in low abundance	58	9.4%	Mayflies: *Caudatella*, *Epeorus grandis group* Stoneflies: *Kathroperla, Megarcys, Yoraperla*, Caddisflies: *Apatania, Oligophlebodes*, *Rhyacophila hyalinata group*, Beetles: *Ampumixis dispar*, Flies: *Cricotopus (Nostococladius), Oreogeton*
III	Intermediate sensitive taxa	168	27.3%	Mayflies: *Drunella doddsii, Cinygmula, Epeorus,* Stoneflies: *Skwala*, *Isoperla, Pteronarcys, Zapada oregonensis group*, Caddisflies: *Arctopsyche, Micrasema, Rhyacophila betteni group,* Beetles: *Heterlimnius corpulentus*, Flies: *Cladotanytarsus, Glutops, Pagastia*
IV	Taxa of intermediate tolerance	187	30.4%	Mayflies: *Acentrella turbida, Baetis tricaudatus complex, Cinygma, Rhithrogena*, Stoneflies: *Zapada cinctipes, Sweltsa, Malenka*, Caddisflies: *Hydropsyche, Glossosoma*, *Wormaldia,* Beetles: *Optioservus, Lara, Zaitzevia, Narpus*, Flies: *Simulium, Ceratopogoninae, Brillia, Eukiefferiella, Micropsectra*, Snails: *Juga*, Bivalves: Sphaeriidae, Worms, Mites
V	Tolerant native taxa	40	6.5%	Mayflies: *Caenis, Fallceon, Maccaffertium*, Caddisflies: *Cheumatopsyche, Parapsyche almota*, Flies: *Chironomus, Cryptochironomus, Procladius, Hemerodromia, Paratanytarsus,* Snails: *Physidae,* Leeches: *Erpobdella,* Amphipods: *Crangonyx, Gammarus*, Isopods: *Caecidotea*
VI	Non-native tolerant taxa	2	0.3%	New Zealand mudsnail: *Potamopyrgus antipodarum,* Bivalve: *Corbicula*
NA	No attribute assignment (insufficient information)	124	20.2%	Attributes were often not assigned (NA) for taxa that occurred infrequently, were unfamiliar to the experts, or were of ambiguous taxonomic resolution
Totals		***615***	***100***	

3.2 BCG Scoring

During the calibration and confirmation rounds, panelists made BCG level assignments on 173 total samples. Of those, we ended up excluding 18 from BCG model calibration, due primarily to coarse levels of taxonomic resolution for Chironomidae (which reduced richness metrics and biased BCG scores downward). Others were clear outliers (e.g., extreme dominance by one or two taxa). Table 11 shows how the remaining 155 assessed samples were distributed across the three gradient/elevation classes and BCG levels. Some samples were assessed more than once. In those situations, we only used the BCG levels from the last assessment. Locations of the assessed samples (color-coded by BCG level) are shown in Figure 12.

No sites were assigned to BCG level 1 (naturally occurring, no impact from anthropogenic stress) (Davies and Jackson 2006). Participants agreed that all sites in the study area had some degree of disturbance, including legacy effects from agriculture, forestry and urbanization, so BCG level 2 samples represent the waters in this exercise that were closest to natural. Of the 155 samples that were assessed, 29 were assigned to BCG level 2. Only two Level 2 samples were from the low gradient/low elevation class. Fourteen were scored as BCG level 6, which represents the most altered condition. Only four samples in the high gradient/high elevation class were assigned to BCG level 5 or 6. The greatest number of samples (n=54) were assigned to BCG level 3 (Table 11). BCG level assignments for all assessed samples are provided in Attachment C.

Table 11. Number of calibration and confirmation samples that were assessed, organized by BCG class and level (panelist median).

BCG level	# Assessed samples			Total
	Low Gradient Low Elevation	**High Gradient Low Elevation**	**High Gradient High Elevation**	
2	2	13	14	***29***
2/3 tie	1	2	1	***4***
3	20	18	16	***54***
3/4 tie	1	0	2	***3***
4	5	12	7	***24***
4/5 tie	1	0	1	***2***
5	8	13	3	***24***
5/6 tie	1	0	0	***1***
6	8	5	1	***14***
Total	***47***	***63***	***45***	***155***

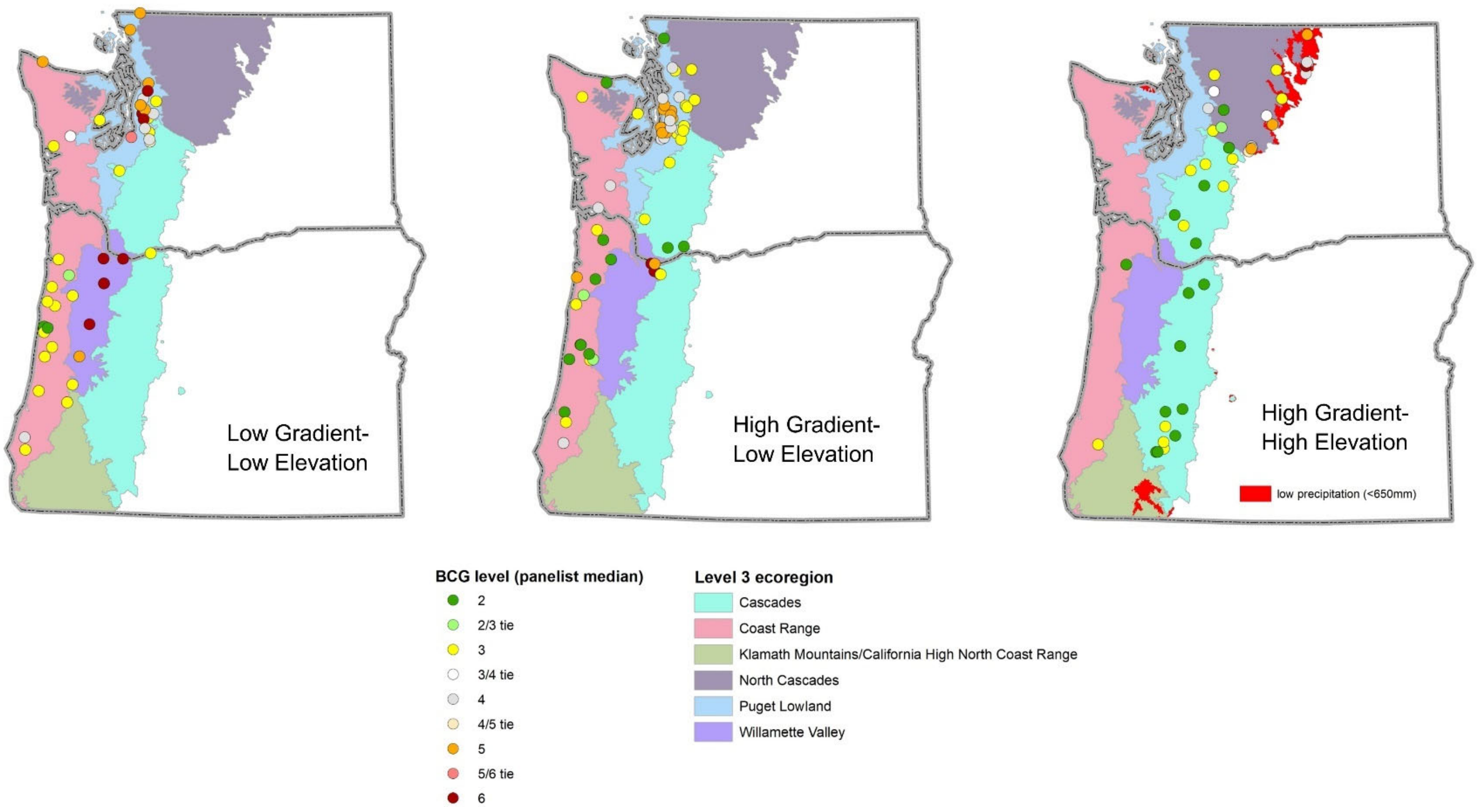


Figure 12. Locations of samples assessed during BCG model calibration and confirmation, color-coded by panelist BCG level assignment (group median), for low gradient (<1%)/lower elevation (<750 m) (left), higher gradient (≥1%)/lower elevation (<750 m) (middle) and higher gradient (≥1%)/higher elevation (≥750 m) (right).

3.3 BCG Rules

The basis of the BCG rules is the expert defined changes in biota associated with increase anthropogenic stress. In most places, but not all, the observed changes along gradient of stress is loss of sensitive taxa and an increase in tolerant and often non-native taxa. Table 12 contains the narrative and quantitative rules for the PMNR BCG levels 2 through 6, along with the narrative description of BCG level 1 (Section 2.2.1), which is based on expert panel judgement and results of the historic literature review (Hafele et al. 2022; Appendix D). Box plots like the ones shown in Figures 13 & 14 were used to help select metrics that best discriminated between BCG levels. The plots show distributions of metric values for all three classes as conditions worsen from BCG level 2 down to level 6.

There are lots of similarities in BCG rules across classes. For example, BCG level 2 and 3 streams in all three classes are expected to support a diverse, balanced community (as measured by the Shannon-Wiener diversity index) with moderate to high numbers of total taxa. Almost half of the taxa in BCG level 2 samples are sensitive (Attribute Ii, II, III) and there are high numbers of taxa representing key functional feeding groups (predators, scrapers and shredders), which indicate intact ecosystem functions and food chains. There are also moderate to high numbers of semivoltine taxa, which require perennial flow and high-quality habitat to complete their life cycle. Non-insect taxa that are non-native or moderately to highly tolerant (Attribute IV, V, VI) make up a small proportion (<15%) of the community.

In BCG level 3 samples, sensitive taxa occur in reduced numbers but are still common and abundant, as are key functional feeding groups. The community still has good diversity, with over a third of the taxa comprised of mayflies, stoneflies and caddisflies (EPT taxa). Attribute IV+V+VI non-insect taxa occur in slightly higher numbers but still make up a small proportion of the community. BCG level 4 samples continue to lose diversity and sensitive taxa, but still retain moderate numbers of EPT taxa, as well as predators, scrapers and shredders. Major changes in structure and function become evident in the BCG level 5 samples. Mayfly, stonefly and caddisfly taxa are still present but in greatly reduced numbers, and total taxa richness may drop to low to moderate numbers. Percent Attribute IV+V+VI non-insect taxa continue to increase, but still do not dominate, comprising less than 25 or 35% of the community, depending on the class.

In BCG level 5 samples, total richness varies widely. Also, the number of metrics is reduced (5-7 metrics compared to 10-11 in the BCG level 2 rules). Samples have markedly reduced numbers of EPT taxa and predators, scrapers and shredders, and Attribute IV+V+VI non-insect taxa continue to increase. A new metric - total number of individuals - is introduced in BCG level 5. The rule requires at least 400 total individuals (-20% of the subsampling target) to be present in 500-organism samples. The panelists decided not to put in a similar requirement for BCG levels 2-4 because it would cause some samples with diverse communities and well-represented sensitive taxa to be assigned to BCG level 6. Instead, the panelists decided to flag BCG level 2-4 samples that have fewer than 450 total individuals for further evaluation. Section 3.3.1 describes this as well as other flagging criteria.

Differences in metrics and thresholds across BCG classes are highlighted below -

Higher gradient/higher elevation

- support naturally less diverse communities, so thresholds are lower for richness metrics (for example, 45 versus 50 total taxa in BCG level 2 samples)
- are comprised of higher proportions of highly sensitive and sensitive, regionally endemic taxa so thresholds for Attribute Ii+II metrics are higher (for example, six versus three Attribute Ii+II taxa in BCG level 2 samples)
- have fewer Attribute IV+V+VI non-insect taxa, so thresholds are lower (for example, 15 versus 20% in BCG level 3 samples)

Lower gradient/lower elevation

- have more tolerant and non-native taxa than the other classes; the % Attribute V+VI individuals metric is used exclusively for this class and is important for discriminating between all the BCG levels
- expect slightly fewer EPT taxa so thresholds are lower in BCG level 3 and 5 rules

Higher gradient/lower elevation

- the '% Attribute IV+V+VI non-insect individuals, excluding Juga and mites' metric is used exclusively for this class and helps discriminate between BCG levels 2-4. The Juga are excluded because they can have patchy distributions and samplers sometimes hit a 'clump' of them during collection. Panelists were comfortable with Juga being present in high numbers as long as the rest of the assemblage was diverse and sensitive taxa were well-represented. Mites were excluded because their numbers were highly variable, due in part to differences across organizations in level of collection effort and taxonomic resolution. Due to these inconsistencies, we had to collapse them to Order-level, which provides limited information on sensitivity to pollution and makes mite data difficult to interpret.

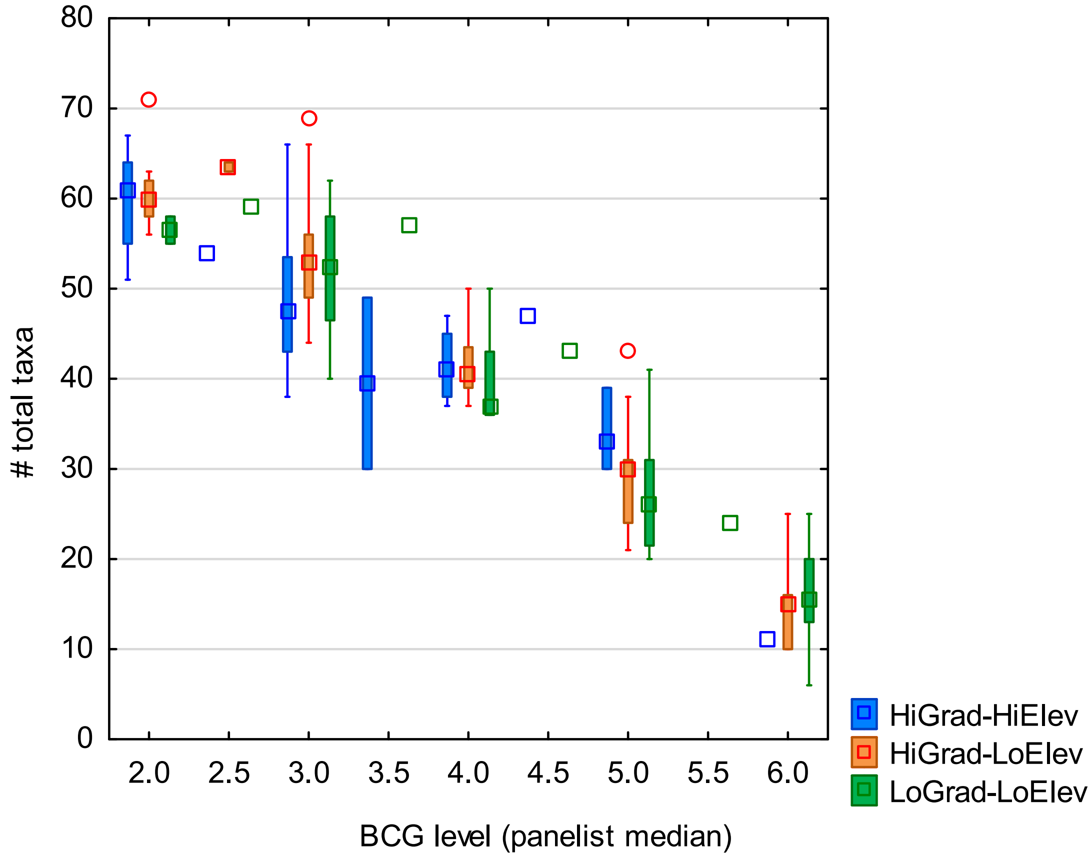


Figure 13. Box plot of number of total taxa vs BCG level based on the panelist median. The half-level increments represent ties between levels (e.g., BCG level 2.5 is a tie between levels 2 & 3). Data points were marked as outliers or extremes if they were more than 1.5 x IQR above the 75th percentile or below the 25th percentile. The whiskers (non-outlier range) cover 1 x IQR above the 75th percentile or below the 25th percentile. Plots are based on assessed samples: low gradient/lower elevation (sample size = 47); higher gradient/lower elevation (sample size = 63); and higher gradient/higher elevation (sample size = 45). Box plots for all metrics that were evaluated can be found in Appendix F.

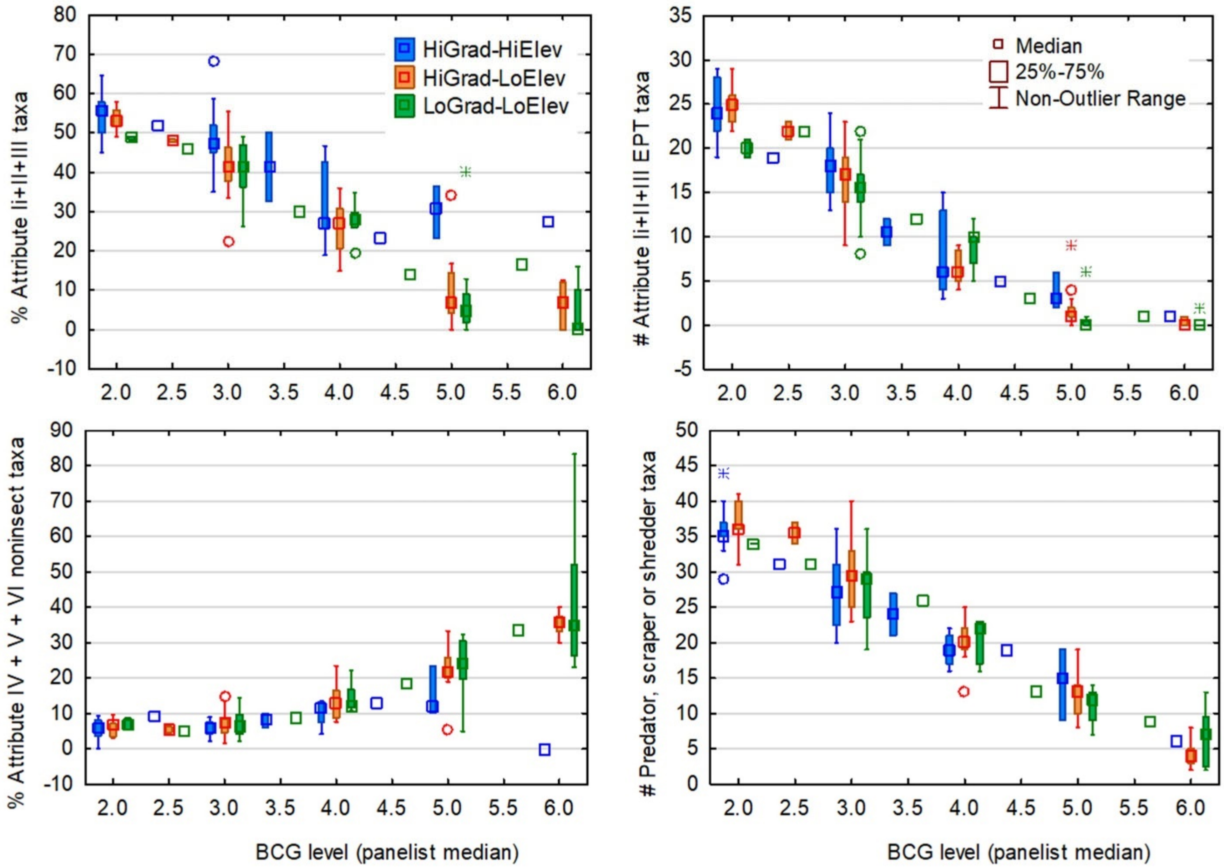


Figure 14. Box plots of % Attribute Ii+II+III % taxa, % Attribute IV+V+VI non-insect taxa, number of sensitive EPT taxa and number of predator, scraper and shredder taxa vs BCG level based on the panelist median. The half-level increments represent ties between levels (e.g., BCG level 2.5 is a tie between levels 2 & 3). Data points were marked as outliers or extremes if they were more than 1.5 x IQR above the 75th percentile or below the 25th percentile. The whiskers (non-outlier range) cover 1 x IQR above the 75th percentile or below the 25th percentile. Plots are based on assessed samples: low gradient/lower elevation (sample size = 47); higher gradient/lower elevation (sample size = 63); and higher gradient/higher elevation (sample size = 45). Box plots for all metrics that were evaluated can be found in Appendix F.

Table 12. BCG level 1 narrative description and quantitative decision rules for levels 2-6 for the low gradient/lower elevation, higher gradient/lower elevation, and higher gradient/higher elevation classes. The numbers in parentheses represent the lower and upper bounds of the fuzzy sets (for more details, see Section 2.3.1). Cells colored with beige shading are 'best 2 of 3' rules (where samples can fail one of the rules and still meet the requirements for that BCG level). BCG levels 2-4 rules for higher gradient/lower elevation streams include the '% Attribute IV+V+VI non-insect individuals, excluding Juga and mites' metric. The Juga are excluded because they can have patchy distributions and samplers sometimes hit a 'clump' of them during collection. The panelists are comfortable with Juga being present in high numbers as long as the rest of the assemblage is diverse and sensitive taxa are well-represented. Mites were excluded because their numbers were highly variable, due in part to differences across organizations in level of effort to collect and identify them. Moreover, when collapsed to Order-level, they provide limited information on sensitivity to pollution, so high numbers are difficult to interpret.

BCG level 1: Natural or native condition. Native structural, functional and taxonomic integrity is preserved; ecosystem function is preserved within the range of natural variability. Level 1 represents biological conditions as they existed (or still exist) in the absence of measurable effects of stressors and provides the basis for comparison to the next five levels.
There are no quantitative rules for BCG level 1 at this time because there were no sites in the model calibration dataset that met the conceptual definition. Instead, we are using this narrative description developed by the expert panel (see Hafele et al. 2022; Appendix D) to provide a conceptual understanding of BCG level 1. However, a numeric index has been proposed to identify sites that support macroinvertebrate communities with high and/or unique taxa richness (see Appendix D). The index could be used to "flag" streams that may be close to or comparable to a BCG level 1 stream in the Pacific Northwest Maritime Region and encourage consideration for protection or restoration. The index was developed for moderate to high gradient, low to mid-elevation streams in the PNMR with coarse mineral substrates and extensive erosional habitat and would need to be adjusted for the low gradient/lower elevation and higher gradient/higher elevation stream classes. The index is an outgrowth of panel deliberations on how to address BCG level 1 in regions where there are no undisturbed or minimally disturbed streams remaining and is offered as a hypothesis for future testing and revision. *Streams with high habitat complexity; natural disturbance regimes to refresh micro-habitats; year-round flow without anthropogenic impacts to hydrology, temperature, or water quality; water often dominated by cool-cold flow from springs, groundwater accretion, and/or natural runoff; high resilience to disturbance including drought and flood extremes; intact riparian and aquatic habitat that provide longitudinal and lateral connectivity; exemplary biological diversity with high taxa richness of rheophilic[4], lotic-depositional, and micro-habitat specialist macroinvertebrates; non-native invasive species absent; biotic community supports all ecosystem functions.*

[4] Rheophilic taxa prefer to live in fast moving water

Table 12 continued…

BCG level 2: Minimal changes in structure of the biotic community and minimal changes in ecosystem function. A diverse community reflects high habitat complexity and resilience; regionally endemic and highly sensitive taxa are maintained with some changes in biomass and/or abundance; sensitive taxa contribute a high proportion of the total taxa richness and individuals; ecosystem functions are fully maintained within the range of natural variability.

Narrative Descriptions	Metric	Numeric rules		
		Low Gradient Low Elevation	**High Gradient Low Elevation**	**High Gradient High Elevation**
Diverse assemblage with moderate to high numbers of total taxa	Number of total taxa	≥ 50 (45-55)	≥ 50 (45-55)	≥ 45 (40-50)
Highly sensitive taxa (II) and sensitive, historically documented, regionally endemic (Ii) taxa are well-represented, particularly in high gradient/high elevation streams	Number of Attribute Ii+II taxa	≥ 3 (1-5)	≥ 3 (1-5)	≥ 6 (3-9)
	% Attribute Ii+II % individuals	≥ 2% (1-3)	≥ 2% (1-3)	≥ 5% (3-7)
Sensitive mayfly, stonefly and caddisfly taxa are present in high numbers	Number of Attribute Ii+II+III EPT taxa	≥ 18 (13-23)	≥ 20 (15-25)	≥ 18 (13-23)
Nearly half (or more) of the taxa are sensitive	% Attribute Ii+II+III % taxa	≥ 45% (40-50)	≥ 45% (40-50)	≥ 45% (40-50)
Sensitive individuals comprise a third or more of the community	% Attribute Ii+II+III % individuals	≥ 30% (25-35)	≥ 30% (25-35)	≥ 35% (30-40)
Non-insect taxa that are non-native or have moderate to high tolerance to pollution make up a small proportion of the community. Juga and mites are excluded from the % individual metric (see caption)	% Attribute IV+V+VI non-insect taxa	≤ 15% (10-20)	≤ 15% (10-20)	≤ 15% (10-20)
	% Attribute IV+V+VI non-insect individuals, excluding Juga and mites	--	≤ 15% (10-20)	--
Tolerant and non-native taxa (not limited to non-insect only) make up a very small fraction of the individuals or are absent	% Attribute V+VI individuals	< 1% (0-1)	--	--
High numbers of taxa representing key functional feeding groups indicate intact ecosystem functions and food chains	Number of predator, scraper and shredder taxa	≥ 30 (25-35)	≥ 30 (25-35)	≥ 25 (20-30)
Excellent diversity and balance	Shannon-Wiener diversity index (base 2)	≥ 4.5 (4.3-4.7)	≥ 4.5 (4.3-4.7)	≥ 4.0 (3.8-4.2)
Good representation of taxa that require more than one year to complete their life cycle, which indicate perennial flow and high-quality habitat	Number of semi-voltine taxa	≥ 15 (10-20)	≥ 15 (10-20)	≥ 12 (9-15)

Best 2 of 3

Table 12 continued…

BCG level 3: Evident changes in structure of the biotic community from reduced habitat complexity and resilience, with minimal changes in ecosystem function. Diminished biodiversity due to loss of some regionally endemic and highly sensitive taxa. Some shift towards dominance by common, widespread, less sensitive taxa. Intermediate sensitive taxa are still common and abundant, and ecosystem functions are fully maintained.

Narrative Descriptions	Metric	Numeric rules		
		Low Gradient Low Elevation	High Gradient Low Elevation	High Gradient High Elevation
Moderate to high numbers of total taxa	Number of total taxa	≥ 40 (35-45)	≥ 40 (35-45)	≥ 35 (30-40)
Moderate to high numbers of mayfly, stonefly and caddisfly taxa	Number of EPT taxa	≥ 15 (10-20)	≥ 20 (15-25)	≥ 15 (10-20)
At least one highly sensitive taxa (II) or sensitive, regionally endemic (Ii) taxon is present in high gradient/high elevation streams	Number of Attribute Ii+II taxa	--	--	≥ 1 (0-1)
Sensitive taxa occur in reduced numbers are still common and abundant	% Attribute Ii+II+III % taxa	≥ 20% (15-25)	≥ 20% (15-25)	≥ 30% (25-35)
	% Attribute Ii+II+III % individuals	≥ 10% (5-15)	≥ 10% (5-15)	≥ 20% (15-25)
Non-insect taxa that are non-native or moderately to highly tolerant occur in slightly higher numbers but still make up a small proportion of the community. Juga and mites are excluded from the % individual metric for reasons described in the caption	% Attribute IV+V+VI non-insect taxa	≤ 20% (15-25)	≤ 20% (15-25)	≤ 15% (10-20)
	% Attribute IV+V+VI non-insect individuals, excluding Juga and mites	--	≤ 20% (15-25)	--
Moderate to high numbers of taxa from key functional feeding groups indicate that ecosystem functions are fully maintained	Number of predator, scraper and shredder taxa	≥ 18 (13-23)	≥ 25 (20-30)	≥ 20 (15-25)
Sensitive EPT taxa are present in moderate to high numbers	Number of Attribute Ii+II+III EPT taxa	≥ 10 (5-15)	≥ 10 (5-15)	≥ 10 (5-15)
Good diversity and balance	Shannon-Wiener diversity index (base 2)	≥ 4 (3.8-4.2)	≥ 4 (3.8-4.2)	≥ 3.5 (3.3-3.7)
Tolerant and non-native individuals make up a very small fraction of the community	% Attribute V+VI individuals	≤ 5% (3-7)	--	--

Best 2 of 3

Table 12 continued…

BCG level 4: Moderate changes in structure of the biotic community and minimal changes in ecosystem function. Moderate changes in structure due to replacement of some intermediate sensitive taxa by more tolerant taxa, but reproducing populations of some sensitive taxa are maintained; overall balanced distribution of all expected major groups; ecosystem functions largely maintained through redundant attributes				
Narrative Descriptions	**Metric**	**Numeric rules**		
		Low Gradient Low Elevation	**High Gradient Low Elevation**	**High Gradient High Elevation**
Moderate numbers of total taxa	Number of total taxa	≥ 30 (25-35)	≥ 30 (25-35)	≥ 25 (20-30)
Sensitive taxa occur in reduced numbers but still comprise at least a tenth of the taxa	% Attribute Ii+II+III % taxa	≥ 10% (5-15)	≥ 10% (5-15)	≥ 10% (5-15)
Non-insect taxa that are non-native or moderately to highly tolerant become more prevalent but comprise less than a quarter of the assemblage.	% Attribute IV+V+VI non-insect taxa	≤ 25% (20-30)	≤ 25% (20-30)	≤ 20% (15-25)
Non-insect taxa that are non-native or moderately to highly tolerant may comprise up to half the individuals. Juga and mites are excluded for reasons described in the caption	% Attribute IV+V+VI non-insect individuals, excluding Juga and mites	--	≤ 50% (45-55)	--
Moderate numbers of mayfly, stonefly and caddisfly taxa	Number of EPT taxa	≥ 9 (6-12)	≥ 9 (6-12)	≥ 9 (6-12)
Moderate numbers of taxa from key functional feeding groups indicate that ecosystem function is largely maintained	Number of predator, scraper and shredder taxa	≥ 12 (7-17)	≥ 15 (10-20)	≥ 15 (10-20)
Tolerant and non-native individuals may be present in higher numbers but still comprise a small proportion of the community	% Attribute V+VI individuals	≤ 15% (10-20)	--	--

Best 2 of 3

Table 12 continued...

BCG level 5: Major changes in structure of the biotic community and moderate changes in ecosystem function. Sensitive taxa are markedly diminished; conspicuously unbalanced distribution of major groups from that expected; organism condition shows signs of physiological stress; system function shows reduced complexity and redundancy; increased build-up or export of unused materials				
Narrative Descriptions	**Metric**	**Numeric rules**		
		Low Gradient Low Elevation	**High Gradient Low Elevation**	**High Gradient High Elevation**
At least -20% of the subsampling target is achieved (based on 500-count samples)	Number of total individuals	≥ 400 (390-410)	≥ 400 (390-410)	≥ 400 (390-410)
Mayfly, stonefly and caddisfly taxa still present but in greatly reduced numbers	Number of EPT taxa	≥ 1 (0-1)	≥ 3 (1-5)	≥ 3 (1-5)
Non-insect taxa that are non-native or moderately to highly tolerant are more prevalent, but comprise a third or less of the community	% Attribute IV+V+VI non-insect taxa	≤ 35% (30-40)	≤ 35% (30-40)	≤ 25% (20-30)
Reduced number of taxa from key functional feeding groups indicate moderate changes in ecosystem function	Number of predator, scraper and shredder taxa	≥ 5 (3-7)	≥ 5 (3-7)	≥ 5 (3-7)
At a minimum, low to moderate number of taxa (can range widely)	Number of total taxa	≥ 20 (15-25)	≥ 15 (10-20)	≥ 15 (10-20)
Sensitive individuals still present but in greatly reduced numbers	% Attribute Ii+II+III % individuals	≥ 1% (0-2)	--	--
Up to a third of total individuals may be tolerant or non-native	% Attribute V+VI individuals	≤ 30% (25-35)	--	--

best 2 of 3

3.3.1 Flagging Criteria

In addition to BCG rules, the panelists established criteria for flagging samples to aid in interpretation of the results (the flags are not intended to be grounds for discarding samples). The flags are included in the BCG model output from the BCGcalc R package. Criteria and rationale are summarized in Table 13. Flags are grouped into several categories. Some alert users to site or sample characteristics that are outside the experience of the BCG calibration dataset (e.g., samples are not collected within the normal index period, or sites are located outside the PNMR). Other flags mark samples with potential classification issues (e.g., if the site is borderline between two BCG classes – for example, has a 0.9% slope, in which case we recommend running it through both the low and high gradient BCG models). The other types of flags are for unusual characteristics (e.g., lower than normal number of organisms, dominance of one or two taxa), or signals about potential stressors (e.g., urban, nutrient). Flags will need to be evaluated and updated as panelists gain more experience with the BCG models.

Table 13. Samples are flagged based on the criteria below, to aid in interpretation of BCG model results.

Flag type	Flag	Criteria	Rationale
Site or sample characteristics outside the experience of the model	Outside study area	Not located in the PNMR Level 3 ecoregions (Coast Range, Puget Lowland, Willamette Valley, Cascades, and North Cascades)	The BCG models have not been tested on streams outside the study area
	Xeric/low precipitation	Mean annual precipitation < 1000 mm, based on PRISM 1982-2010 data (source: EPA StreamCat, NHDPlusV2 local catchment)	During sample assessments, macroinvertebrate communities in xeric areas (in particular, with < 650 mm mean annual precipitation) seemed noticeably different. The BCG model may need to be adjusted for xeric streams.
	Brackish	*Americorophium*, *Gnorimosphaeroma, Mysis* or Mysidae are present	Brackish water will affect the freshwater community. Note – Rammellogammarus (may be ID'd as Anisogammaridae or Eogammarus in some data) is estuarine, but it does often penetrate upstream into pure fresh water.
	Too many organisms	> 600 total individuals in 500-count samples	Samples should subsampled before metric calculation, otherwise richness metrics may be inflated (the larger the subsample size, generally the higher the richness)
	Timing of collection	Collection month before June and after October	Seasonality affects which taxa are present.
	Reduced sampling effort	Sampling area ≥ 8 ft^2	Higher likelihood of collecting a poor sample and missing sensitive taxa
	Size/drainage area	Very small (< 2 mi^2 drainage area)	Very small streams have a higher likelihood of going dry in certain years, and may have a higher variation in insect abundance
		Too large (> 100 mi^2 drainage area)	The BCG models have not been tested on streams > 100 mi2. Larger stream
Classification	Gradient	NHDPlusV2 flowline slope ≥ 8%	For HiGrad-HiElev streams; ≥ 8% is on the high end for this stream class and suggests a cascading stream with high levels of natural disturbance
		NHDPlusV2 flowline slope ≥ 5% or < 1.2%	Transitional HiGrad-LoElev; if ≥ 5% (high for this stream class), consider also running through the HiGrad-HiElev model; if <1.2% (low for this stream class), consider also running through the LoGrad-LoElev model
		NHDPlusV2 flowline slope ≥ 0.7%	Transitional LoGrad-LoElev; if ≥ 0.7 (high for this stream class), consider also running through the HiGrad-LoElev model

Table 13 continued…

Flag type	Flag	Criteria	Rationale
Classification	Soft bottom substrate signal	% Odonata individuals ≥ 10%	Suggests soft-bottom stream (which could be natural or unnatural); we had limited high quality, naturally soft-bottom streams in the BCG calibration dataset. If more data become available, the BCG model may need to be adjusted for soft-bottom streams.
		% Corbiculidae + Sphaeriidae individual ≥ 10%	
Unusual characteristics	Lower than normal number of organisms	Raw counts < 450 total individuals in 500-count samples	The 450-individual threshold allows for a 50-organism ‘buffer’ in case the taxonomist rejects some organisms that the sorting technician includes. Something odd is going on if you don't get at least 450 organisms.
		Subsample_percent ≥ 75%	Proxy for density; high proportion picked indicates lower than normal number of organisms
		Density < 1000 organisms/m^2	Density is not available for all sites, otherwise this metric may be preferred over raw counts because it takes into account differences in sampling area and subsampling effort. Moving ahead, we recommend that density be included as a standard output and that this threshold be further explored and refined.
	Dominance of one or two taxa	The two most dominant taxa comprise ≥ 50% of the total individuals	Resampling should be performed if resources permit. Dominance of one or two taxa affects richness metrics in particular. There are no easy answers on how to address it (if labs kept picking beyond 500 organisms, they’d likely get more of the same - e.g., 600 *Baetis tricaudatus* instead of 300). If this occurs in BCG level 3 samples, panelists suspect that sensitive taxa are present but are being missed due to the dominant taxa; as sites get more degraded, panelists don’t expect as many taxa are being missed (they expect more of the same tolerant taxa). The dominance issue accounts for a fair amount of year-to-year variability (which affects IBI scores as well as BCG scores).
	Missing BCG attribute category	One or more taxa without a BCG attribute	Most of the BCG rules are based on taxa attribute metrics; the more taxa missing assignments, the more it will affect BCG model outputs (especially if the taxon comprises a large percentage of individuals)
	High mites	% Mite individuals ≥ 10%	Mites abundances were highly variable, due in part to differences across organizations in level of effort to collect and identify them. High numbers limits representation of other taxa in the sample
	High Juga	% Juga + Fluminicola individuals ≥ 25%	These can have patchy distributions; samplers sometimes hit a ‘clump’ of them during collection which limits representation of other taxa in the sample

Table 13 continued…

Flag type	Flag	Criteria	Rationale
Signal	Disturbance	% Baetis tricaudatus complex + Simuliidae individuals ≥ 35%	Signal of high flow event (which could be natural or anthropogenic – e.g., stormwater induced)
		% Multivoltine taxa ≥ 50%	High proportion of rapidly reproducing taxa indicates disturbance
	Nutrient	% Collector gatherer and filterer individuals ≥ 50%	Nutrient enrichment often triggers elevated numbers of these functional feeding groups
	Urban	% Crangonyx + Caecidotea + Gammarus individuals ≥ 10%	Tolerant of degraded conditions associated with urban environments
	Stressors	No mayfly taxa	Something odd is going on if you don't get at least one taxon in each of these Orders.
		No stonefly taxa	
		No caddisfly taxa	
		No beetle taxa	
		% Chironomidae individuals ≥ 30%	high proportions tend to indicate disturbance (but not always – it depends on the taxa – e.g., certain types of worms, like the genus Mesenchytraeus, are found in high quality, cold water streams)
		% Oligochaeta individuals ≥ 30%	

3.4 Model Performance

Overall, the quantitative models were 95% accurate in replicating the panel assessments within ± 0.5 BCG level for both the calibration and confirmation data sets (Table 14). No BCG model outputs were more than 1 level off from the panelist median calls. When broken down across BCG classes -

Lower gradient/lower elevation. Of the 37 calibration samples, 33 (89%) were exact matches. The remaining 4 samples were within 0.5 levels, with slight bias towards the model scoring samples worse. However, the confirmation dataset showed the opposite pattern. Seven of the 10 samples were exact matches, and for the remaining three, the model scored samples 0.5 to 1 level better than panelist median.

Higher gradient/lower elevation. Of the 42 calibration samples, 36 (86%) were exact matches. For the remaining six, the model scored samples 0.5 to 1 level worse. In the confirmation dataset, 19 of the 21samples (90.5%) were exact matches. The two remaining samples were split, with the model scoring one better and one worse than the panelist median.

Higher gradient/higher elevation. The calibration dataset had the lowest percentage of exact matches (23 of 32 = 72%) across classes. Model bias was not evident in the remaining samples. Of the nine samples, the model scored four worse and five better than the panelist median. In contrast to the calibration dataset, this class had the highest percentage of exact matches in its confirmation dataset (12 of 13 = 92.3%). A likely contributing factor was that the calibration dataset included some sites from unique Level 4 ecoregions like the Okanogan Pine/Fir Hills that had low mean annual precipitation (< 1000 mm, or in some cases, < 650 mm). Those sites were not included in the confirmation dataset, which likely improved model performance.

Of the 25 anomalous samples (samples where the BCG model and panelist consensus were not exact matches), 18 were mismatched by ± 0.5 BCG level. The panel felt that a difference of ± 0.5 BCG level was small and did not think it was necessary to modify the rules to try and get all the results to align exactly. In fact, some warned against this, expressing concerns about overfitting the BCG model. Table 15 contains a list of the samples that did not match exactly. When the model probability of membership and panelist pluses/minuses are considered, most of the differences are smaller than Table 14 implies. For example, Sample Tt_654 had a panelist median score of 4+; based on the model output, it has 0.7 probability of membership in level 3 and 0.4 probability of membership in level 4. So while not an exact match, the difference is small.

Table 14. Performance of BCG models, grouped by calibration and confirmation samples, and BCG class: lower gradient/lower elevation; higher gradient/lower elevation; and higher gradient/higher elevation. "Better" and "worse" indicate model assessment of stream condition compared to panelist median scores (e.g., "better" if model assessed BCG level 2, but panel assessed BCG level 3, and so forth).

BCG class	Dataset	Unit	Difference					
			Model 1 level better	Model 1/2 level better	Exact match	Model 1/2 level worse	Model 1 level worse	*Total*
Low Gradient Low Elevation	Calibrate	Number		1	33	3		***37***
		Percent		2.7%	89.2%	8.1%		***100%***
	Confirm	Number	1	2	7			***10***
		Percent	10.0%	20.0%	70.0%			***100%***
High Gradient Low Elevation	Calibrate	Number		1	36	3	2	***42***
		Percent		2.4%	85.7%	7.1%	4.8%	***100%***
	Confirm	Number		1	19		1	21
		Percent		4.8%	90.5%		4.8%	***100%***
High Gradient High Elevation	Calibrate	Number	3	2	23	4		***32***
		Percent	9.4%	6.3%	71.9%	12.5%		***100%***
	Confirm	Number			12	1		***13***
		Percent			92.3%	7.7%		***100%***
Total	***Calibrate***	***Number***	***3***	***4***	***92***	***10***	***2***	***111***
		Percent	***2.7%***	***3.6%***	***82.9%***	***9.0%***	***1.8%***	***100%***
	Confirm	***Number***	***1***	***3***	***38***	***1***	***1***	***44***
		Percent	***2.3%***	***6.8%***	***86.4%***	***2.3%***	***2.3%***	***100%***

Table 15. Samples in the PNMR dataset where the BCG model output did not match exactly with the panelist consensus.

BCG class	Organization	Tetra Tech ID	Panelist median	Model	Model primary membership	Model secondary membership	Difference between panelist median and model
Low Gradient Low Elevation	King County	Tt_450	4/5 tie	4-	4 (0.6)	5 (0.4)	model 1/2 level better
	King County	Tt_654	4+	3-	3 (0.7)	4 (0.3)	model 1 level better
	WA ECY	Tt2_185	3/4 tie	3-	3 (0.7)	4 (0.3)	model 1/2 level better
	ODEQ	Tt_1255	3-	3/4 tie	3 (0.56)	4 (0.44)	model 1/2 level worse
	ODEQ	Tt_735	2-	2/3 tie	2 (0.59)	3 (0.41)	model 1/2 level worse
	ODEQ	Tt_919	3-	3/4 tie	3 (0.5)	4 (0.5)	model 1/2 level worse
	ODEQ	Tt_993	6	5/6 tie	5 (0.52)	6 (0.48)	model 1/2 level better
High Gradient Low Elevation	ODEQ	Tt_1019	3+	2/3 tie	3 (0.54)	2 (0.46)	model 1/2 level better
	WA ECY	Tt2_108	2-	3+	3 (0.77)	2 (0.23)	model 1 level worse
	WA ECY	Tt2_367	4-	5+	5 (0.70)	4 (0.30)	model 1 level worse
	ODEQ	Tt_1355	3+	2/3 tie	2 (0.5)	3 (0.5)	model 1/2 level better
	WA ECY	Tt2_188	3-	4+	4 (0.6)	3 (0.4)	model 1 level worse
	ODEQ	Tt_1200	2/3 tie	3+	3 (0.86)	2 (0.14)	model 1/2 level worse
	King County	Tt_500	3-	3/4 tie	4 (0.59)	3 (0.41)	model 1/2 level worse
	ODEQ	Tt_1180	2/3 tie	3+	3 (0.62)	2 (0.38)	model 1/2 level worse
High Gradient High Elevation	WA ECY	Tt2_125	3/4 tie	4	4 (1)	NA	model 1/2 level worse
	WA ECY	Tt2_582	2-	2/3 tie	2 (0.5)	3 (0.5)	model 1/2 level worse
	WA ECY	Tt2_713	3/4 tie	3-	3 (0.77)	4 (0.23)	model 1/2 level better
	WA ECY	Tt2_628	4-	4	4 (0.9)	5 (0.1)	model 1/2 level better
	WA ECY	Tt2_552	4	3-	3 (0.7)	4 (0.3)	model 1 level better
	ODEQ	Tt_1395	2-	2/3 tie	2 (0.59)	3 (0.41)	model 1/2 level worse
	WA ECY	Tt2_715	4	4/5 tie	4 (0.5)	5 (0.5)	model 1/2 level worse
	WA ECY	Tt2_227	4	4/5 tie	4 (0.5)	5 (0.5)	model 1/2 level worse
	WA ECY	Tt2_209	3	2-	2 (0.7)	3 (0.3)	model 1 level better
	WA ECY	Tt2_131	3+	2-	2 (0.73)	3 (0.27)	model 1 level better

3.5 Panelist Variability

Overall, there were fairly high levels of agreement in the individual assessments. For the calibration samples, 82.3% of individual assessments were within 0.5 BCG level of the group median, and 96.9% were within 1 BCG level (Table 16, Figure 15). Level of agreement was even higher for the confirmation samples: 92% of ratings were within 0.5 BCG level of the panel median, and 98.9% were within 1 BCG level. This was expected, given how panelists had gained more experience and familiarity with the process and samples at that point. Level of agreement was fairly consistent across BCG levels, although in the calibration dataset, Level 4 samples had slightly lower agreement (~75% of the scores were within 0.5 BCG levels of the group median versus > 80% in the other nominal levels; this is not unexpected, given how distinguishing between samples in the middle of the stressor gradient tends to be more challenging than samples on the extreme ends of the stressor gradient. Some of the BCG 'tie' levels (e.g., 3/4 tie and 4/5 tie) had lower levels of agreement as well, but those were driven by low sample sizes (<10 samples).

Table 16. Individual panelist scores were placed into five 'bins' based on the difference between the individual scores and panelist median scores. The numbers in this table represent the percentage of individual panelist scores falling within each 'bin.' Scores are based on the plus/minus scoring scheme[5]. Number of panelists per sample varied, ranging from 4 to 9.

Dataset	# Individual panelist scores	Panelist median BCG level	Difference between individual panelist scores and panelist median scores				
			0	± 0.1 to 0.5	± 0.6 to 1.0	± 1.1 to 1.5	± 1.6 to 2
Calibration	156	2	38.5 %	55.8 %	5.8 %	0.0 %	0.0 %
	34	2/3 tie	44.1 %	55.9 %	0.0 %	0.0 %	0.0 %
	261	3	34.5 %	48.7 %	14.9 %	1.5 %	0.4 %
	8	3/4 tie	37.5 %	50.0 %	12.5 %	0.0 %	0.0 %
	114	4	36.0 %	38.6 %	20.2 %	3.5 %	1.8 %
	7	4/5 tie	14.3 %	57.1 %	28.6 %	0.0 %	0.0 %
	150	5	42.7 %	39.3 %	14.7 %	2.7 %	0.7 %
	8	5/6 tie	0.0 %	62.5 %	25.0 %	12.5 %	0.0 %
	85	6	43.5 %	42.4 %	9.4 %	3.5 %	1.2 %
Total	***823***	***Overall mean***	***32.3 %***	***50.0 %***	***14.6 %***	***2.6 %***	***0.4 %***
Confirmation	76	2	44.7 %	51.3 %	3.9 %	0.0 %	0.0 %
	136	3	37.5 %	54.4 %	8.1 %	0.0 %	0.0 %
	15	3/4 tie	26.7 %	66.7 %	6.7 %	0.0 %	0.0 %
	67	4	31.3 %	52.2 %	14.9 %	1.5 %	0.0 %
	9	4/5 tie	33.3 %	66.7 %	0.0 %	0.0 %	0.0 %
	20	5	40.0 %	45.0 %	15.0 %	0.0 %	0.0 %
	17	6	52.9 %	41.2 %	0.0 %	5.9 %	0.0 %
Total	***340***	***Overall mean***	***38.1 %***	***53.9 %***	***6.9 %***	***1.1 %***	***0.0 %***

[5]Plus/minus scores were converted to numeric scores by 0.25 increments. For example, 2+ = 1.75, 2 = 2, 2- = 2.25, 2/3 tie = 2.5, 3+ = 2.75, 3 = 3, 3- = 3.25, 3/4 tie = 3.5, 4+ = 3.75, 4 = 4, and so on.

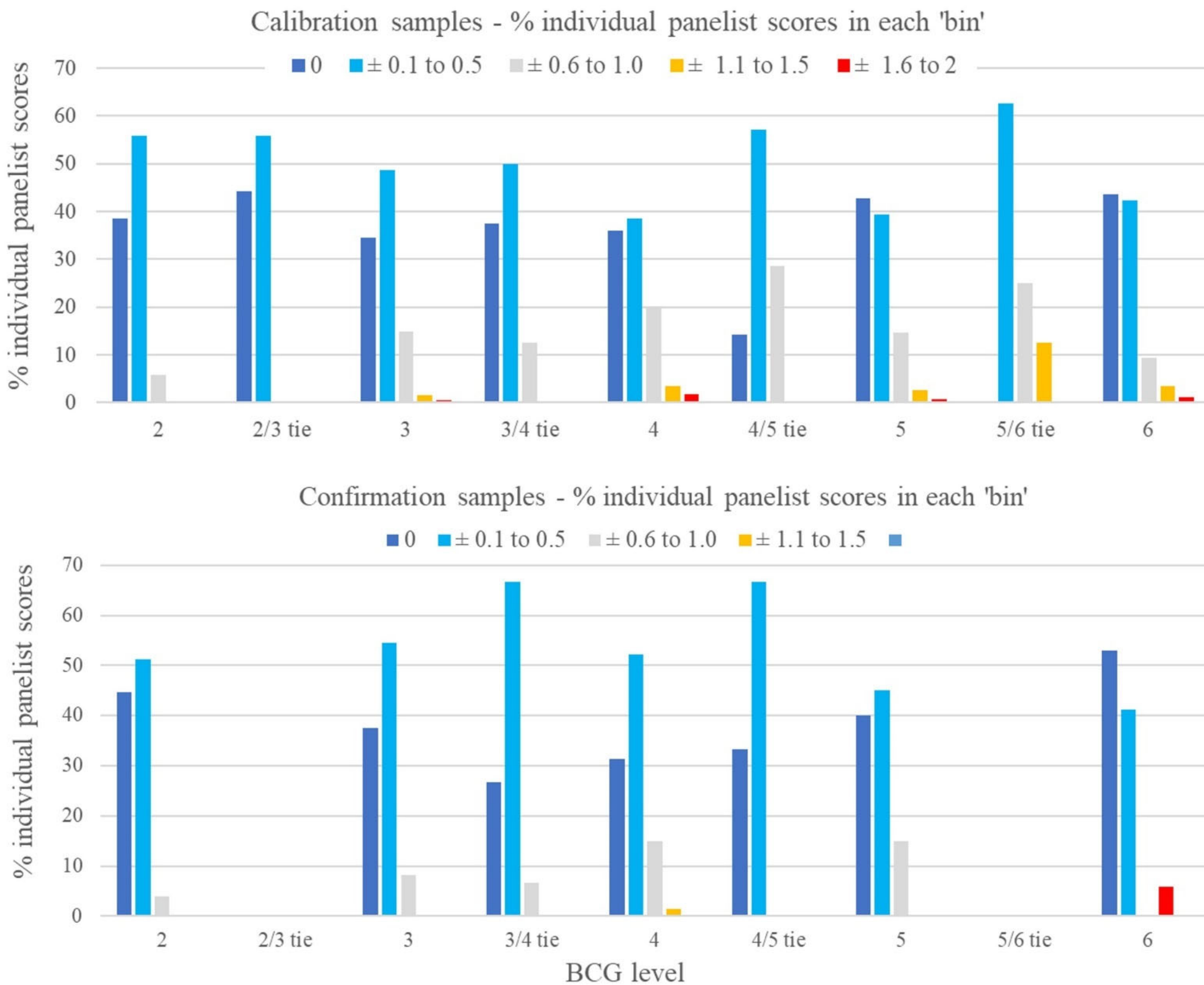


Figure 15. Distribution of individual panelist BCG scores expressed as the difference from the group median, for calibration (top) and confirmation (bottom) samples (sample size for the calibration dataset = 823; confirmation dataset = 340). Number of panelists per sample varied, ranging from 4 to 9.

3.6 Indicator Taxa

Many of the PNMR BCG rules are based on sensitive taxa metrics (Ii, II, III) (Table 12). The percent Attribute IV+V+VI non-insect taxa is also used in rules for all BCG levels and all classes. Here we highlight some of the most common taxa that go into these BCG rules and note whether any of the taxa show strong associations with a particular BCG gradient/elevation class. The full list of taxa and BCG taxa attribute category assignments can be found in Attachment B.

Sensitive (Attribute Ii+II+III) taxa

- **Attribute Ii** (sensitive, regionally endemic) – only 34 of the 615 taxa were assigned to this category. Of those, only one (Trichoptera: Pedomoecus sierra) occurred in 5% or more of the samples in the PNMR BCG calibration dataset, showing a strong association with higher gradient/higher elevation samples. The next most common was Trichoptera: Rhyacophila vetina complex, which occurred in 4.7% of the higher gradient/higher elevation samples and 2-3% of the samples from the other two gradient/elevation classes.

- **Attribute II** (highly sensitive) - 59 taxa were assigned to this category, the majority of which (46) were EPT taxa. Forty-one occurred in ≥ 5% of higher gradient/higher elevation samples. Of these, three Plecoptera (Doroneuria, Visoka cataractae and Yoraperla) occurred in ~50% or more of the samples. Attribute II taxa were less common in the other two gradient/elevation classes, with 23 occurring in ≥ 5% of the higher gradient/lower elevation samples and eight occurring in ≥ 5% in lower gradient/lower elevation samples. The following three stonefly and caddisfly taxa occurred in ≥ 5% of samples in all three gradient/elevation classes: Plecoptera: Yoraperla, Doroneuria and Perlinodes aurea; and Trichoptera: Rhyacophila hyalinata group, Apatania and Uenoidae.

- **Attribute III** (sensitive ubiquitous) – this category included 92 taxa, 66 of which occurred in at least 5% of samples in all three gradient/elevation classes. Three of the most common taxa (all of which occurred in ≥ 50% of samples in all three classes) were Trichoptera: Rhyacophila betteni group, Coleoptera: Heterlimnius corpulentus and Ephemeroptera: Neoleptophlebia/Paraleptophlebia.

Attribute IV+V+VI taxa

The most commonly occurring non-insect taxa in this group were Trombidiformes, Oligochaeta, Nemata, Juga and Sphaeriidae, which were all Attribute IV taxa. Attribute V and VI taxa were much less common than Attribute IV taxa, even when looking beyond non-insect taxa alone. Only two taxa were assigned to Attribute VI (Bivalvia: Corbicula and Gastropoda: Potamopyrgus antipodarum) and they occurred in less than 1% of the samples in all classes. Forty taxa were assigned to Attribute V. Of those, only one (Diptera: Chironomidae: Limnophyes) occurred in at least 5% of samples in all three gradient/elevation classes. Eighteen Attribute V taxa occurred in ≥ 5% of the lower gradient/lower elevation samples, and eight occurred in ≥ 5% of the higher gradient/lower elevation samples. The most common Attribute V taxa were

Trichoptera: Cheumatopsyche, Prostoma (ribbon worms) and Isopoda:Caecidotea, followed by mostly Diptera: Chironomidae (Cricotopus bicinctus group, Cryptochironomus, Procladius and Chironomus).

4 DISCUSSION

The conceptual BCG model was derived from widespread empirical experience of working aquatic ecologists from across the country (Davies and Jackson 2006). The calibration process of the index is simultaneously quantitative, empirical, and conceptual. The BCG is calibrated using a data set, but also requires ecological considerations with wide expert agreement from biologists familiar with the resources. The BCG requires descriptions of the levels from pristine to degraded. Documentation of the rationale for making BCG level determinations provides the foundation for building robust quantitative models and ensures that future information and discoveries can be related back to the baseline level descriptions. BCG levels are intended to be universal, so that a BCG level 3 assessment ideally means the same in the Pacific Northwest as it does elsewhere such as in the Midwest, Southeast or New England. The BCG taxon attribute categories are intended to apply in all regions as well (such that taxa assigned to attribute II are highly sensitive (as described in Table 1) regardless of the location).

From 2017-2019, aquatic biologists from federal, state and county agencies, taxonomists and other local experts (Table 2) partnered to calibrate a BCG for macroinvertebrate communities in low and high gradient freshwater wadeable streams in the Puget Lowlands and Willamette Valley (Stamp and Gerritsen 2019). From 2020-2022, they extended model development to the Pacific Northwest Maritime Region within Oregon and Washington state boundaries. This was a collective exercise among regional biologists to develop consensus on assessments of samples. The model was based on a fuzzy model approach. One of the strengths of the fuzzy model approach is that the set of quantitative rules that are developed are transparent and can, in principle, be followed by anyone who is making an assessment of a site. Although it may seem exotic to those not familiar with the approach, the fuzzy model rules are fully laid out and are not hidden in a statistical model or in artificial machine learning. Thus, it should be possible for someone with basic knowledge of aquatic organisms to follow the BCG rules to replicate decisions of the experts.

During the BCG process, the rules that the biologists used to assess the samples were elicited, and then a set of quantitative decision criteria rules were developed for assigning samples to BCG levels. There was fairly high agreement among the biologists performing the assessments, and very high concordance between the expert assessment and the quantitative BCG model. The panelists were able to achieve consensus despite using a variety of different approaches to arrive at their respective ratings (e.g., some focused largely on specific taxa, while others focused more on metrics (e.g., EPT, measures of trophic balance). The concordance increases our confidence in the consensus professional judgment to assess sites, and in the quantitative model to replicate that judgment.

Generation one of the Pacific Northwest Maritime Region (PNMR) BCG model supersedes the Puget Lowland/Willamette Valley BCG model that was developed during the previous phase of work. The PNMR model was calibrated for macroinvertebrate communities in three classes of freshwater wadeable streams: low gradient (<1%)/lower elevation (<750 m); higher gradient (≥1%)/lower elevation (<750 m); and higher gradient (≥1%)/higher elevation (≥750 m). When applying the BCG, users should keep in mind that they can run any samples through the model and get a result. However, if samples do not meet the criteria below, results should be interpreted with caution because they are outside the experience of the model.

Criteria:

- Freshwater wadeable streams ranging in size from 5 to 260 km^2 (2 to 100 mi^2) in five Omernik Level 3 ecoregions: Coast Range, Puget Lowland, Willamette Valley, Cascades, and North Cascades, excluding xeric regions (< 650 mm mean annual precipitation)
- 500-count samples (number of individuals should not exceed 600; if it does, random subsampling should be applied)
- At least 8 ft^2 sampling area
- Lowest practical taxonomic resolution based on current Standard Taxonomic Effort guidelines (Wisseman et al. 2015)
- Collection gear: D-Frame kick-nets or Surber net with 500-micrometer mesh net
- Collection method: targeted-riffle (in keeping with methods used by ODEQ, King County and Snohomish County) or reachwide, multihabitat (in keeping with WA ECY's protocol) (Hayslip 2007)
- Collection period: summer baseflow period (typically June through mid-October)

Results should also be interpreted with caution if they are flagged for any of the criteria listed in Table 13 (e.g., brackish influence, extreme dominance by one or two taxa).

Practitioners can use the BCGcalc R package, which is available on GitHub[6], to calculate BCG model outputs. A web-based (R Shiny) version is currently under development and is scheduled for completion by spring 2023. Users will be able to upload input files to the website, which will generate the BCG model output and sample reports without the use of R code (and does not require users to install R software on their computers). Thus, we'll have options for users who prefer to work with R code and those who do not.

Before running the BCG model, the user needs to assign samples to the proper gradient/elevation class. The source of the gradient and elevation data used for BCG model calibration was NHDPlusV2 and EPA StreamCat, respectively[7]. The gradient/elevation thresholds should be regarded as fuzzy versus distinct lines, as they represent transitional zones where streams likely share characteristics of both gradient/elevation classes. Because of this, we recommend that users run sites that are close to the thresholds through the BCG models that are closest to the thresholds and evaluate both sets of results (for example, run data for a 350-m site with 0.9%

[6] https://github.com/leppott/BCGcalc (Leppo 2018)

[7] Source of gradient data: NHDPlusV2 flowline slope from the NHDPlus Attributes table (McKay et a. 2016; https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data); source of elevation data: EPA StreamCat elevation, local catchment scale (Hill et al. 2016; https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset)

slope through both the lower gradient/lower elevation and higher gradient/lower elevation models.

We encourage biomonitoring programs to test the BCG models on their own data and provide feedback on questions such as: Are results are in keeping with expectations? Where is the model performing well? Where is it performing poorly and why? This feedback will help inform future improvements, additions and explorations, including:

- **Substrate and slope data**: if regionally consistent substrate and reach-scale slope data become available, stream classification should be revisited and potentially improved, and differences between low gradient, soft bottom and high gradient macroinvertebrate assemblages should be further evaluated.
- **Taxonomic resolution:**
    - Identify taxa for which species-level identifications are particularly important. For example, *Parapsyche*, which commonly occurs in the dataset, was not given an attribute assignment due to differences in species-level tolerances (*Parapsyche elsis* is more sensitive (Attribute II) than *Parapsyche almota* (Attribute IV)). Critical information may be lost if identifications are only taken to the genus-level for this as well as other taxa.
    - Mites and worms: for generation one of the PNMR BCG model, mites were collapsed to the Order-level and worms to subclass. Moving ahead, if greater regional consistency can be achieved with mite and worm collection and identification, and a second version of the BCG model is someday calibrated, we recommend that higher resolution data be used for these taxonomic groups.
- **Year-to-year variability**: some sites were sampled multiple years and did not have the same BCG level assignment each year (but nearly all were within ±1 BCG level). The same thing can occur with B-IBI scores, which has implications for bioassessment programs. Where multiple years of data exist, we encourage exploration of ranges in year-to-year variability and potential reasons behind the variability (such as an extreme weather event).
- **Depositional habitats**: If the opportunity arises, it may be beneficial to collect some additional data from depositional habitats. While the riffle and reach-wide collection methods suit bioassessment purposes well, it is unlikely that they capture the full diversity of taxa in depositional habitats, which limits our ability to characterize BCG level 1 in low gradient streams.

The BCG outputs can be used to supplement and enhance existing tools that practitioners currently use to assess stream health (such as the Puget Sound Lowlands B-IBI and ODEQ's PREDATOR model). The BCG can also be an effective tool for communicating resource condition to the public and for informing management decisions to protect or remediate water resources. It can allow for practical and operational implementation of multiple aquatic life uses in a state's water quality criteria and standards. For example, a number of sites were assigned to BCG level 2, which, based on participants' input, represent present-day waters in the PNMR that are closest to natural. Development of quantitative BCG models provides a technical tool for identifying and potentially protecting this region's highest quality streams, as well as developing realistic restoration goals for waters impacted by legacy activities, such as ditching, impoundments, and urban and agricultural land use.

5 LITERATURE CITED

Agresti, A. 2013. Categorical Data Analysis. 3rd Edition. John Wiley and Sons, New York.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish. Second Edition. EPA/841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Borja, A., Franco, J., Muxika, I., 2004. The Biotic Indices and the Water Framework Directive: the required consensus in the new benthic monitoring tools. Marine Pollution Bulletin 48 (3–4), 405–408

Castella, E. and M.C.D. Speight. 1996. Knowledge representation using fuzzy coded variables: an example based on the use of Syrphidae (Insecta, Diptera) in the assessment of riverine wetlands. Ecological Modelling 85:13-25.

Davies, S. B., and S. K. Jackson. 2006. The Biological Condition Gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251–1266.

Demicco, R.V. and G.J. Klir. 2004. Fuzzy Logic in Geology. Elsevier Academic Press, San Diego, CA.

Droesen, W.J. 1996. Formalisation of ecohydrological expert knowledge applying fuzzy techniques. Ecological Modelling 85:75-81.

Flotemersch, J.E., Leibowitz, S.G., Hill, R.A., Stoddard, J.L., Thomas, M.C., & Tharme, R.E. 2016. A watershed Integrity Definition and Assessment approach to Support Strategic Management of Watersheds. *River Research and Applications,* 32, 1654–1671.

Gerritsen, J., R.W. Bouchard Jr., L. Zheng, E.W. Leppo, and C.O. Yoder. 2017. Calibration of the biological condition gradient in Minnesota streams: a quantitative expert-based decision system. Freshwater Science, 36(2), 427-451.

Gerth, W.J. and A.T. Herlihy. 2006. The effect of sampling different habitat types in regional macroinvertebrate bioassessment surveys. Journal of the North American Benthological Society 25:501–512.

Gotelli, N.J. and G.R. Graves. 1996. Null models in ecology. Washington, DC: Smithsonian Institution Press.

Hafele, R., Plotnikoff, R. and R. Wisseman. 2022. Disturbance History and Natural Condition Assessment for the Pacific Northwest Maritime Region. Prepared for U.S. Environmental Protection Agency.

Hausmann, S., Charles D.F., Gerritsen J., and T.J. Belton. 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Sci Total Environ. 562:914-927. doi: 10.1016/j.scitotenv.2016.03.173.

Hayslip, Gretchen, editor, 2007. Methods for the collection and analysis of benthic macroinvertebrate assemblages in wadeable streams of the Pacific Northwest. Pacific Northwest Aquatic Monitoring Partnership, Cook, Washington. www.pnamp.org/web/workgroups/SC/meetings/2007_0821/2007_0531PNAMP_macroinvert_protocol_final.pdf

Hill, R.A., Weber, M.H., Leibowitz, S.G., Olsen, A.R., & Thornbrugh, D.J. 2016. The Stream-Catchment (StreamCat) dataset: a database of watershed metrics for the conterminous United States. *Journal of the American Water Resources Association*, 52, 120-128.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354.

Hubler, S. 2008. PREDATOR: Development and use of RIVPACS-type macroinvertebrate models to assess the biotic condition of wadeable Oregon streams. DEQ08-LAB-0048-TR. Available online:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.366.4212&rep=rep1&type=pdf

Ibelings, B.W., M Vonk, H.F.J. Los, D.T. Van Der Molen, and W.M. Mooij. 2003. Fuzzy modeling of Cyanobacterial surface waterblooms: validation with NOAA-AVHRR satellite images. Ecological Applications 13:1456-1472.

Johnson, Zachary & G. Leibowitz, Scott & Hill, Ryan. (2018). Revising the index of watershed integrity national maps. Science of The Total Environment. 10.1016/j.scitotenv.2018.10.112.

King County. 2014. Recalibration of the Puget Lowland Benthic Index of Biotic Integrity (B-IBI). Prepared by Jo Opdyke Wilhelm, (Water and Land Resources Division [WLRD]); Leska Fore (Statistical Design), Deb Lester (WLRD) and Elene Dorfmeier (WLRD). Seattle, Washington

Klir, G.J. 2004. Fuzzy Logic: A Specialized Tutorial. In Fuzzy Logic in Geology, in R.V. Demicco and G.J. Klir (eds.), pp. 11-61. Elsevier Academic Press, San Diego, CA.

Kopf, R.K., C.M. Finlayson, P. Humphries, N.C. Sims, and S. Hladyz. 2015. Anthropocene baselines: Assessing change and managing biodiversity in human-dominated aquatic ecosystems. Bioscience. 65(8): 798-811 (August 2015).

Larson, C.A., Merritt, G., Janisch, J., Lemmon, J., Rosewood-Thurman, M., Engeness, B., Polkowske, S., Onwumere, G. (2019). The first statewide stream macroinvertebrate bioassessment in Washington State with a relative risk and attributable risk analysis for multiple stressors. *Ecological Indicators*, *102*, 175-185. https://doi.org/10.1016/j.ecolind.2019.02.032

Leppo, E. 2018. BCGcalc R package (https://github.com/leppott/BCGcalc)

Leppo, E. 2020. BioMonTools R package (https://github.com/leppott/BioMonTools)

May, C., C. Luce, J. Casola, M. Chang, J. Cuhaciyan, M. Dalton, S. Lowe, G. Morishima, P. Mote, A. Petersen, G. Roesch-McNally, and E. York, 2018: Northwest. In Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp. 1036–1100. doi: 10.7930/NCA4.2018.CH24

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., Reah, A., 2012. NHDPlus Version 2: User Guide. U.S. Environmental Protection Agency. https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data

Omernik, J.M. 1987. Ecoregions of the Conterminous United States. Annals of the Association of American Geographers 77(1): 118-125.

Paul, M.J. et al. 2020. Characterizing benthic macroinvertebrate and algal biological condition gradient models for California wadeable Streams, USA. Ecological Indicators 117: 106618

Puget Sound Stream Benthos B-IBI. Accessed 28 September 2022. https://pugetsoundstreambenthos.org/

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rehn, A., Ode, P. & Hawkins, C. 2007. Comparisons of targeted-riffle and reach-wide benthic macroinvertebrate samples: Implications for data sharing in stream-condition assessments. Journal of the North American Benthological Society. 26. 0-0. 10.1899/0887-3593(2007)26[332:COTARB]2.0.CO;2.

Stamp, J., and J. Gerritsen. 2019. Calibration of the Biological Condition Gradient (BCG) for Macroinvertebrate Assemblages in Puget Lowland/Willamette Valley Freshwater Wadeable Streams. Prepared for US EPA Office of Science and Technology and US EPA Region 10.

Steedman, R.J. 1994. Ecosystem health as a management goal. Journal of the North American Benthological Society. 13(4):605–610 ter Braak, CJF; Looman, CWN. 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio 65:3–11.

Thornbrugh, D. J., Leibowitz, S.G., Hill, R. A., Weber, M. H., Johnson, Z.C. Olsen, A. R., Flotemersch, J. E., Stoddard, J. L., & Peck, D. V. 2018. Mapping watershed integrity for the conterminous United States. *Ecological Indicators*, 85, 1133-1148.

USEPA (US Environmental Protection Agency). 2016. A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems. Office of Water, Washington DC. EPA 842-R-16-001.

Washington Department of Ecology (WA ECY). 2010. Quality Assurance Monitoring Plan - Ambient Biological Monitoring in Rivers and Streams: Benthic Macroinvertebrates and Periphyton. Publication No. 10-03-109. Available online: https://fortress.wa.gov/ecy/publications/summarypages/1003109.html

Weisberg, S.B., B. Thompson, J.A. Ranasinghe, D.E. Montagne, D.B. Cadien, D.M. Dauer, D. Diener, J. Oliver, D.J. Reish, R.G. Velarde, and J.Q. Word. 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. *Ecological Indicators* 8:389–394.

Wisseman, R., Sullivan, S., Pfeiffer, J. and S. Salter. 2015. Northwest Standard Taxonomic Effort. https://www.pnamp.org/project/northwest-standard-taxonomic-effort

Wisseman, R. 2020. Biological Condition Gradient (BCG) Level 1 Biodiversity Index - Draft for review by the Pacific NW BCG Expert Panel.

Yuan, Lester. 2006. Estimation and Application of Macroinvertebrate Tolerance Values. Report No. EPA/600/P-04/116F. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.