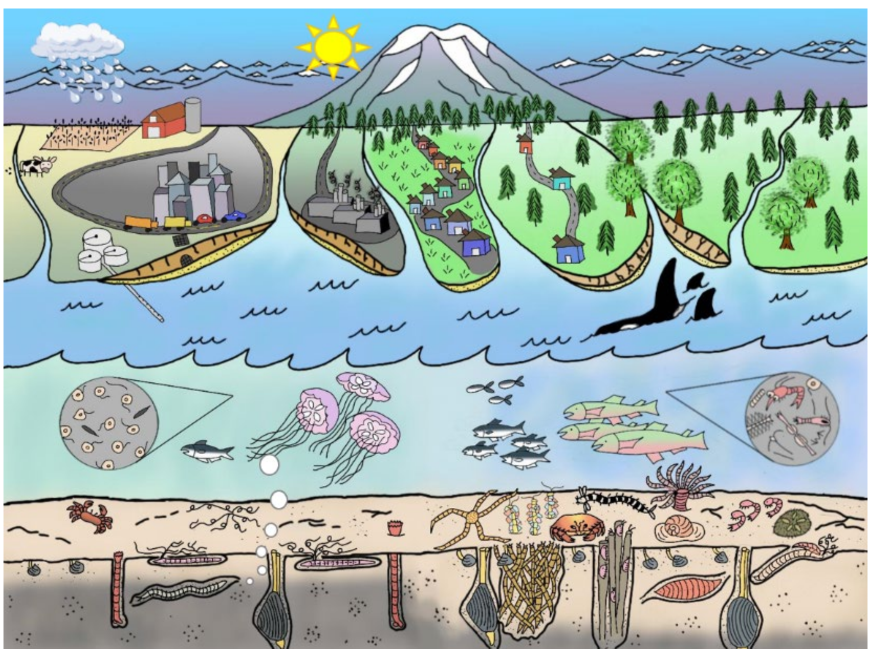
DEPARTMENT OF
ECOLOGY
State of Washington

Quality Assurance Project Plan

Puget Sound *Marine Benthic Index* and Graphical Causal Model



May 2022

Publication 22-03-105

Publication Information

Each study conducted by the Washington State Department of Ecology must have an approved Quality Assurance Project Plan (QAPP). The plan describes the objectives of the study and the procedures to be followed to achieve those objectives. After completing the study, Ecology will post the final report of the study to the Internet.

This QAPP was approved to begin work in May 2022. It was finalized and approved for publication in May 2022.

The final QAPP is available on Ecology's website at https://apps.ecology.wa.gov/publications/SummaryPages/2203105.html.

Suggested Citation
Partridge, V.A. and D.R. Schoolmaster, Jr. 2022. Quality Assurance Project Plan: Puget Sound *Marine Benthic Index* and Graphical Causal Model. Publication 22-03-105. Washington State Department of Ecology, Olympia. https://apps.ecology.wa.gov/publications/SummaryPages/2203105.html.

Data for this project are available in Ecology's EIM Database. Search Study IDs:
PSAMP_HP (Historical program), PSAMP_LT (Long-Term program through 2015), PSEMP_LT (Long-Term program from 2016), PSAMP_SP (Regional program), UWI (Urban Bays program).

The Activity Tracker Code for this study is 23-002.

Contact Information

Publications Team
Environmental Assessment Program
Washington State Department of Ecology
P.O. Box 47600
Olympia, WA 98504-7600
Phone: 360 407-6764

Washington State Department of Ecology: https://ecology.wa.gov

- Headquarters, Olympia 360-407-6000
- Northwest Regional Office, Shoreline 206-594-0000
- Southwest Regional Office, Olympia 360-407-6300
- Central Regional Office, Union Gap 509-575-2490
- Eastern Regional Office, Spokane 509-329-3400

COVER ILLUSTRATION: Conceptual model of benthic invertebrate communities in relation to human and natural pressures. Illustration drawn by Margaret Dutch, digitally colored by Grace McKenney.

Any use of product or firm names in this publication is for descriptive purposes only and does not imply endorsement by the author or the Department of Ecology.

To request ADA accommodation for disabilities or printed materials in a format for the visually impaired, call the Ecology ADA Coordinator at 360-407-6831 or visit https://ecology.wa.gov/accessibility.
People with impaired hearing may call Washington Relay Service at 711.
People with speech disability may call TTY at 877-833-6341.

Quality Assurance Project Plan

Puget Sound *Marine Benthic Index* and Graphical Causal Model

by Valerie Partridge and Donald Schoolmaster Jr.

May 2022

Approved by:

Signature:	Date:
Jenna Judge, Client, Puget Sound Partnership	
Signature:	Date:
Nathalie Hamel, Client, Puget Sound Partnership	
Signature:	Date:
Valerie Partridge, Author / Project Manager / Co-Principal Investigator, EAP	
Signature:	Date:
Donald Schoolmaster Jr., Author / Co-Principal Investigator, USGS	
Signature:	Date:
Julianne Ruffner, Author's Unit Supervisor, EAP	
Signature:	Date:
Stacy Polkowske, Author's Section Manager, EAP	
Signature:	Date:
Arati Kaza, Ecology Quality Assurance Officer	

Signatures are not available on the Internet version.
EAP: Environmental Assessment Program
USGS: United States Geological Survey

1.0 Table of Contents

List of Figures

List of Tables

2.0 Abstract

Marine benthic invertebrates (benthos) are key components of the Puget Sound ecosystem, with many species providing important services such as nutrient renewal, food for other species (including human consumption), and ecosystem engineering. Benthos are valuable sentinels of ecosystem health because of their direct association living in, and sometimes consuming, sediments. Their sedentary lifestyle means they are unable to escape exposure to stressors such as contaminated sediments, organic matter enrichment, habitat alteration, ocean acidification, and low dissolved oxygen in the water. Some species are intolerant of such stressors, while other species may thrive in such challenging conditions. In addition, the relatively short lifespan of many species mean that their responses to stressors may be reflected as changes to the community structure and function in as little as one year or less.

Indicators of benthic invertebrate community health can serve as direct measures of sediment and water quality. These can help to answer key questions such as:

- What is the current condition of the benthic habitat, including sediments and porewater and their associated invertebrate assemblages?
- How does benthic condition change over time in response to natural and human disturbances, especially as climate change modifies the ecosystem?

However, no widely accepted benthic infaunal indices equivalent to those developed elsewhere (e.g., AMBI) have yet been adopted for sediment regulatory or ambient monitoring work in Puget Sound.

As part of the June 2020 revision of the Puget Sound Vital Signs, the Puget Sound Partnership adopted an as-yet-unspecified *Marine Benthic Index* as an indicator of the Marine Water Vital Sign to fill a recognized gap in indicators of Puget Sound health. This project will develop not only the *Marine Benthic Index* to assess and communicate status and trends of benthic invertebrate community health, but also a graphical causal model that will be used to estimate linkages and test hypotheses of causation.

3.0 Background

3.1 Introduction and problem statement

Benthic invertebrates are key components of marine ecosystems, such as Puget Sound. Many species provide crucial ecosystem services such as nutrient renewal and prey for higher trophic levels, including culturally and commercially important species ranging from oysters and geoduck to salmon. However, changes to the environment may alter the ability of the ecosystem to provide ecosystem services and resources (WSAS, 2012).

Benthos are valuable sentinels of ecosystem health because of their direct association living in, and sometimes consuming, sediments (Dauvin et al., 2010). Their sedentary lifestyle means they are unable to escape stressors such as contaminated sediments, organic matter enrichment, habitat alteration, ocean acidification, and low dissolved oxygen in the water. Some species are intolerant of such stressors, while other species may thrive in such challenging conditions. Many benthic species have relatively short lifespans, implying that their responses to stressors may be reflected as changes to the community structure and function in as little as one year.

The need for a Puget Sound-specific measure of benthic health as an indicator of habitat condition has long been recognized (e.g., Levin et al., 2011; Ranasinghe et al., 2013; O'Neill et al., 2018). In addition to indicating effects of human disturbances on biota, a well-designed benthic index would have multiple downstream benefits, such as:

- Improving understanding of the entire ecosystem.
- Informing science-based best management practices for improving water quality.
- Understanding changes that affect tribal treaty rights to marine resources.
- Clean Water Act §303(d)/§305(b) reporting.

As part of the June 2020 revision of the Puget Sound Vital Signs, the Puget Sound Partnership (PSP) adopted an as-yet-unspecified *Marine Benthic Index* as an indicator of the Marine Water Vital Sign to fill this recognized gap in indicators of Puget Sound health (McManus et al., 2020).

Although tools for assessing ecological quality based on benthic invertebrates have been put into practice elsewhere (e.g., Borja et al., 2019), such tools have not been adapted for Puget Sound. A decade ago, Ecology's Marine Sediment Monitoring Team (MSMT) had five off-the-shelf benthic indices calibrated for use in Puget Sound, but the validation was limited (Ranasinghe et al., 2013) and the indices were therefore never employed. Other attempts to develop methods for assessing benthic community health are detailed in Appendix A.

In the absence of a widely accepted benthic indicator for Puget Sound, the MSMT has for many years been classifying invertebrate assemblages as either "*adversely affected*" or "*unaffected*" in a binary Sediment Benthic Index (SBI) for Puget Sound. The determination is based on best professional judgment of a suite of numeric abundance and diversity measures calculated for each benthos sample, along with the presence (or absence) and abundance of stress-tolerant and stress-sensitive species (Dutch et al., 2018). The *adversely affected/unaffected* determination deliberately does not attempt to assign cause as being natural or anthropogenic. SBI categories have been mapped, and spatial extents calculated, for examination of spatial and temporal patterns in sediment quality (e.g., Weakland et al., 2018).

While informative, the binary nature of the SBI classification has revealed some limitations, including the following:

- Intermediate-condition benthic communities are forced into the categories of the two extremes, leading to potential misrepresentation.
- The SBI does not take into account the structuring effects of the habitat (the physico-chemical environment, including sediment, water, temperature, oxygen, etc.) on the benthos or attempt to determine causes for benthic condition.
- The SBI determination procedure requires a degree of expertise with Puget Sound-specific benthic invertebrate communities, which limits its applicability by others who may desire to adopt a benthic index. It is also subjective and may drift over time as the benthic experts gain experience.
- The SBI is not statistically reproducible. Efforts to replicate the binary SBI decisions of the benthic experts using the same data with statistical tools such as constrained analysis of principal coordinates (Anderson et al., 2008) have had only moderate success (Partridge, unpublished data).

Thus, a new approach is needed. We need an objective, quantitative benthic indicator which can also be used to delineate qualitative categories for conveying broad levels of benthic invertebrate condition. In addition to the criteria listed by Levin et al. (2011) for scientifically credible indicators for Puget Sound, we add these specifics for a benthic index:

- Beyond distinguishing the extremes of very disturbed and pristine conditions, it is important to be able to distinguish intermediate categories of condition (e.g., low and moderate levels of disturbance). Extremes of condition are usually readily apparent in sample assemblages. What is more difficult to determine is whether an assemblage exhibiting some degree of disturbance is only slightly disturbed or more heavily disturbed.
- We need to detect changes and trends. If healthy benthic communities are deteriorating in condition, there should be indications of shifts well before sudden change from *unaffected* to *adversely affected* under a binary index. Conversely, a change from *adversely affected* to *unaffected* could give a rosier impression of improvement in benthic condition than warranted when the binary classification is the only way to express improvement from more impacted to less impacted (but still impacted).
- We want to distinguish between natural and human-caused stressors. Environmental managers and regulators want scientifically sound information so they can apply adaptive management to reduce human-caused stressors and improve conditions. Differentiating between natural and human stressors, however, is challenging without a rigorous statistical framework.

3.2 Study area and surroundings

Details of the study area are described in the Puget Sound Sediment Monitoring Program Quality Assurance Monitoring Plan (Dutch et al., 2018), the parent QAPP for this program. We use the term "Puget Sound" here as shorthand for the marine/estuarine portion of the southern Salish Sea, from the Canadian border to the southern terminus of Puget Sound proper, east from Port Angeles, inclusive (Figure 1).

Figure 1. Puget Sound Sediment Monitoring Program study area.

3.2.1 History of study area

The history of the study area is described in the Puget Sound Sediment Monitoring Program Quality Assurance Monitoring Plan (Dutch et al., 2018).

3.2.2 Summary of previous studies and existing data

There have been multiple efforts in the past to develop benthic indices and biological-effects criteria for Washington State and Puget Sound, including those listed below. Details are provided in Appendix A.

- Ecology, 1991 – Washington State Sediment Management Standards, WAC 173-204-320(3): Marine sediment quality standards biological effects criteria.
- EPA, Puget Sound Water Quality Authority, and Ecology, 1993 – National Benthic Experts Workshop, Recommendations for assessing adverse benthic effects in Puget Sound.
- Ecology Toxics Cleanup Program Sediment Management Unit, 1995-2000 – Development of Reference Value Ranges for assessment of benthos.
- Ecology Environmental Assessment Program Marine Sediment Monitoring Team, 2003 – Development of a Sediment Benthic Index with binary categories: *unaffected*, *adversely affected* communities, based on best professional judgment after examination of univariate metrics (abundance, taxa richness, evenness, dominance) and presence/abundance of sensitive and tolerant species.
- Ecology Environmental Assessment Program Marine Sediment Monitoring Team and Southern California Coastal Water Research Project, 2008-2013 – Calibration and partial validation of five multivariate benthic indices used elsewhere for use with Puget Sound benthos.

3.2.3 Parameters of interest and potential sources

The parameters of interest are the environmental and biological variables detailed in Section 4.3, below. The sources are Ecology's Marine Monitoring Unit (MMU) data from 30+ years of long-term monitoring of Puget Sound marine waters and sediments.

3.2.4 Regulatory criteria or standards

No assessments for compliance will be conducted. The *Marine Benthic Index* is being developed for the Puget Sound Ecosystem Monitoring Program (PSEMP) sediment monitoring component, which is not part of any regulatory activity by the Department of Ecology. However, other programs in Ecology may be interested in potential use of benthic indicators developed, for regulatory purposes. For example, if a defensible marine benthic index is available, Ecology's Water Quality Program would use it for narrative Water Quality Assessments (Ecology, 2020; Brown, *pers. comm.*). In addition, a Puget Sound benthic index is of interest to the Puget Sound Nutrient Source Reduction Project for incorporation into a eutrophication index (Bilhimer, *pers. comm.*).

The PSP has included a placeholder for a *Marine Benthic Index* in its suite of revised Vital Signs[1] (McManus et al., 2020). While not regulatory, targets are set for some Vital Signs indicators for measuring progress in the revitalization of Puget Sound (PSP, 2020).

[1] This *Marine Benthic Index* is planned to be an indicator for the Marine Water Vital Sign. A separate indicator of abundance and biomass of benthic marine invertebrates is planned as an indicator for the Groundfish & Benthic Invertebrates Vital Sign (McManus et al., 2020), based on by-catch in groundfish trawls.

4.0 Project Description

4.1 Project goals

Project goals are to develop:

- A *Marine Benthic Index* suitable for use as a Puget Sound Marine Water Vital Sign indicator (McManus et al., 2020) and in PSEMP ambient sediment monitoring. The indicator must be capable of objectively and reliably distinguishing benthic communities with no, low, moderate, and high levels of disturbance. The *Marine Benthic Index* would be used to document changes over time and space in benthos assemblage structure both for individual monitoring stations and for large areas, such as Puget Sound's urban bays or all of Puget Sound.
- A graphical causal model that can be used to test causal hypotheses of effects of human disturbances, including management actions, on subtidal benthos.

4.2 Project objectives

Project objectives are listed below. More detail is provided in Section 7.

- Develop causal sub-models for benthic organisms as functions of environmental variables.
- Develop software to solve model to obtain Disturbance Index for locations. This will be the tool for assessing condition of benthic communities.
- Validate the Disturbance Index as the *Marine Benthic Index* and determine thresholds for qualitative categories and recommended target value.
- Develop graphical causal model to test hypotheses of drivers causing benthic condition and to estimate relationships among variables. This "what-if" tool will be used both for screening potential drivers and for predicting effects of management actions.
- Apply *Marine Benthic Index* to sediment monitoring data to assess condition of Puget Sound benthos.

4.3 Information needed and sources

Existing biological data, sediment physical and chemical data, and station characteristics from Ecology's Puget Sound Marine Sediment Monitoring Program (the sediment monitoring component of PSEMP) will be the primary variables used. Marine waters data, information on climate drivers (boundary conditions), river flow data, or Salish Sea model outputs may be incorporated into the causal models to be developed if we are able to determine appropriate mappings in space and time to benthos samples. Additional data sources are beyond the scope of this project.

Inclusion or exclusion of parameters listed below does not imply inclusion or exclusion of the data in the models; rather, these are lists of potential data sources, described to the level of knowledge of the authors of this QAPP. While there are undoubtedly other possible factors influencing or describing Puget Sound benthos, these are the data at hand.

4.3.1 Puget Sound marine sediment data

All of the variables listed in this subsection are already collected, and have been collected for many years, by Ecology's Marine Sediment Monitoring Program (MSMP). Over 3,300 benthic invertebrate samples have been taken from some 900 locations throughout Puget Sound since 1989. Sediment samples have been collected once a year, either in spring (generally April) for the long-term stations or in June for rotating regional and bay-wide surveys.

All of the data except benthos functional ecology, calculated variables, and a few field measurements are available from Ecology's Environmental Information Management System (EIM) database. Information on the functional ecology of the benthos is available in published journal articles and databases; examples include functional feeding guilds developed by Macdonald et al., 2010, and multiple traits housed in EPA's Coastal Biodiversity Risk Assessment Tool (CBRAT) database (Lee et al., 2015, 2017).

The variables listed in Table 1 characterize benthic invertebrate communities. All numerical values are relative to 0.1 m^2, the surface area of the van Veen grabs used to collect the samples. In addition, we will calculate numerical values standardized to sample volume, based on grab penetration depth for each sample.

Table 1. Benthic invertebrate community parameters (Dutch et al., 2018).

Category	Parameter	Purpose/Concern
Measured (each individual animal in the sample)	Taxonomic identification[2] to lowest level (to species level if possible)	Characterization of benthos assemblages
	Size class[3]	
Calculated for each sample	Count of animals for each taxonomic identification	Characterization of benthos assemblages
	Numeric benthic measures including: total abundance, major taxa abundance, taxa richness, Pielou's evenness, Swartz dominance index	Characterization of benthos assemblages
	Biomass estimates[2]	Estimation of biomass of individual organisms and whole benthos samples using biomass measurements taken from a 2016 Puget Sound-wide benthos reference collection; useful in understanding carbon budget of ecosystem
	Functional ecology (e.g., feeding guilds, bioturbation potential)	Application of information on ecological functions (e.g., feeding guilds - Macdonald et al., 2010, 2012) to each species to obtain better understanding of benthos population dynamics

[2] Nomenclature in EIM may not be kept up to date.
[3] Available only from 2016 on.

The habitat variables listed in Table 2 are known or hypothesized to affect benthic invertebrates.[4] All but the calculated variables and some field measurements are stored in EIM.

Table 2. Habitat parameters (Dutch et al., 2018).

Category	Parameter	Purpose/Concern
Location and time	Station depth	Characterization of habitat
	Station coordinates (latitude, longitude)	
	Date of sampling	
Sea bottom	Sediment grain size distribution	Characterization of habitat
	Sediment temperature[5]	
	Salinity of overlying water in sediment grab[5]	
	Grab penetration depth[5]	
Biogeo-chemistry[6] (surface sediments)	Total carbon	Determination of organic composition and quality in sediments; lability and availability of nutrients to benthos; identification of sources of organic matter; potential eutrophication exposure
	Total organic carbon	
	Total inorganic carbon (calculated)	
	Total nitrogen	
	C:N ratios (calculated)	
	δ13C and δ15N stable isotopes	Determination of relative proportion of terrestrial vs. marine organic input (i.e., nutrient sources); trophic structure
	Total sulfides	Determination of sediment quality with respect to reduced condition and toxicity to benthos
	Biogenic silica	Proxy for diatom microfossil abundance in sediments; relationship to diatom abundance in water column and food web implications

[4] This is not a complete list of variables that can affect the benthos; it is only a list of data we have in the MSMP.
[5] Field measurement, not in EIM.
[6] Except for TOC, these were not analyzed before 2017 (C, N, C:N) or 2018 (isotopes, sulfides, silica).

The sediment contaminants listed in Table 3 potentially affect benthic invertebrates. Analyses of existing data (see Section 7.3.2.4, below) will determine which, if any, may be incorporated into benthic indicators. The raw concentration data are stored in EIM.

Table 3. Sediment chemistry parameters[7] (Dutch et al., 2018).

Category	Parameter	Purpose/Concern
Chemistry (surface sediments)	Metals	Determination of degree of anthropogenic chemical contamination in bulk sediments; better understanding of benthic/pelagic food web links and contaminant transfer through the food web
	Polycyclic aromatic hydrocarbons	
	Polychlorinated biphenyls	
	Polybrominated diphenyl ethers	
	Phthalates	
Derived variables	SQS quotients	Ratios of chemical contaminants to their respective Sediment Quality Standards (Ecology, 2013)

4.3.2 Puget Sound marine waters data[8]

The marine waters variables listed in Table 4 are known or hypothesized to affect benthic invertebrates.[9] All but the calculated variables and some field measurements are stored in EIM or Ecology's Environmental Assessment Program Marine Waters (EAP-MW) database.

The Marine Waters Monitoring Program consists of 38 core stations, 30 of which are located throughout the southern Salish Sea (Bos et al., 2022), sampled monthly for discrete and continuous water quality profiling. Particulates are sampled at 10 meters and near-bottom at a subset of those locations in close proximity to 20 of the 50 current Long-term sediment monitoring stations.

The spatial, temporal, and depth pairing of marine waters and sediment stations, other than the 20 that are essentially co-located, have yet to be determined (see map in Appendix B). In general, water moves and sediment and benthos do not, so the currents will factor into the pairing. Benthos integrate environmental conditions over their lifetimes. Analyses would be required to determine what selection and form of marine waters parameters (e.g., dissolved oxygen minimum and duration) would best inform effects on the benthos.

[7] May not be available for all samples.
[8] It has not yet been determined whether to use marine waters data; this is listed as a potential data source.
[9] This is not a complete list of parameters available or that may affect the benthos; it is our limited understanding of the data that may be available from Ecology's Marine Waters Monitoring Program.

Table 4. Marine waters parameters (Keyzers et al., 2020; Bos et al., 2022).

Category	Parameter	Purpose/Concern
Location and time	Station depth	Characterization of habitat
	Station coordinates (latitude, longitude)	
	Date of sampling	
Water column (continuous profile)[10]	Chlorophyll fluorescence	Characterization of habitat
	Conductivity or salinity	
	Density (calculated)	
	Dissolved oxygen	
	Light transmission (NTU)	
	Nitrate (SUNA sensor)	
	PAR (photosynthetically active radiation)	
	Pressure	
	Stratification (calculated)	
	Temperature	
	Turbidity	
Water column (discrete depths)[10]	Nitrate	Characterization of habitat, including nutrients and biogeochemistry
	Nitrite	
	Ammonium	
	Orthophosphate	
	Silicate	
	Chlorophyll-*a*	
	Salinity	
Particulates[11] (10-m and near-bottom)	Total organic carbon	Characterization of habitat, including nutrients and biogeochemistry
	Total nitrogen	
	Particulate nitrogen	
	Particulate organic carbon	
	C:N ratios (calculated)	

[10] Not all parameters available for all stations, all years, or all depths.
[11] Available only for 20 stations and only from 2016 on.

4.3.3 Salish Sea Model output[12]

The Salish Sea Model (Khangaonkar et al., 2018) is a tool used to predict impacts of nutrient sources, ocean acidification, or climate change on dissolved oxygen in the Puget Sound region (McCarthy et al., 2018; Pelletier et al., 2017a,b; Ahmed et al., 2019, 2021). The resolution of the model is 10 layers x thousands of nodes (Ahmed et al., 2019). Model predictions may be useful for filling gaps for sediment monitoring stations not co-located or within the same water masses as the marine waters monitoring stations. Analyses would be required to determine appropriate pairing of essentially continuous modeled waters results with annual benthos samples.

Available model outputs of potential interest for use could include variables such as water residence time and estimated sediment fluxes, as well as bottom-water dissolved oxygen.

Another use of outputs from the Salish Sea Model might be as reasonableness checks, e.g., as a qualitative check of whether benthos predictions are consistent with water-condition predictions.

4.3.4 Boundary conditions data[13]

Benthic habitats and invertebrates are affected by large-scale drivers such as weather, hydrology, and climate.[14] As with the marine waters monitoring data, analyses would be required to determine appropriate pairing of monthly values with annual benthos samples.

Publicly-available data sources of oceanographic indices, large river flows, and weather, already used by Ecology's Marine Waters Monitoring Program, are listed in Appendix C.

4.4 Tasks required

Tasks required are outlined below. Details are given in Section 7.3.2.

- Choose independent calibration and validation datasets.
- Prepare the data.
- Determine which commonly collected benthic taxa to use.
- Determine causal sub-models for the chosen taxa.
- Select prior distributions for fitted parameters.

4.5 Systematic planning process

Not applicable — QAPP is sufficient.

---

[12] It has not yet been determined whether to use Salish Sea Model results; this is listed as a potential data source.
[13] It has not yet been determined whether to use boundary conditions data; these are listed as potential data sources.
[14] This is not a complete listing of drivers or datasets; it is our limited understanding of data that are readily available and already used by Ecology's Marine Waters Monitoring Program.

5.0 Organization and Schedule

5.1 Key individuals and their responsibilities

Table 5 shows the responsibilities of those who will be involved in this project.

Table 5. Organization of project staff and responsibilities.

Staff	Title	Responsibilities
Jenna Judge Puget Sound Partnership	Client	Clarifies scope of project. Provides external review of QAPP and approves final QAPP. Funds portion of project. Reviews Vital Sign report.
Nathalie Hamel Puget Sound Partnership	Client	Clarifies scope of project. Provides external review of QAPP and approves final QAPP. Funds portion of project. Reviews Vital Sign report.
Valerie Partridge MMU, WOS	Project Manager, Co-Principal Investigator	Leads project. Writes QAPP. Conducts QA review of data. Organizes and leads validation exercise. Co-authors manuscript. Drafts Vital Sign report.
Donald Schoolmaster Jr. Wetland and Aquatic Research Center, U.S. Geological Survey	Contractor, Co-Principal Investigator	Co-authors QAPP. Develops Disturbance Index and R code. Develops graphical causal model and R code. Writes manuscript.
Sandra Weakland MMU, WOS	Benthic Ecologist	Database management, EIM data entry; validation of *Marine Benthic Index*
Dany Burgess MMU, WOS	Taxonomist	Benthic invertebrate expert; validation of *Marine Benthic Index*
Mike McHugh Tulalip Tribes	Tribal Biologist	Validation of *Marine Benthic Index*
Oliver Miler Northwest Indian Fisheries Commission	Tribal Biologist	Validation of *Marine Benthic Index*
Tommy Moore Northwest Indian Fisheries Commission	Tribal Biologist	Validation of *Marine Benthic Index*
Blair Paul Skokomish Tribe	Tribal Biologist	Validation of *Marine Benthic Index*
Other biologists (not yet identified)	Benthic Ecologists	Validation of *Marine Benthic Index*
Interested scientists and stakeholders from multiple agencies	Technical Advisory Group (TAG)	Peer review
Julianne Ruffner MMU, WOS	Unit Supervisor for Project Manager	Provides internal review of the QAPP, approves the budget, and approves the final QAPP.
Stacy Polkowske WOS	Section Manager for Project Manager	Reviews the project scope and budget, tracks progress, reviews the draft QAPP, and approves the final QAPP.
Arati Kaza	Ecology QA Officer	Reviews and approves the draft QAPP and the final QAPP.

MMU – Marine Monitoring Unit
WOS – Western Operations Section
QAPP – Quality Assurance Project Plan

5.2 Special training and certifications

Very specific expertise is required for the tasks involved in this project:

- Experience developing and using environmental models: Donald Schoolmaster
- Statistics: Donald Schoolmaster
- Benthic invertebrate taxonomy: Dany Burgess
- Benthic ecology: all
- R programming: Donald Schoolmaster

5.3 Organization chart

Figure 2 shows the organizations that will be involved in this project.

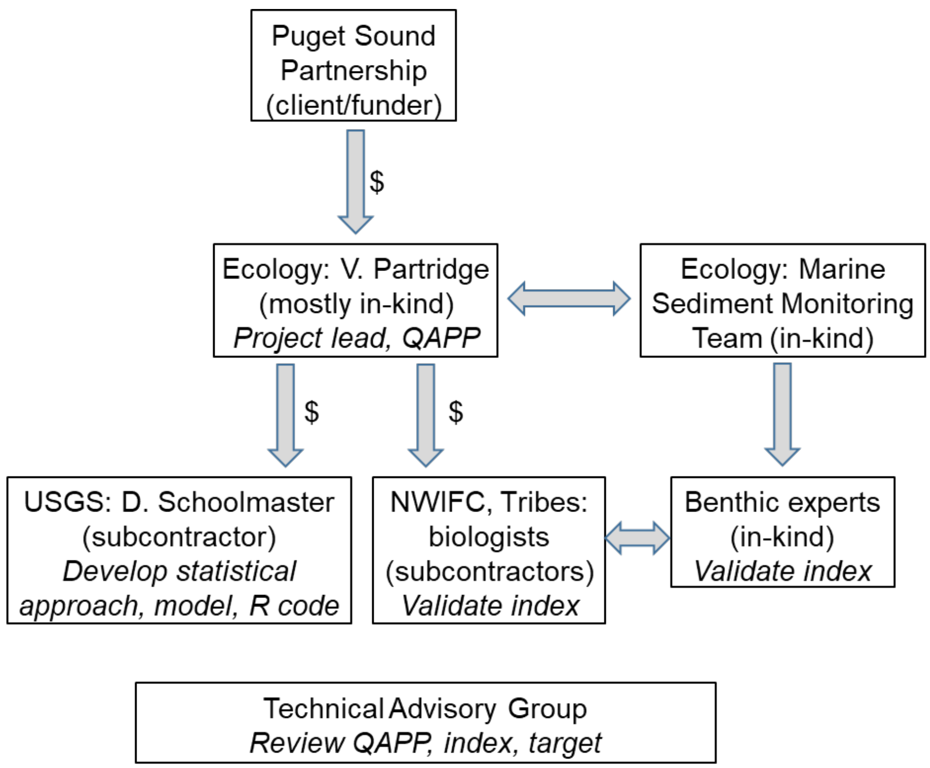


Figure 2. Organization chart for this project.

5.4 Proposed project schedule

Table 6 lists key activities, due dates, and lead staff for this project.

Table 6. Proposed schedule for completing work and reports.

Work type	Due date	Lead staff
Field and laboratory work		
Field work completed	NA – no new field work required for this study	
Laboratory analyses completed	NA – no new lab analyses required for this study	
Environmental Information System (EIM) database		
EIM data loaded	NA – this study will use existing data, already in EIM, and will not create data	
EIM data entry review		
EIM complete		
Preparation		
Determine habitat and assemblage types	3/31/2022	Marine Sediment Monitoring Team, D. Schoolmaster
Define reference conditions		
Choose independent calibration and validation datasets		
Gather data from multiple databases	3/31/2022	V. Partridge
Prepare (scrub, QA) data	3/31/2022	V. Partridge
Develop Disturbance Index (USGS)		
Refine statistical theory	6/15/2022	D. Schoolmaster
Program R code		
Test code as user	6/15/2022	V. Partridge
Finalize R code	6/30/2022	D. Schoolmaster
Validation of Disturbance Index for use as *Marine Benthic Index* (Ecology)		
Plan meetings	7/31/2022	V. Partridge
Provide Disturbance Index results for validation dataset	7/31/2022	D. Schoolmaster
Anonymize validation dataset	7/31/2022	V. Partridge
Send data and instructions to participants	8/15/2022	V. Partridge
Assess condition of invertebrate communities in anonymized samples, rank communities, assign condition category to each community	9/15/2022	Marine Sediment Monitoring Team, Tribal biologists, other benthic ecology experts
Analyze results to determine correspondence between calculated index and experts' judgments	9/30/2022	V. Partridge
If needed, hold follow-up Kappa exercise to improve agreement amongst the experts	10/15/2022	Marine Sediment Monitoring Team, Tribal biologists, other benthic ecology experts
Analyze final results	10/31/2022	V. Partridge
Determine thresholds for categories based on statistical analysis of the index values and the expert judgments	10/31/2022	V. Partridge
Recommend target value = lower limit of no-disturbance category	10/31/2022	V. Partridge
Develop graphics for displaying both the quantitative and qualitative results for individual locations and for large areas	10/31/2022	Marine Sediment Monitoring Team
Draft report	11/30/2022	V. Partridge

Work type	Due date	Lead staff
Review *Marine Benthic Index* as Marine Water Vital Sign indicator	1/31/2023	Technical Advisory Group (TAG)
Develop Graphical Causal Model (USGS, while Ecology is in Validation exercise)		
Specify model with direct and indirect relationships between the benthos and the environment	11/30/2022	Marine Sediment Monitoring Team, D. Schoolmaster
Refine statistical theory	12/31/2022	D. Schoolmaster
Program R code to calculate Disturbance Index for locations		
Test code as user	12/31/2022	V. Partridge
Finalize R code	1/31/2023	D. Schoolmaster
Draft report describing model, with a worked example	2/28/2023	D. Schoolmaster
Train MSMT how to use Graphical Causal Model to perform hypothesis-testing	2/28/2023	D. Schoolmaster
Train MSMT how to update parameters in R code for Disturbance Index		
Communication		
Draft manuscript for submission to scientific journal	4/30/2023	D. Schoolmaster, V. Partridge
Draft *Marine Benthic Index* indicator status and trends report for Vital Signs report and website	4/30/2023	Marine Sediment Monitoring Team
Review *Marine Benthic Index* indicator report	5/15/2023	Technical Advisory Group (TAG)
Revise *Marine Benthic Index* indicator report based on reviews	5/31/2023	Marine Sediment Monitoring Team

NA - Not Applicable

Table 7. Schedule for final report.

Task	Due date	Lead staff
Draft to supervisor	4/30/2023	V. Partridge
Draft to client/ peer reviewer	4/30/2023	
Draft to external reviewers	4/30/2023	
Final draft to publications team	5/31/2023	
Final report due on web	7/31/2023	

Figure 3 shows a timeline of key activities and participating groups for this project.

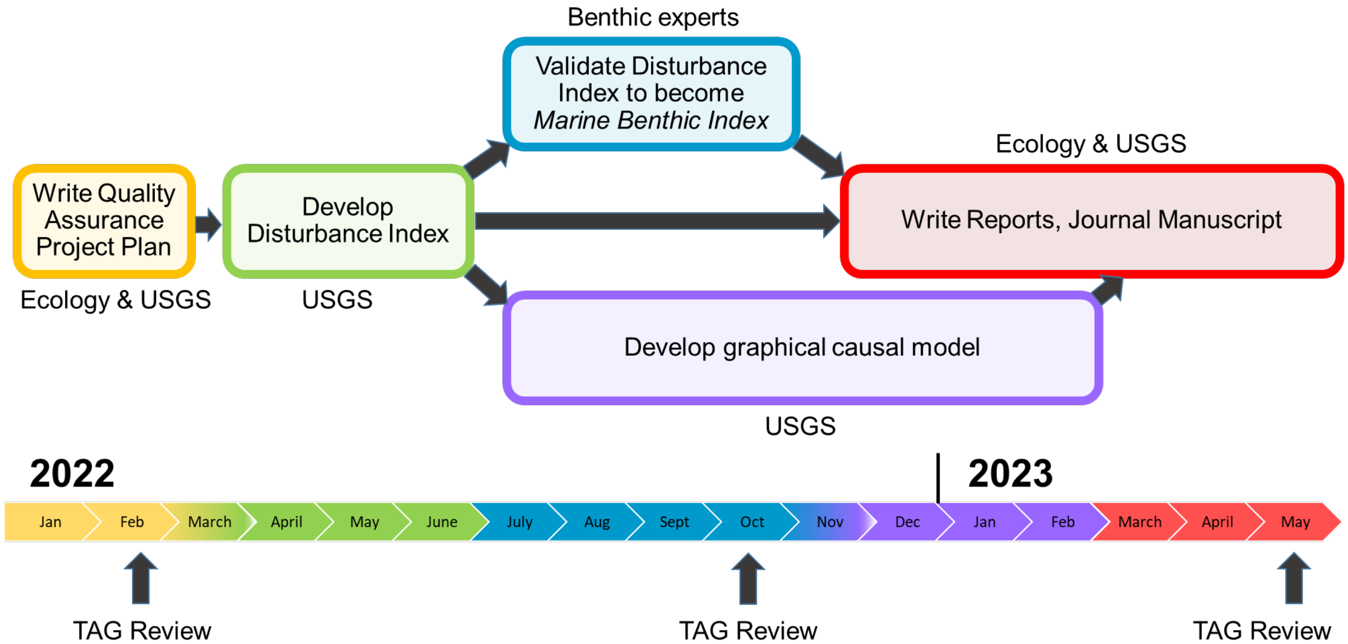


Figure 3. Timeline for this project.

5.5 Budget and funding

The overall budget for the project is shown in Table 8.

Table 8. Project budget and funding.

Item	Puget Sound Partnership Grant	Ecology In-Kind
Salary, benefits, and indirect/overhead	$2,934	$106,237
Equipment	$0	$0
Travel and other	$0	$0
Contracts	$64,840	$0
Laboratory	$0	$0

6.0 Quality Objectives

6.1 Data quality objectives

The main data quality objective for this project is to use the benthic invertebrate and habitat data to generate a Disturbance Index. Sediment contaminant data and potentially marine waters and other data sources will be used to screen parameters as indicative of human disturbance and test hypotheses of effects of management interventions. Upon validation of the Disturbance Index as the *Marine Benthic Index*, benchmark values delimiting multiple categories of benthic assemblage condition will be determined.

6.2 Measurement quality objectives

See also Section 6.4, Model quality objectives.

6.2.1 Targets for precision, bias, and sensitivity

6.2.1.1 Precision

The Disturbance Index will include 95% confidence intervals.

Marine Benthic Index values and condition categories for replicate samples should be no more variable than the variability of the invertebrate assemblages; in other words, the index should not introduce additional variability.

6.2.1.2 Bias

There are multiple sources of bias, some of which can be avoided, or at least minimized, by adherence to statistical procedures. Of particular relevance to the task of developing benthic indicators is bias induced by failure to take into account structuring habitat variables. By modeling species occurrence, abundance, or biomass as a function of environmental variables known or suspected to be fundamental (e.g., depth, grain size), it is possible to reduce masking or exaggerated responses.

In addition, although some developers of biotic indicators use only uncontaminated, pristine or near-pristine samples for calibrating indicators, doing so can introduce bias (Schoolmaster et al., 2013b).[15] Because the relationships among the habitat, biota, and stressor variables are complex, it would be better to use the entire range of samples to develop indicators.

A third source of bias is missing data. The degree of bias and ways to mitigate it are functions of how many measurements are missing, why they are missing, whether they are missing at random, and how important they are. Mitigation techniques include pairwise deletion of variables and samples (i.e., separately for each variable, delete only those samples which are missing), listwise deletion (delete from all samples missing any values), imputation (estimation of missing values by means, for example), use of model outputs, and other statistical techniques such as bootstrapping (large numbers of simulations of sampling with replacement from data

[15] That said, clear definition of reference conditions is important to determining thresholds for condition categories (Borja et al., 2019) (Section 7.3.2.6). See also discussion of reference conditions for model development in Section 7.3.2.1.

available). The method chosen must be one appropriate to the particular variable and situation, or other bias can be introduced.

6.2.1.3 Sensitivity

The *Marine Benthic Index* should be able to produce distinct outputs given the range of the spatial and temporal variation in the system.

Sensitivity will be used to inform model selection and evaluation at both the species level and the index level. For the species-level models we will use the sensitivity across the values of the environmental variables as a criterion for using the species in the index; for example, we will include species that show at least 10% response to the range of each environmental variable in our screening analyses.

At the index level we will use the SBI to identify sites that are scored as *unaffected* and *adversely affected*. We will use environmental variable values from each these sites (three highest and three lowest, for example) to anchor the ends of the ranges and set the criterion that the index should be able to show significant differences between these groups. In this analysis, significance will be determined by the uncertainty in the index values. For example, using the full distributions predicted by the index, we will estimate the probability that these samples came from the same underlying distribution. If the index fails to distinguish between these sites, we will revisit the underlying species models, parameterizations, and included environmental variables to improve the sensitivity.

In addition, it is a goal that the *Marine Benthic Index* be able to correspond significantly to judgments of undisturbed and minimally, moderately, and highly disturbed benthic assemblages in the validation dataset, using the definitions in Ranasinghe et al. (2013), as determined independently by marine benthic experts during the validation exercise. Consistency of rankings and categorical groupings will be measured by kappa analysis.

6.2.2 Targets for comparability, representativeness, and completeness

See also Section 6.4, Model quality objectives.

6.2.2.1 Comparability

Ecology's marine sediments and marine waters monitoring programs are long-term programs which have been using the same standard sampling and laboratory analytical methods for years (Dutch et al., 2018; Keyzers et al., 2020; Bos et al., 2022). Thus, the data to be used in this project already are comparable to each other over time.

Standardization of taxonomy, to ensure comparability of taxonomic identification across years, will be accomplished using the method described in Burgess (2019). Because Ecology's marine sediment monitoring program has involved the same few taxonomists for many years and kept up with internationally-recognized nomenclature changes over the years, the number of organisms and species affected is minimal.

Validation of the Disturbance Index as the *Marine Benthic Index* will be accomplished using the methods of Weisberg et al. (2008), which involve measuring consistency of best professional judgment by a panel of benthic experts, as described in Section 13.4.1.4, below.

6.2.2.2 Representativeness

The data to be used for this project are from samples collected throughout the study area annually (sediment) or monthly (water) for 30 years. The sediment monitoring surveys have been designed to estimate conditions for large geographical areas, from bay-scale to Sound-wide, as well as for individual long-term stations (Dutch et al., 2018). The marine waters surveys have been designed to estimate conditions for the entire southern Salish Sea (greater Puget Sound area), constituent oceanographic regions/water masses, and individual long-term stations (Keyzers et al., 2020; Bos et al., 2022).

6.2.2.3 Completeness

It should be possible to characterize benthic community condition using numerical and qualitative indicators for at least 95% of samples, preferably 99% or more. For cases in which indicators cannot be calculated (e.g., too few taxa or individuals), rules will be developed to specify default values. For example, azoic samples (such as have been found a few times in Budd Inlet) will be classified as extremely disturbed.

6.3 Acceptance criteria for quality of existing data

The data used for this study have already met the quality and usability acceptance criteria for Ecology’s Puget Sound Sediment Monitoring Program (Dutch et al., 2018) and Ecology’s Marine Waters Monitoring Program (Keyzers et al., 2020; Bos et al., 2022).

There are several data gaps, the primary of which are:

- Benthos variables (e.g., biomass) or sediment variables (e.g., nutrients) that were not measured until recent years.
- Lack of water-column measurements corresponding to many sediment stations.

To map the representativeness of the marine waters station network for coverage of the sediment monitoring station network will require specific study of the teams involved. It may be possible to predict past conditions from Salish Sea Model output to fill gaps.

6.4 Model quality objectives

Goals for a new method of assessing benthic condition will include:

- Be useful in cases where the mechanisms of human disturbance are unknown. In other words, this does not require prior knowledge/assumptions of disturbance mechanisms.
- Be able to deal with high-dimensional natural environmental gradients as a matter of design.
- Have a clear, quantitative definition of “reference condition” everywhere on the manifold of environmental conditions.
- Avoid the problem of over-interpretation of metrics that is common with indices created from community-level metrics. For example, if species diversity is a metric, then it is not the case that increasing the species diversity, regardless of how achieved, will reduce the ecological impact of human disturbance.
- Provide estimates of uncertainty for the indices derived from the “lack-of-fit” of individual species models and parameter estimates.

- Have a quantitative relationship to other common assessment methods.

Evaluation of the Disturbance Index would include:

- Whether, given the uncertainty, the estimates are precise enough to be useful, i.e., is there variation among sites within years as well as between years.
- Alignment with expert opinion. For example, if we are missing an important environmental variable, site ranks may be in error.
- Comparison with existing assessment devices for the same sites and years, either based on the same benthic communities or possibly other biotic communities thought to be strongly correlated with the benthic communities.

See also Section 7.3.2.5, Validate Disturbance Index as *Marine Benthic Index*, for the formal mechanism for comparing the model outputs to expert opinion.

7.0 Study Design

7.1 Study boundaries

Details of the study area are described in the Puget Sound Sediment Monitoring Program Quality Assurance Monitoring Plan (Dutch et al., 2018) and shown in Figure 1.

7.2 Field data collection

7.2.1 Sampling locations and frequency

The data to be used for this study are the already-collected long-term monitoring data from Ecology's Puget Sound Sediment Monitoring Program (Dutch et al., 2018) and Ecology's Marine Waters Monitoring Program (Keyzers et al., 2020; Bos et al., 2022). The selection of data to be used for model development is addressed in Section 7.3.2.2, below. No new field sampling will be required.

7.2.2 Field parameters and laboratory analytes to be measured

The data to be used for this study are the already-collected long-term monitoring data from Ecology's Puget Sound Sediment Monitoring Program (Dutch et al., 2018) and Ecology's Marine Waters Monitoring Program (Keyzers et al., 2020; Bos et al., 2022). No new field measurements or laboratory analyses will be required.

7.3 Modeling and analysis design

7.3.1 Analytical framework

7.3.1.1 Analytical framework: *Marine Benthic Index*

The Disturbance Index that will be the basis for the *Marine Benthic Index* will be developed using methods adapted from the fields of machine learning (artificial intelligence) and causal inference (statistical modeling). The approach will use broadscale environmental drivers and patterns of benthic species occurrence and abundance to inform an estimate of human-caused impact.

Traditional methods of developing indices are based on a selection of metrics based on pre-determined hypothesized stressors which are correlated with some measure of human disturbance (Figure 4). Often, the human disturbance is not known but assumed (Schoolmaster et al., 2012a). Because candidate metrics are often based on a mix of expert opinion and tradition, they may not reflect the most ecologically important aspects of the benthic ecosystem (Schoolmaster et al., 2012a). Furthermore, incorporating environmental variables with metrics calculated from observed species can either obscure or exaggerate human influences (Schoolmaster et al., 2013a).

The method to be used in this case uses a different approach. It begins instead with the occurrence and abundance of the benthic invertebrate species, and lets the benthos inform us of the condition of the habitat, thereby providing a tool by which potential stressors can be discovered.

The method builds causal sub-models for many of the commonly collected benthic organisms as functions of two sets of variables (Figure 5): environmental variables unlikely to be directly affected by human disturbance (*E*) and environmental variables likely to be affected by human disturbance (*D*). The variables in set *E* would be entered as inputs, while those in *D* would remain latent (unobserved). Fitting this set of models simultaneously for each taxon at each site will result in an estimate of *D* for each site, and estimated parameter α_j for every taxon. The estimates of *D* are directly transformed to create the Disturbance Index for locations. The parameters α_j contain information on the sensitivity of the taxa to disturbance and can be used by benthic experts to further understand the system. In other words, we will be using the occupancy patterns of each species to obtain a "crowdsourced" estimate of the impact of human disturbance at each sampling site. This approach is described in detail in Section 7.3.2.4.

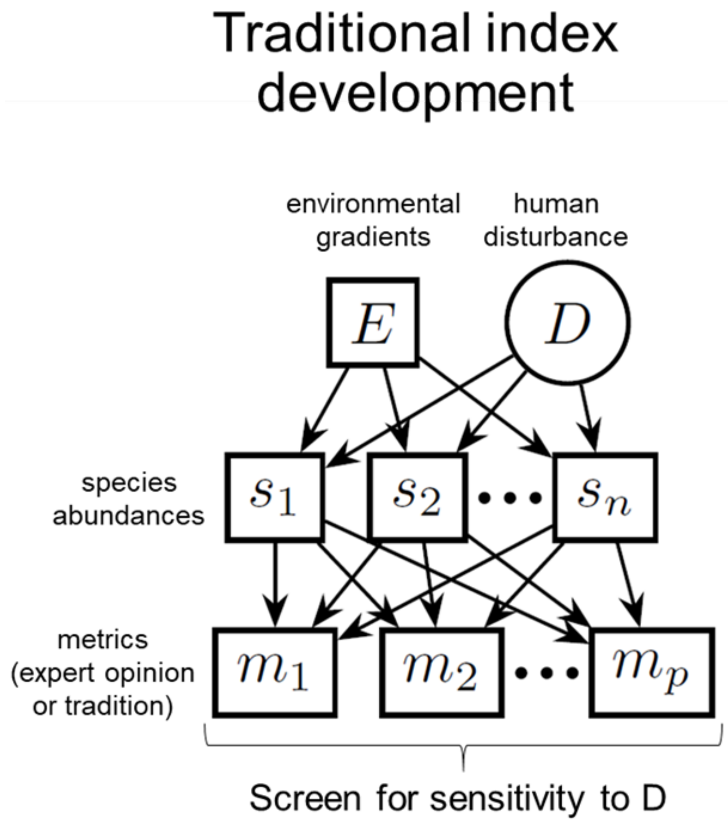


Figure 4. Traditional method of developing biotic indices.

The variables are:
E = matrix of known environmental gradients;
D = matrix of unknown disturbances;
s_1, ..., s_n = abundance of species 1, 2, ..., n;
m_1, ..., m_p = metric 1, 2, ..., p.

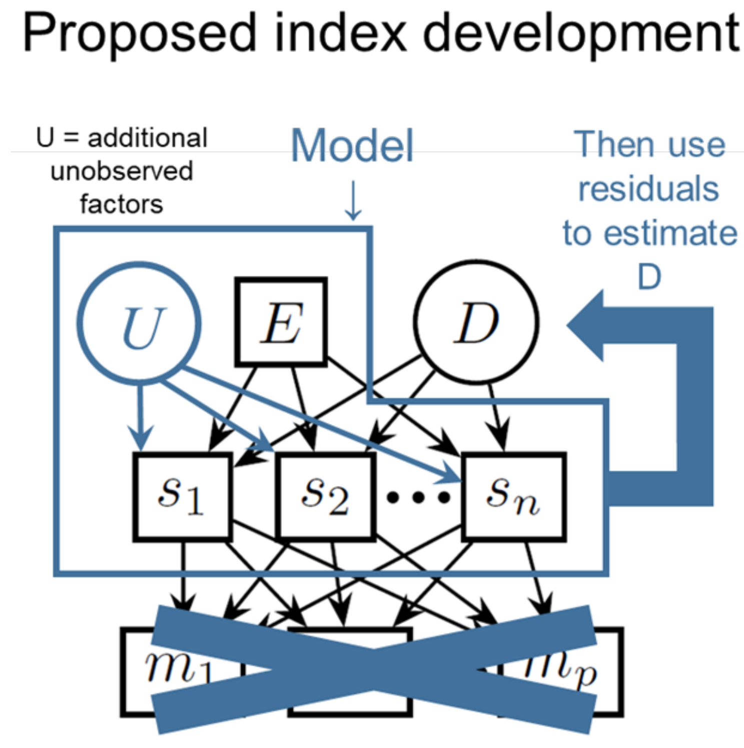


Figure 5. Proposed model approach for Disturbance Index.

The variables are:
E = matrix of known environmental gradients;
D = matrix of unknown disturbances;
U = matrix of additional unknown factors;
s_1, ..., s_n = abundance of species 1, 2, ..., n.
See Section 7.3.2.4 for model specification details.

Benefits of this approach include using information efficiently to estimate both species sensitivity and measures of *D*, not relying on correctly estimating which composites to calculate as potential metrics, and resulting in a direct estimate of *D* with uncertainty, which can be used to screen hypothesized mechanisms of human disturbance. We are not assuming anything about site disturbance *a priori*, but using the patterns of species composition to estimate it. A pilot study of the proposed approach using only species presence/absence resulted in relatively good correspondence with best professional judgment of Ecology's benthic invertebrate experts (Appendix D).

7.3.1.2 Analytical framework: Graphical Causal Model

A graphical causal model (GCM) is a probabilistic model. The nodes represent random variables and the arrows represent causal influence among them. The GCM encodes the conditional independence structure of the variables and characterizes the factorization of the joint probability distribution into the product of distributions represented by the nodes (Pearl, 2009). As such, the GCM is the graphical representation of an underlying set of causal functions. For example, the set of equations: $z = f(x),\ x = f(y)$, encodes the graphical model $y \rightarrow x \rightarrow z$.

The structure of the graphical causal model used for testing hypotheses of causation and estimating magnitudes of influence of multiple stressors in the Puget Sound benthic environment will be informed by the PSEMP-derived Driver-Pressure-State-Impact-Response (DPSIR) model of the Puget Sound benthic ecosystem (Dutch et al., 2018) and the Vital Signs Conceptual Models (McManus et al., 2020).

The graphical causal model represents a multivariate causal hypothesis. Fitting the model to data, which can be done using any of a number of numerical methods including, commonly, maximum likelihood or Bayesian estimation, gives feedback at two levels:

- How strong is the quantitative evidence in the data for causal drivers proposed in the model?
- The set of variables not directly connected represent a hypothesis of conditional independence between them. In practical terms, this means that after fitting the model there should be no residual relationship among the variables. If and how these hypotheses fail will give feedback on the hypothesized causal structure of the model as a whole and can guide revision for testing with future data.

GCM is further explained in Section 7.3.2.7, below.

7.3.2 Model setup and data needs

The proposed approach involves several steps:

- Selection of a baseline year or years: Because 'ideal' conditions for each species and community may not ever be manifested as a function of continuous, interacting drivers, basing indices on an absolute target is impossible. For this reason, all the measures must be based on a scale relative to some baseline. This requires selecting a year or set of years to use to parameterize the model, and as the yardstick for comparison for other years. Details are provided in Sections 7.3.2.1 and 7.3.2.2.
- Selection of a suite of species to use: The goal for this step is to include a range of species with different habitat requirements, specificities, and commonness/rarity. The species have to

be common enough that we can get stable estimates for the individual species models, but variable enough that there is variation among sites in presence/abundance.

- Selection of a set of environmental variables to use for species models: This step is not completely separable from the one above. It will be determined by a number of factors, such as spatial variability, importance as a driver of species presence/abundance, and likelihood of availability across years.

7.3.2.1 Define reference conditions

A reference condition provides the context from which a measured condition can be evaluated (Hawkins et al., 2010). As reviewed by Hawkins et al. (2010), there are two general approaches for defining reference condition: classifications based on natural environmental settings and models that use continuous values of environmental measurements as inputs. It is the latter approach that we will use for this study.

For the classification approach, reference sites are often selected to represent a least-disturbed condition. For the continuous input approach, however, this requires the assumption that the effect of D=d is known for each species for every point on the n-space defined by the environmental input variables. For a ecosystem as complex as Puget Sound, we feel that this is the not a feasible approach to employ. Instead, we use the distribution of disturbance observed in the training set to provide the context for evaluating the measured condition.

Specifically, we will use the training set (Section 7.3.2.2, below) to estimate the values d_{ref}, assuming they follow a standard normal distribution (i.e., $d_{ref} \sim N(0,1)$). Then, given an observation outside of the training set we can estimate d_{meas}, which can be placed in context as an index value in the range (0,1) as $d_{ind} = \int_{x=-\infty}^{d_{meas}} f_{D_{ref}}(x)dx$. Thus, d_{ind} can be interpreted as the percentile rank of d_{meas} relative to the distribution given by the training set. For example, for $d_{ind} = 0.9$ and training set from 2010, d_{meas} can be understood as being in a condition that would have placed it in top 10% of most disturbed sites in 2010.

This approach allows the reference condition to be updated over time, as necessary, and avoids the problematic assumption that the quantitative effect of disturbance is known for each species in all combinations of environmental conditions.

7.3.2.2 Choose independent calibration and validation datasets

A good dataset will need to have 1) variation in habitat that is linked to variables measured and 2) variation in the populations of the organisms to be modeled. In addition, it may be useful to identify 3) a few species that are good indicators of undisturbed conditions in different habitats or of disturbed conditions in different habitats, or both.

The choice of calibration and validation datasets will be based on a number of factors, including completeness of dataset, quality of taxonomy, and computational tractability.

In addition to the fact that changes in sampling methods and long-term trends in community membership and species adaptation could bias results, small directional shift in human disturbance can also affect the efficacy of the assessment (Figure 6).

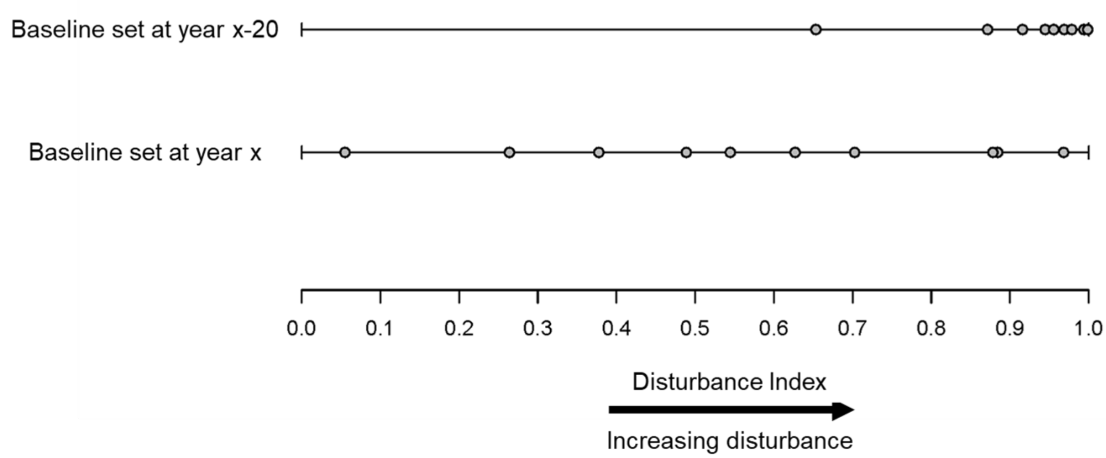


Figure 6. Illustration of rationale for selecting recent dataset to "train" disturbance model.

The disturbance model (i.e., the model used to develop the Disturbance Index) will be "trained" with benthos and habitat data from the 2016 and 2017 Puget Sound-wide sediment sampling events (Figure 7). In 2016, three replicate samples were taken at each of 22 stations consisting of the 10 long-term stations that the sediment monitoring program has sampled annually since 1989, plus 12 stations added at or near 12 of the long-term marine waters sampling stations. In 2017, three independent replicates were sampled at 28 stations from the random sample draw provided by EPA (Dutch et al., 2018), and single samples were taken at the 22 stations sampled in triplicate the previous year. For the purpose of training the model, all replicates will be treated as separate samples. This selection of samples to develop the model will ensure a wide variety of habitats, assemblages, and conditions in a time window short enough to minimize temporal changes.

The rationale for the choice of recent years as the baseline is illustrated in Figure 6. Imagine that we set the baseline for the disturbance model at year *x*-20, where *x* is the current year. Imagine also that there is some directional drift over that time in the effect of disturbance on the system. Since the Disturbance Index maps the observed values of *D* to the percentile rank of the baseline, values for year *x*+2 have very little relative difference and are crowded toward the upper extreme of the index scale (top line in Figure 6). Setting the baseline at year *x* (lower line in Figure 6) results in a spread in the points that shows the relative difference between them. In addition, the spans of researcher careers and the limits of human memory are such that the phrase "The *D* observed in site *y* would have been at the 10th percentile of those in 2016" is more intuitively meaningful than "The *D* observed in site *y* would have been at the 10th percentile of those in 1996."

Benthos and habitat data from the 50 stations sampled (no replicates) during the 2018 Puget Sound-wide sediment sampling event will be used to test the disturbance model developed with the 2016-2017 data.

A random selection of at least 40 samples from the 1997-2019 collection of benthos samples (>3000 samples) will be used for validation. The samples will represent a wide variety of habitats, assemblages, and conditions throughout Puget Sound.

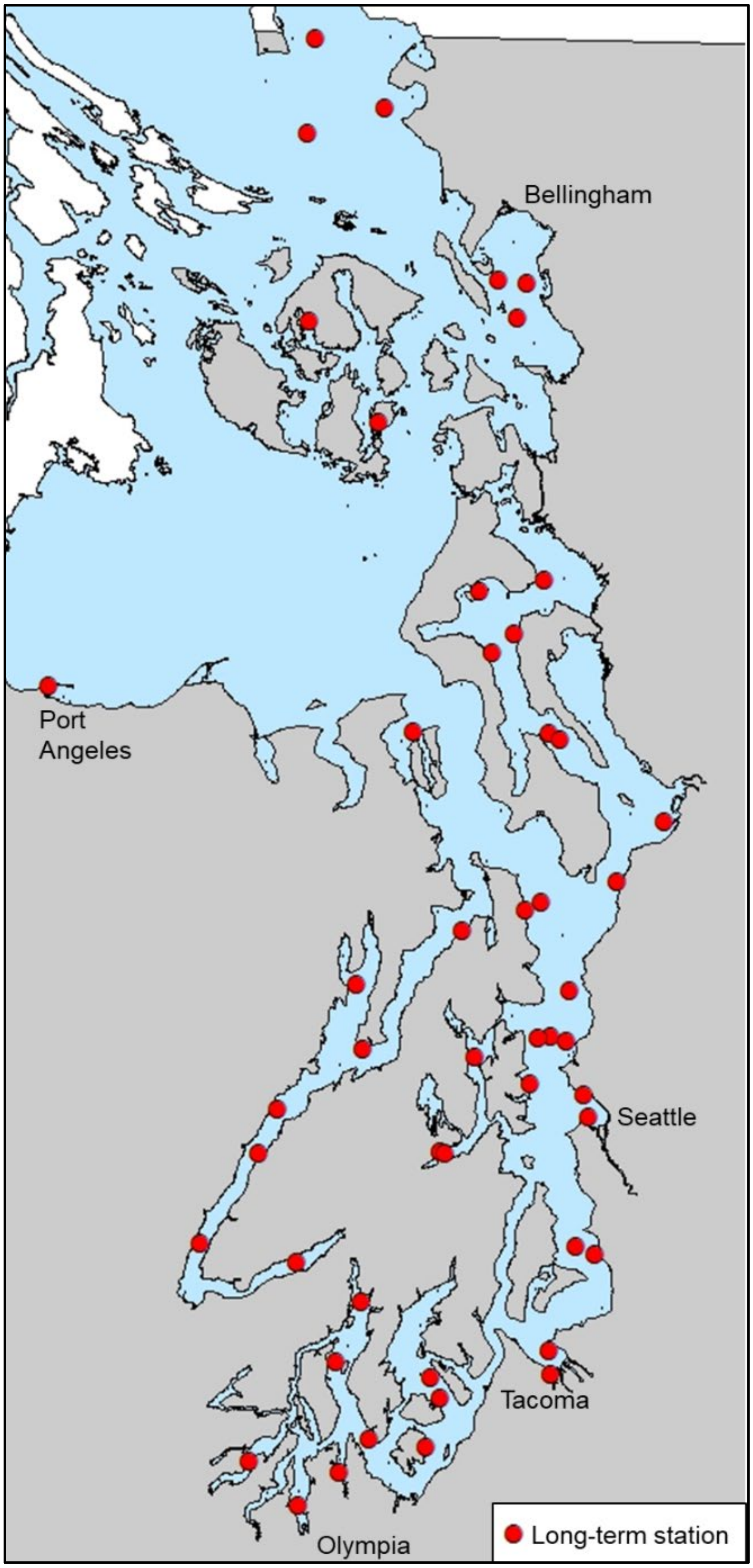


Figure 7. Puget Sound Sediment Monitoring Program Long-term stations.

7.3.2.3 Prepare data

It will be necessary to standardize (harmonize) the taxonomy across the calibration and validation datasets. The challenge is to preserve as much detailed information as possible. Ecology's Marine Sediment Monitoring Team developed and adopted a standardization protocol (Burgess, 2019) which balances retention, deletion, and summarization of species- and higher-level data.

7.3.2.4 Specify benthos-environment models, Disturbance Index

Using modern estimation methods, such as Markov Chain Monte Carlo (MCMC) or Structural Equation Models (SEM), it is possible to use the causal structure of the hypothesized data-generating mechanism to estimate the implied values for unobserved, often called latent, variables.

The logic of this approach is demonstrated through the following example:

Figure 5 hypothesizes that the abundance of each species $i = \{1, 2, \ldots, n\}$ in sample $j = \{1, 2, \ldots, k\}$ can be understood as being co-determined by three factors: (1) natural environmental gradients, (2) human disturbance, and (3) additional un-modeled factors. In symbols,

$$\boldsymbol{S}_{[k\times n]} = \boldsymbol{E}_{[k\times(p+1)]}\boldsymbol{B}_{[(p+1)\times n]} + \boldsymbol{D}_{[k\times 1]}\boldsymbol{\alpha}_{[1\times n]} + \boldsymbol{U}_{[k\times n]},$$

where the subscripted brackets give the dimensions of each matrix. In this equation, $\boldsymbol{S}$ is a matrix of species abundances, $\boldsymbol{B}$ is a matrix of coefficients describing the response of each of n species to the p measured environmental gradients in matrix $\boldsymbol{E}$. The vector $\boldsymbol{\alpha}$ contains the response of each species to the unknown disturbances $\boldsymbol{D}$, and $\boldsymbol{U}$ is a matrix of the variation from unspecified causes.

If a model, $\hat{s}_i$, of the observed variables, s_i, is fit, then the residuals will be

$$\hat{s}_i - s_i = \alpha_i \boldsymbol{D} + \varepsilon_i ,$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$ are the errors of the fitted model.

If we assume (1) that the remaining covariation between species is due to the joint dependence on $\boldsymbol{D}$ and (2) no covariation between $\boldsymbol{D}$ and ε_i[16], then the covariance of the residuals $[\alpha_1 D+ \varepsilon_1, \alpha_2 D+ \varepsilon_2, \ldots, \alpha_n D+\varepsilon_n]^{\mathrm{tr}}{}_{[n\times 1]}$, $\boldsymbol{\Sigma}_{[n\times n]}$, is equivalent to

$$\Sigma = \begin{bmatrix} \alpha_1^2\sigma_D^2 + \sigma_1^2 & \alpha_1\alpha_2\sigma_D^2 & \ldots & \alpha_1\alpha_n\sigma_D^2 \\ \vdots & \ddots & & \\ \alpha_1\alpha_n\sigma_D^2 & \ldots & \ldots & \alpha_n^2\sigma_D^2 + \sigma_n^2 \end{bmatrix},$$

where σ_D^2 is the variance in $\boldsymbol{D}$ and σ_i^2 is the residual variance in the abundance of species i.

Assuming a value for one of the α_i or σ_D^2 allows us to set the scale (i.e., units) on $\boldsymbol{D}$ and solve for the rest of the coefficients in $\boldsymbol{\Sigma}$.

[16] The assumption of no covariation between $\boldsymbol{D}$ and ε_i will be true by construction for any instantiation of the model, so cannot be checked empirically.

From this set of coefficients, the estimates $\widehat{\boldsymbol{D}}$ (with a measure of uncertainty) of $\boldsymbol{D}$ can be calculated for each site. $\widehat{\boldsymbol{D}}$ can be used to derive a corresponding set of index values as $DI_i = \int_{-\infty}^{D_i} f(D)dD$, which takes on values between 0 and 1 (see example in Figure 6). **Thus, we call *DI* the Disturbance Index.**

An example of the application of this approach is outlined in Appendix D, the results of the pilot study.

Solving for the coefficients also provides estimates $\widehat{\boldsymbol{\alpha}}$ (with a measure of uncertainty) of $\boldsymbol{\alpha}$, which can provide new information to benthic biologists about species' habitat preferences.

The benefits of this approach are (1) it uses information efficiently to estimate both species sensitivity $\boldsymbol{\alpha}$ and measures of $\boldsymbol{D}$, (2) it does not rely on correctly estimating which composites to calculate as potential metrics, and (3) it results in a direct estimate of $\boldsymbol{D}$ with uncertainty, which can be used to screen hypothesized mechanisms of human disturbance.

The Disturbance Index will undergo validation by a team of benthic invertebrate experts (Section 7.3.2.5, below); once validated, it will become the *Marine Benthic Index* Vital Signs Indicator.

The example model above is to provide insight into the approach. During model fitting we will, based on the data, decide if linear, generalized linear, or non-linear specifications are the best approach. In theory, the best set of variables, *E*, and choice of specification can be chosen independently for each species. The scale of the latent variable *D* will be determined by balance of interpretability and generality for the result.

7.3.2.5 Validate Disturbance Index as *Marine Benthic Index*

In order to assure that the Disturbance Index, to be reported as the *Marine Benthic Index*, is an accurate indicator of condition along a disturbance gradient, it is necessary to validate the index with an independent dataset. Furthermore, it will be useful to develop index value ranges that can be used as qualitative condition categories for communication.

The index will be validated using a peer-reviewed and well-accepted process involving agreement of best professional judgment of multiple benthic ecology experts (Teixeira et al., 2010; Dauvin et al., 2010). A panel of Ecology, Tribal, and external specialists will conduct a two-step validation using a separate subset of PSEMP data independent of the dataset used to calibrate the Disturbance Index. Those with expertise in subtidal benthic assemblage structure will independently assess the condition of invertebrate communities using best professional judgment. Thresholds for categories will be based on analysis of the index values and the expert judgments. Application of the newly developed index will be tested by the validation team for ease of use and accuracy.

The experts will be given species composition and abundance data, as well as data on habitat variables (excluding location), from anonymized samples in the validation dataset. Based on that information, they will be asked to determine the condition of the benthic community at each site. They will be asked to express the condition two ways: by ranking the relative condition of each site from best to worst, and by assigning to each site a condition category, defined as follows (Ranasinghe et al., 2013):

- Undisturbed: a community that would occur at a pristine site.
- Low Disturbance: a community that shows some indication of stress, but could be within measurement error of undisturbed condition.
- Moderate Disturbance: a community that shows clear evidence of physical, chemical, natural, or anthropogenic disturbance.
- High Disturbance: a community with a high magnitude of disturbance.

The benthic community experts will be free to use whichever community attributes they choose, though they will be asked also to list the attributes they used to determine site rankings and condition categories and to rate the importance of the attributes, from very important to very low importance. If they identify indicator species as one of the attributes used in their assessment, they will be asked to list the organisms used as indicator species and to rank the species importance using the same scale. They will not be given any information on sediment chemistry or toxicity, nor asked to differentiate among potential causes for affected condition.

Level of agreement among the experts will be determined by Spearman correlation of the sample rankings. For the categories, a weighted Kappa analysis will be used to compare category assignments. The purpose for the weights is to take into account greater levels of disagreement for categories farther apart than for adjacent categories (Weisberg et al., 2007).

In the event that agreement among the experts is low, a "Delphi" procedure, an iterative procedure of ranking and feedback (Hsu and Sandford, 2007), can be used to improve agreement. The previously-mentioned requested rationale for the rankings and condition categories will be shared so that the experts can learn from each other, generate additional insights, and improve agreement on subsequent, though still independent, assessments of the validation dataset, until consensus is reached (Ranasinghe et al., 2013).

7.3.2.6 Determine category thresholds, target value, graphical displays

Thresholds for categories will be based on analysis of the index values and the expert judgments from the validation exercise. The calculated indicator values will be ranked and compared to the rankings of the experts by Spearman correlation and graphically.

The calculated indicator values will be divided into four or five categories using weighted Kappa analysis. Percent agreement with the categorical assessments of the experts will be determined.

Application of the newly developed index will be tested by the validation team for ease of use and accuracy.

A target value for the *Marine Benthic Index* as a Vital Sign indicator to be used in assessing progress toward rejuvenation of Puget Sound – likely the threshold value for the undisturbed category – will be recommended.

Once the *Marine Benthic Index* has been validated, the MSMT, in consultation with the Technical Advisory Group (consisting of interested members of Puget Sound Partnership, PSEMP workgroups, tribes, others), will review the index, target values, and graphical displays for communication of both the quantitative and qualitative scales of the index.

7.3.2.7 Develop graphical causal model

A causal model is a multivariate set of hypotheses about the cause-effect relationships in a system. The existence and direction of each link can be determined by different criteria, such as logic or prior studies, and is determined *a priori* (Pearl, 2009). A causal model partitions the joint probability distribution of the included variables and can thus be used to provide criteria for identification for estimation of specific causal effects, and provide estimates for causal queries.

As before, let D be the latent (unknown) **disturbance** variables, E be the natural **environmental** variables, and S be the **species** occurrences. The probability of occurrence of the species given the environment and the disturbance is expressed as $Pr(S|E,D)$.

The procedure described in the previous section to develop the Disturbance Index is based on an underlying causal hypothesis which represents the joint cause of unspecified disturbance mechanisms. With the Disturbance index, we can screen data of potential proximal mechanisms of the human disturbance, where the set of proximal mechanisms will be determined by the combination of expert system knowledge and data availability. For example, in the pilot study (Appendix D), we found that the total nitrogen content of the sediment was moderately correlated with the Disturbance Index ($r=0.71$). This suggests that nitrogen enrichment might be one of the ways that humans are altering the system to negatively affect the benthic organisms.

Let P denote **mechanisms of human disturbance**, i.e., those factors in the environment that **have been altered by disturbance** and **have subsequently altered the benthic communities**. This causal hypothesis can be expressed graphically as $D \rightarrow P \rightarrow S \rightarrow M$, where M are metrics of interest that can be calculated from the species data, such as species diversity or number of invasive species.

The complexity and specificity of the $D \rightarrow P \rightarrow S \rightarrow M$ sub-model will be determined by the patterns of correlations among potential disturbance mechanisms P, and our ability to formulate a hypothetical relationship among them that satisfies both that pattern of correlations and the prior understanding of system experts.

Once we have screened for the set of potential mechanisms, we can test their individual or joint ability to explain the relationship between the measures of D and S using a method called the test of mediation (Baron and Kenny, 1986). For example, if mechanism p lies on the causal path between D and S, then the estimated coefficient associated with D in the model $Pr(S|E,D,P)$ would be smaller than in the model $Pr(S|E,D)$. That is, adding the potential mechanism to the model would decrease the explanatory power of D on S, because part (or all of it) is explained by the mechanism p.

When a final set of mechanisms are identified, the model can be used in a prospective manner by quantifying causal queries. For example, an interventional query that could be calculated, with a measure of uncertainty, is: "How would metric m change if disturbance mechanism p were reduced to p'?" As another example, one might query the effect of two alternative restoration strategies: "Are affected communities restored more by changing the level (e.g., concentration) of mechanism p_1 to p_1' or by changing the level of p_2 to p_2'?"

7.4 Assumptions underlying design

- The scale for the Disturbance Index will be set by conditions present at the time of collection of the data used for calibration. Any previous system-wide degradation of conditions prior to data collection could result in values biased downward compared to more pristine conditions that, while theoretically possible, are no longer observed. This issue is not unique to this method. All biological or ecological quality indices are defined relative to a chosen "reference condition" (Hawkins et al., 2010).
- There is a balance to choosing the number of species to include in the analysis. Rare species might be very informative of system condition, but due to sparsity in the reference data set, are hard to model accurately.
- The method for development of the Disturbance Index treats the presence/absence of species as occurring independently of one another, necessary for making the model tractable to parametrize. Models of similar structure, called Naïve Bayes Classifiers, have proved very successful approximations in a variety of contexts (Schoolmaster et al., 2012b; Tsangaratos and Ilia, 2016).
- Additionally, this model makes the strategic assumption that environmental drivers of benthic communities can be usefully categorized into two groups: 1) those that are unlikely to be altered by human disturbance on the temporal scales considered here, like depth and sediment texture, and 2) those that are modified by human activity and thus act as the mechanisms for human disturbance. We recognize that some drivers may be hard to categorize. In those cases, where possible, we will make the assumption that those variables belong in the latter category.

  We also note that the success and usefulness of the model will be robust to moderate violations of this assumption. For example, if there is a strong correlation between *D* and the variables in *E*, the result will be the inability of the Disturbance Index to make distinctions among them; the noise will be larger than the signal. As such, the ability of the Disturbance Index distinguish among sites will be a continuous function of the degree of unintended correlation between *D* and *E*. As a result, the usefulness of the index acts an estimate of the veracity of the assumption that *D* and *E* are independent.

7.5 Possible challenges and contingencies

- Estimates of ***D*** are powered by relative differences among sites, not an absolute value of ***D***. It will be necessary to set an interpretable scale. The exercise with the benthic experts to validate the Disturbance Index, ***D***, will result in setting an interpretable scale for the *Marine Benthic Index*.
- In selecting which species to include, the species need to be common enough to get good statistical fits but variable enough to be informative. Species which are too uncommon are too difficult to fit, while species which occur everywhere do not provide useful information.
- With big samples and low signal-to-noise ratio, it can be tricky to achieve statistical identification, in other words, to find a unique solution. It may be necessary to simplify the model specification or reduce the dataset used.
- Training a model based on values from one method then using it on values derived from a different method is a big source of potential bias or error.

7.5.1 Logistical problems

Not applicable — no sampling will be required.

7.5.2 Practical constraints

- Data missing or unusable: Statistical techniques may be used to impute missing data, or results of multiple analyses with and without a problematic variable will be compared to determine the analyses' sensitivity to that variable.
- Imperfect alignment of sediment and water data: Water parameters have been measured more frequently, but less geographically densely, than the sediment and benthos parameters. Although water moves, assumptions as to representativeness of sampled water parameters to small-scale geographic areas may be required. Marine oceanographic experts will be consulted as to representativeness of samples relative to given water masses and conditions.
- Taxonomic drift: The taxonomic identification of benthic invertebrate organisms will be standardized using the procedures of Burgess (2019).
- Lack of agreement among benthic experts during validation exercise: In the event that agreement among the experts is low, a "Delphi" procedure, an iterative procedure of ranking and feedback (Hsu and Sandford, 2007), can be used to improve agreement (Ranasinghe et al., 2013). Disagreement that cannot be resolved using the Delphi process may result in a reduced effective validation dataset.
- Disturbance Index not matching expert judgment well: Because the model will consider many species and their associated habitats simultaneously, it will likely find patterns that humans would be unable to. Specific differences between the Disturbance Index and expert opinion will be the starting point for explorations to understand them. Adjustments to the parameterization of the model may be necessary.
- Model complexity: Due to complexity and size, the model might not be mathematically tractable or might not converge. Adjustments to the parameterization of the model may be necessary.
- Model errors: Types of errors that can occur in developing structural equation models include errors of model specification and errors of interpretation (Grace, 2008). Given that the PSEMP DPSIR, Vital Signs, and Salish Sea models have been vetted by many Puget Sound subject-matter experts, the foundation of the graphical causal model should avoid most such errors.

7.5.3 Schedule limitations

Due to the complexity and unique nature of the statistical models, extra time may be required for the Technical Advisory Group to sufficiently understand the approach to approve the QAPP and the *Marine Benthic Index*.

To map the representativeness of the marine waters station network for coverage of the sediment monitoring station network will require specific study of the teams involved. Marine waters data will be incorporated as time allows.

8.0 Field Procedures

Not Applicable – No additional samples will be collected.

9.0 Laboratory Procedures

Not Applicable – No additional samples will be analyzed.

10.0 Quality Control Procedures

10.1 Table of field and laboratory quality control

Not Applicable – No additional samples will be taken or analyzed.

10.2 Corrective action processes

- Check for and correct input data problems: Thoroughly scrub and correct or standardize (for sums of percentages) data prior to use. Data-scrubbing will be accomplished by orthographical, numerical, graphical, and cartographical means. Data-scrubbing includes, but is not limited to, checks for:
  - Inconsistent spelling, capitalization, and spacing (important for case-sensitive programs).
  - Nonsense values (e.g., negative percentages).
  - Percentages not summing to 100.
  - Missing values.
  - Inconsistent units of measure.
  - Incorrect location coordinates.
- Check data for completeness: Check data for completeness before, during, and after computations. If needed data are dropped (vs. deliberately excluded) during computational procedures, stop. Figure out the problems (e.g., uncaught data error, programming bug) and fix them. Rerun the computations. Repeat.
- Check results for inconsistencies: Check computational results for inconsistent or nonsense values. If any are found, stop. Figure out the problems (e.g., uncaught data error, programming bug) and fix them. Rerun the computations. Repeat.

11.0 Data Management Procedures

11.1 Data recording and reporting requirements

Not Applicable – The data to be used for this study have already been transferred to EIM (sediment parameters, benthos, CTD data) and to EAP-MW (discrete water parameters).

11.2 Laboratory data package requirements

Not Applicable – No additional samples will be analyzed.

11.3 Electronic transfer requirements

Not Applicable – No additional samples will be analyzed.

11.4 EIM/STORET data upload procedures

Not Applicable – The data to be used for this study have already been transferred to EIM (sediment parameters, benthos, CTD data) and to EAP-MW (water parameters).

11.5 Model information management

Data used to develop, calibrate, and validate indicators will be stored in Access databases, Excel files, or csv or text files, depending on the stage of analysis. Data will be manipulated and calculated in Excel and R. At this writing, the benthos data comprise over 170,000 records (taxon x sample); the sediment chemistry/biogeochemistry/grain size data comprise over 475,000 records (parameter x sample).

Model code and outputs will be hosted on an online repository (GitHub or GitLab, or alternative.) These repositories allow for version tracking as well as collaborative coding and can be either publicly accessible or restricted to a defined user base. The final decisions on where the code and output are hosted will be determined to maximize access to end-users and meet any requirements of the State of Washington and USGS.

12.0 Audits and Reports

12.1 Field, laboratory, and other audits

Two stages of audits will be conducted: (1) the validation of the Disturbance Index by benthic ecology experts (Section 7.3.2.5) and (2) the review and vetting of the *Marine Benthic Index* and recommended target value by the Technical Advisory Group.

12.2 Responsible personnel

Ecology and Tribal benthic ecology experts will validate the Disturbance Index. The Technical Advisory Group, composed of interested stakeholders from multiple organizations, will review and vet the *Marine Benthic Index* and recommended target value for use as a Marine Water Vital Sign indicator for the Puget Sound Partnership.

12.3 Frequency and distribution of reports

Deliverables reports will be sent to the funder (PSP), as required by the terms of the contract between PSP and Ecology.

A final report will be written to document the development of the Disturbance Index, the datasets used to calibrate and validate the Disturbance Index, the validation and evaluation processes for the *Marine Benthic Index*, thresholds for categorizing *Marine Benthic Index* results as no/low/moderate/high disturbance, and results of application of the *Marine Benthic Index* as a Marine Water Vital Sign index.

A final report will be written to document the development of the graphical causal model and use as a “what-if” tool. The report will include a worked example.

12.4 Responsibility for reports

Valerie Partridge, Don Schoolmaster, Ecology’s Marine Sediment Monitoring Team, and tribal biologists and others who participate in the validation exercise will contribute to the project reports. As Project Manager, Valerie Partridge will be responsible for reports to the funder, per the terms of the contract between the Puget Sound Partnership and Ecology.

13.0 Data Verification

13.1 Field data verification, requirements, and responsibilities

Not Applicable – The data to be used for this study are the already-verified long-term monitoring data from Ecology’s Puget Sound Sediment Monitoring Program (Dutch et al., 2018) and Ecology’s Marine Waters Monitoring Program (Keyzers et al., 2020; Bos et al., 2022).

13.2 Laboratory data verification

Not Applicable – The data to be used for this study are the already-verified long-term monitoring data from Ecology’s Puget Sound Sediment Monitoring Program (Dutch et al., 2018) and Ecology’s Marine Waters Monitoring Program (Keyzers et al., 2020; Bos et al., 2022).

13.3 Validation requirements, if necessary

See Section 13.4.1, below.

13.4 Model quality assessment

Beginning in 1997, the benthic assemblages sampled as part of the Puget Sound Sediment Monitoring Program have been characterized as *unaffected* or *adversely affected* by best professional judgment of Puget Sound benthic experts (largely the same individuals over the years), based on nine univariate measures of abundance and diversity, plus the presence, absence, and abundance of taxa thought to be sensitive or tolerant.

As indicated in Section 6.2.1.3, above, the *Marine Benthic Index* developed should be able to correspond significantly to those binary assessments. Again, the matches will not be exact because the development of the *Marine Benthic Index* will take into consideration additional variables not addressed in the current SBI. Consistency in categorization will be measured by percent agreement.

13.4.1 Calibration and validation

Data from over 3,300 benthic invertebrate samples are available for use for indicator development and calibration, still leaving hundreds for validation.

Agreement of best professional judgment of multiple, independent benthic ecology experts is a recognized and well-accepted method for benthic indicator validation (e.g., Weisberg et al., 2008; Teixeira et al., 2010; Ranasinghe et al., 2013). The procedure is described briefly in Section 13.4.1.4, below, and in detail in Section 7.3.2.5, above.

13.4.1.1 Precision

Precision will be assessed using field replicates: three replicate benthos samples were taken at each of the 10 original long-term stations each year until 2016. With the transition to the current Long-term sediment monitoring program in 2016, three replicates were taken at each new station added to the station network, the first time each station was sampled (but not subsequently).

For numerical indicators, the relative standard deviation, or coefficient of variation, will be calculated as a measure of precision. For categorical indicators, percent agreement among replicates will be the measure of precision.

13.4.1.2 Bias

Bias can be estimated from the performance of the indicators at the original 10 long-term stations from 2016-2018, compared to the same stations from 1989-2015. The benthic communities at the 10 long-term stations have largely remained stable over time (Partridge et al., 2018). Therefore, index results from the training and testing datasets can be compared with index results for other years for those stations, to determine if there is a tendency of the index to under- or overstate the degree of disturbance.

13.4.1.3 Representativeness

The data to be used are from samples collected throughout the study area annually (sediment) or monthly (water) for 30 years. The sediment monitoring surveys have been designed to estimate conditions for large geographical areas, from bay-scale to Sound-wide.

13.4.1.4 Qualitative assessment

The indicator developed will be validated with a separate, independent dataset by best professional judgment of multiple, independent benthic ecology experts. The method to be used for *Marine Benthic Index* validation has been published in peer-reviewed literature and used worldwide. In this method, experts are given species composition and abundance data, as well as data on habitat variables (excluding location), from anonymized samples in the validation dataset, and asked to rank the samples and indicate categories of condition (e.g., no, low, moderate, or high disturbance). Level of agreement is determined by Spearman correlation of the sample rankings and by weighted Kappa analysis for the categories. In the event that agreement among the experts is low, a "Delphi" procedure, an iterative procedure of ranking and feedback, can be used to improve agreement. Details are given in Section 7.3.2.5, above.

13.4.2 Analysis of sensitivity and uncertainty

Sensitivity analyses will be conducted to assess the thresholds for dividing indicator results into classes of disturbance.

14.0 Data Quality (Usability) Assessment

14.1 Process for determining project objectives were met

As indicated in Sections 6.2.1.3 and 13.4, above, the *Marine Benthic Index* developed by this project should result in characterizations of community condition that are consistent with previous assessments in the extremes, should match well with independent assessments by multiple benthic experts, and should make sense. Although the causal submodels will inherently detect patterns that are beyond the capability of humans to detect, specific differences between the index results and human expertise should be explored.

14.2 Treatment of non-detects

Most of the data for this project do not contain non-detects; only the sediment chemistry and particulates nutrient data have non-detects.

The MSMP chemistry results fall into two groups: those with relatively few (generally <10%) non-detects – metals and PAHs – and those with mostly (generally >80%) non-detects – all other organics. No amount of statistical wizardry can overcome >50% non-detects. Therefore, usually only metals and PAHs are ever used in the analyses.

14.3 Data analysis and presentation methods

The *Marine Benthic Index* developed will have two components, a calculated value on a continuous numerical scale, and a category of condition. The condition category would be one of no, low, moderate, or high disturbance (or some such scale). The numerical thresholds of the category boundaries would be a result of the validation step.

For presentation purposes, the numerical component could be re-scaled to 0 to 100, if required. Stoplight colors are often used to display condition categories: green (no or minimal disturbance), yellow (low level of disturbance), orange (moderate disturbance), red (highly disturbed). The Marine Sediment Monitoring Team will work with the Technical Advisory Group to recommend presentations of the *Marine Benthic Index* for publication in Puget Sound Partnership Vital Signs reports.

14.4 Sampling design evaluation

Not Applicable – No additional samples will be collected.

14.5 Documentation of assessment

Data usability will be included in the report written to summarize this project.

15.0 References

Ahmed, A., C. Figueroa-Kaminsky, J. Gala, T. Mohamedali, G. Pelletier, and S. McCarthy. 2019. Puget Sound Nutrient Source Reduction Project Volume 1: Model Updates and Bounding Scenarios. Washington State Department of Ecology Publication 19-03-001. https://apps.ecology.wa.gov/publications/SummaryPages/1903001.html

Ahmed, A., J. Gala, T. Mohamedali, C. Figueroa-Kaminsky, S. McCarthy. 2021. Technical Memorandum: Puget Sound Nutrient Source Reduction Project Phase II - Optimization Scenarios (Year 1). Washington State Department of Ecology, Olympia. https://www.ezview.wa.gov/Portals/_1962/Documents/PSNSRP/OptimizationScenarioTechMemo_9_13_2021.pdf

Anderson, M.J., R.N. Gorley, and K.R. Clarke. 2008. PERMANOVA+ for PRIMER: Guide to software and statistical methods. PRIMER-E, Plymouth, UK.

Baron, R.M., and D.A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 5:1173-1182.

Bilhimer, D. Project Manager, Puget Sound Nutrient Source Reduction Project, Water Quality Program, Washington State Department of Ecology. 2019. Personal communication with V. Partridge.

Borja, Á., G. Chust, and I. Muxika. 2019. Forever young: The successful story of a marine biotic index. *Advances in Marine Biology* 82:93-127.

Bos, J., N. Coleman, M. Horwith, E. Rauschl, S. Albertson, J. Ruffner, and C. Krembs. 2022. Addendum 1 to Quality Assurance Monitoring Plan: Long-Term Marine Waters Monitoring, Water Column Program. Washington State Department of Ecology https://apps.ecology.wa.gov/publications/SummaryPages/2203103.html

Brown, C. Supervisor, Water Quality Management Unit, Water Quality Program, Washington State Department of Ecology. 2019. Personal communication with V. Partridge.

Burgess, D. 2019. Standard Operating Procedure for Taxonomic Standardization of Benthic Invertebrate Data. Environmental Assessment Program Standard Operating Procedure EAP128. Washington State Department of Ecology, Olympia, WA.

Dauvin, J.-C., G. Bellan, and D. Bellan-Santini. 2010. Benthic indicators: from subjectivity to objectivity – where is the line? *Marine Pollution Bulletin* 60(7):947-953.

Dutch, M., V. Partridge, S. Weakland, D. Burgess, and A. Eagleston. 2018. Quality Assurance Monitoring Plan: The Puget Sound Sediment Monitoring Program. Washington State Department of Ecology Publication 18-03-109. https://apps.ecology.wa.gov/publications/SummaryPages/1803109.html

Ecology (Washington State Department of Ecology). 2013. Sediment Management Standards. Chapter 173-204 Washington Administrative Code. Washington State Department of Ecology, Olympia, WA. Publication 13-09-055. https://apps.ecology.wa.gov/publications/SummaryPages/1309055.html http://apps.leg.wa.gov/WAC/default.aspx?cite=173-204

Ecology (Washington State Department of Ecology). 2020. Water Quality Program Policy 1-11 Chapter 1: Washington's Water Quality Assessment Listing Methodology to Meet Clean Water Act Requirements. Washington State Department of Ecology Publication no. 18-10-035 https://apps.ecology.wa.gov/publications/SummaryPages/1810035.html

Grace, J.B. 2008. Structural Equation Modeling for Observational Studies. *Journal of Wildlife Management* 72(1):14-22.

Hawkins, C.P., J.R. Olson, and R.A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society* 29:312-343.

Hsu, C.-C., and B.A. Sandford. 2007. The Delphi Technique: Making sense of consensus. *Practical Assessment, Research & Evaluation* 12:1-8.

Keyzers, M., J. Bos, and S. Albertson. 2020. Quality Assurance Monitoring Plan: Long-Term Marine Waters Monitoring, Water Column Program. Washington State Department of Ecology Publication 21-03-108. https://apps.ecology.wa.gov/publications/SummaryPages/2103108.html

Khangaonkar, T., A. Nugraha, W. Xu, W. Long, L. Bianucci, A. Ahmed, T. Mohamedali, and G. Pelletier. 2018. Analysis of hypoxia and sensitivity to nutrient pollution in Salish Sea. Journal of Geophysical Research: Oceans 123:4735–4761.

Lee II, H., K. Marko, M. Hanshumaker, C. Folger, and R. Graham. 2015. User's Guide & Metadata to Coastal Biodiversity Risk Analysis Tool (CBRAT): Framework for the Systemization of Life History and Biogeographic Information. Environmental Protection Agency, Office of Research and Development, Western Ecology Division. EPA/601/B-15/001.

Lee II, H., C.L. Folger, D.A. Reusser, P. Clinton, and R. Graham. 2017. Predicting Patterns of Vulnerability to Climate Change in Near Coastal Species Using an Algorithm-Based Risk Framework. Environmental Protection Agency, Office of Research and Development, Western Ecology Division. EPA/ORD/R-17/052.

Levin, P.S., A. James, J. Kershner, S. O'Neill, T. Francis, J. Samhouri, C. Harvey, M.T. Brett, and D. Schindler. 2011. Understanding future and desired system states. Chapter 1A in: Puget Sound Science Update. Puget Sound Partnership. pp. 6-144.

Macdonald, T.A., B. Burd, V.I. Macdonald VI, and A. van Roodselaar. 2010. Taxonomic and feeding guild classification for the marine benthic macroinvertebrates of the Strait of Georgia, British Columbia. *Canadian Technical Reports of Fisheries and Aquatic Sciences* 2874:1-63.

Macdonald, T.A., B. Burd, and A. van Roodselaar. 2012. Facultative feeding and consistency of trophic structure. *Marine Ecology Progress Series* 445:129–140.

McCarthy, S., C. Figueroa-Kaminsky, A. Ahmed, T. Mohamedali, and G. Pelletier. 2018. Quality Assurance Project Plan: Salish Sea Model Applications. Washington State Department of Ecology Publication 18-03-111. https://apps.ecology.wa.gov/publications/SummaryPages/1803111.html

McManus, E., K. Durance, and S. Khan. 2020. Revisions to Puget Sound Vital Signs and Indicators. A Collaboration of Ross Strategic and Puget Sound Partnership. Olympia, WA.

O'Neill, S., S. Redman, C. Sullivan, K. Stiles, H. Harguth, and T. Collier. 2018. Evolving the portfolio of indicators to assess and report on the condition and recovery of the Puget Sound ecosystem: Moving from theory to practice. Puget Sound Partnership, Tacoma, WA.

Partridge, V., S. Weakland, M. Dutch, D. Burgess, and A. Eagleston. 2018. Sediment Quality in Puget Sound: Changes in chemical contaminants and invertebrate communities at ten sentinel stations, 1989–2015. Washington State Department of Ecology Publication 18-03-005. https://apps.ecology.wa.gov/publications/SummaryPages/1803005.html

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge University Press, New York.

Pelletier, G., L. Bianucci, W. Long, T. Khangaonkar, T. Mohamedali, A. Ahmed, and C. Figueroa-Kaminsky. 2017a. Salish Sea Model: Ocean Acidification Module and the Response to Regional Anthropogenic Nutrient Sources. Washington State Department of Ecology Publication 17-03-009. https://apps.ecology.wa.gov/publications/SummaryPages/1703009.html

Pelletier, G., L. Bianucci, W. Long, T. Khangaonkar, T. Mohamedali, A. Ahmed, and C. Figueroa-Kaminsky. 2017b. Salish Sea Model: Sediment Diagenesis Module. Washington State Department of Ecology Publication 17-03-010. https://apps.ecology.wa.gov/publications/SummaryPages/1703010.html

PSP (Puget Sound Partnership). 2020. Puget Sound Vital Signs. https://www.psp.wa.gov/evaluating-vital-signs.php

Ranasinghe, J.A., E.D. Stein, M.R. Frazier, and D.J. Gillett. 2013. Development of Puget Sound Benthic Indicators. Report to the Washington State Department of Ecology. SCCWRP Technical Report 755. Washington State Department of Ecology Publication 13-03-035. https://apps.ecology.wa.gov/publications/SummaryPages/1303035.html

Schoolmaster, D.R. Jr., J.B. Grace, and E.W. Schweiger. 2012a. A general theory of multi-metric indices and their properties. *Methods in Ecology and Evolution* 3:773–781.

Schoolmaster, D.R. Jr., J.B. Grace, and E.W. Schweiger. 2012b. Wetland bioassessment using Naïve Bayesian indicator species analysis. 97th Ecological Society of America Annual Convention, Portland, OR, August 2012.

Schoolmaster, D.R. Jr., J.B. Grace, E.W. Schweiger, G.R. Guntenspergen, B.R. Mitchell, K.M. Miller, and A.M. Little. 2013a. An algorithmic and information-theoretic approach to multi-metric index construction. *Ecological Indicators* 26:14-23.

Schoolmaster, D.R. Jr., J.B. Grace, E.W. Schweiger, B.R. Mitchell, and G.R. Guntenspergen. 2013b. A causal examination of the effects of confounding factors on multi-metric indices. *Ecological Indicators* 29:411–419.

Teixeira, H., Á. Borja, S.B. Weisberg, J.A. Ranasinghe, D.B. Cadien, D.M. Dauer, J.-C. Dauvin, S. Degraer, R.J. Diaz, A. Grémare, I. Karakassis, R.J. Llansó, L.L. Lovell, J.C. Marques, D.E. Montagne, A. Occhipinti-Ambrogim, R. Rosenberg, R. Sardá, L.C. Schaffner, and R.G. Velarde. 2010. Assessing coastal benthic macrofauna community condition using best professional judgement – Developing consensus across North America and Europe. *Marine Pollution Bulletin* 60:589-600.

Tsangaratos, P. and I. Ilia. 2016. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena* 145:164–179.

Weakland, S., V. Partridge, M. Dutch, and C. Maloy. 2017. Changes in Puget Sound sediment quality measured over a quarter century. In: C.A. James, J. Lanksbury, D. Lester, S. O'Neill, T. Roberts, C. Sullivan, and J. West, Eds. PSEMP Toxics Work Group. 2016 Salish Sea Toxics Monitoring Review: A Selection of Research. Puget Sound Ecosystem Monitoring Program. Tacoma, WA. pp. 32-34.

Weakland, S., V. Partridge, and M. Dutch. 2018. Sediment Quality in Puget Sound: Changes in chemistry, toxicity, and benthic invertebrates at multiple geographic scales, 1989–2015. Washington State Department of Ecology Publication 18-03-004. https://apps.ecology.wa.gov/publications/SummaryPages/1803004.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. 2007. The level of agreement among experts applying Best Professional Judgment to assess the condition of benthic infaunal communities. SCCWRP Technical Report 523. Southern California Coastal Water Research Project, Costa Mesa, CA.

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.

WSAS (Washington State Academy of Sciences). 2012. Washington State Academy of Sciences Committee on Puget Sound Indicators, Sound Indicators: A Review for the Puget Sound Partnership. Olympia, WA. 101pp.

16.0 Appendices

Appendix A. Previous efforts to develop benthic indices

Listed below are the previous efforts to develop benthic indices and biological-effects criteria for Washington State and Puget Sound, including findings, recommendations, and publications (compiled by Margaret Dutch; links current as of February 2022):

- **1991: Benthos assessment tool in WA State Sediment Management Standards**

  WAC 173-204-320 (3) – Marine sediment quality standards, Biological effects criteria: (c) Benthic abundance: *The test sediment has less than fifty percent of the reference sediment mean abundance of any one of the following major taxa: Class Crustacea, Phylum Mollusca or Class Polychaeta, and the test sediment abundance is statistically different (*t *test,* p ≤ *0.05) from the reference sediment abundance.*

  Reference: Ecology (Washington State Department of Ecology). 2013. Sediment Management Standards. Chapter 173-204 Washington Administrative Code. Washington State Department of Ecology, Olympia, WA. Publication 13-09-055. http://apps.leg.wa.gov/WAC/default.aspx?cite=173-204

- **1993: National Benthic Experts Workshop (EPA, PSWQA, Ecology – Seattle)**

  Purpose: To identify and evaluate technical adequacy of current methods used by Puget Sound programs to assess adverse effects on benthic communities and develop recommendations for improving methods. The participants focused primarily on examining numeric indices and univariate tests; there was limited examination of multivariate techniques and indices.

  Recommendations included: 1) identify to lowest possible taxonomic level; 2) evaluate > 1 univariate benthic endpoint; 3) need to identify indicator species; 4) identify reference conditions; 5) use univariate statistical tests (*t*-tests and ANOVA) to compare study area (contaminant focus) and reference conditions.

  Publication: PTI Environmental Services. 1993. Recommendations for assessing adverse benthic effects in Puget Sound. Prepared for Washington Department of Ecology, Sediment Management Unit, Olympia, Washington. https://apps.ecology.wa.gov/publications/SummaryPages/9309001.html

- **1995-2000: Development of Reference Value Ranges for Assessment of Benthos (Ecology's Toxics Cleanup Program – Sediment Management Unit)**

  Findings: Reference Value Ranges identified for 14 univariate benthic endpoints (abundance, richness, Swartz dominance index, etc.) in four habitat categories (based on % fines) for stations <150 feet deep. These defined the characteristics of "reference" communities. Reference communities compared to test communities with univariate tests (*t*-tests) with recommendations that impact decisions based primarily on the SDI and enhanced polychaete abundances and comparison to exceedances of Sediment Quality Standards (SQS) and Cleanup Screening Levels (CSL) levels. It was noted that these reference ranges and preferred end points may change with increasing database of benthos data from different areas of Puget Sound.

Outcome: Final peer review of reference ranges, preferred benthic endpoints, and threshold values led to concerns: 1) wanted more data to be included in Sediment Quality (SEDQUAL) database, 2) sensitivity and efficiency of recommended benthic endpoint reference ranges was in question, 3) significant time would be needed to modify the Rule and present to public and state.

Bottom Line: No changes to SQS were ever made based on these studies and recommendations.

Publications:

Weston Inc. 1995. Development and Evaluation of Biological Effects Criteria and Reference Area Performance Standards of the Sediment Management Standards Rule - Task 3 Report: Evaluation and recommendation of revised SMS benthic infaunal sediment standards. Prepared for Washington Department of Ecology, Olympia, WA. https://apps.ecology.wa.gov/publications/SummaryPages/9509001.html

Striplin Environmental Associates Inc. 1996. Development of Reference Value Ranges for Benthic Infauna Assessment Endpoints in Puget Sound. Prepared for Washington Department of Ecology, Olympia, WA. https://apps.ecology.wa.gov/publications/documents/9609001.pdf

Striplin Environmental Associations and Roy F. Weston, Inc. 1999. Puget Sound Reference Value Project. Task 3: Development of Benthic Effects Sediment Quality Standards. US Environmental Protection Agency, Region 10, Lacey, Washington. https://apps.ecology.wa.gov/publications/documents/9909001.pdf

Striplin Environmental Associates and MER Consulting. 2000. Peer Review of Ecology's Proposed Benthic Assessment Methods and Endpoints for Use in Regulatory Decisions: Responsiveness Summary. Prepared for Washington Department of Ecology Toxics Cleanup Program, Olympia, WA. https://www.researchgate.net/publication/283090312_Peer_Review_of_Ecology's_Proposed_Benthic_Assessmeent_Methods_and_Endpoints_for_use_in_Regulatory_Decisions

- **2003: Ecology's Environmental Assessment Program Marine Sediment Monitoring Team –** Development of *Sediment Quality Triad* of indicators, including a *Sediment Benthic Index*, with binary categories (*unaffected*, *adversely affected* communities), based on best professional judgment after examination of univariate metrics (abundance, taxa richness, evenness, dominance) and presence of sensitive and tolerant species.

  Publication: Dutch, M., V. Partridge, S. Weakland, D. Burgess, and A. Eagleston. 2018. Quality Assurance Monitoring Plan: The Puget Sound Sediment Monitoring Program. Washington State Department of Ecology Publication 18-03-109. https://apps.ecology.wa.gov/publications/SummaryPages/1803109.html

- **2008-2013: Ecology's Environmental Assessment Program Marine Sediment Monitoring Team –** Development of multivariate Puget Sound benthic index with grant from Puget Soundkeeper Alliance and contract with Southern California Coastal Water Research Project (SCCWRP). Steps included:

- Assess number and distribution of benthic macrofaunal assemblages in Puget Sound.
- Calibrate five benthic indices developed elsewhere for use in Puget Sound:
  - Benthic Response Index (BRI).
  - AZTI Marine Biotic Index (AMBI).
  - Relative Benthic Index (RBI).
  - Benthic Quality Index (BQI).
  - Observed over Expected (O/E) Index (based on RIVPACS approach).
- Create validation data set to evaluate performance of calibrated Puget Sound benthic indices, based on best professional judgment of expert benthic ecologists.
- Evaluate the performance of the calibrated Puget Sound benthic indices.

Findings

- BQI identified as best-performing index.
- AMBI/BQI/RBI and BQI/BRI/RBI index combinations performed 2^{nd}- and 3^{rd}-best, respectively.

Recommendations

- Prepare guidance and documentation to support routine benthic monitoring in Puget Sound.
- Expand confidence in the BQI with further validation testing.
- Further explore the two three-index combination options that performed next-best.

Publication: Ranasinghe, J.A., E.D. Stein, M.R. Frazier, and D.J. Gillett. 2013. Development of Puget Sound Benthic Indicators. Report to the Washington State Department of Ecology. SCCWRP Technical Report 755. Washington State Department of Ecology Publication 13-03-035.
https://apps.ecology.wa.gov/publications/SummaryPages/1303035.html

Appendix B. Sediment and marine waters stations

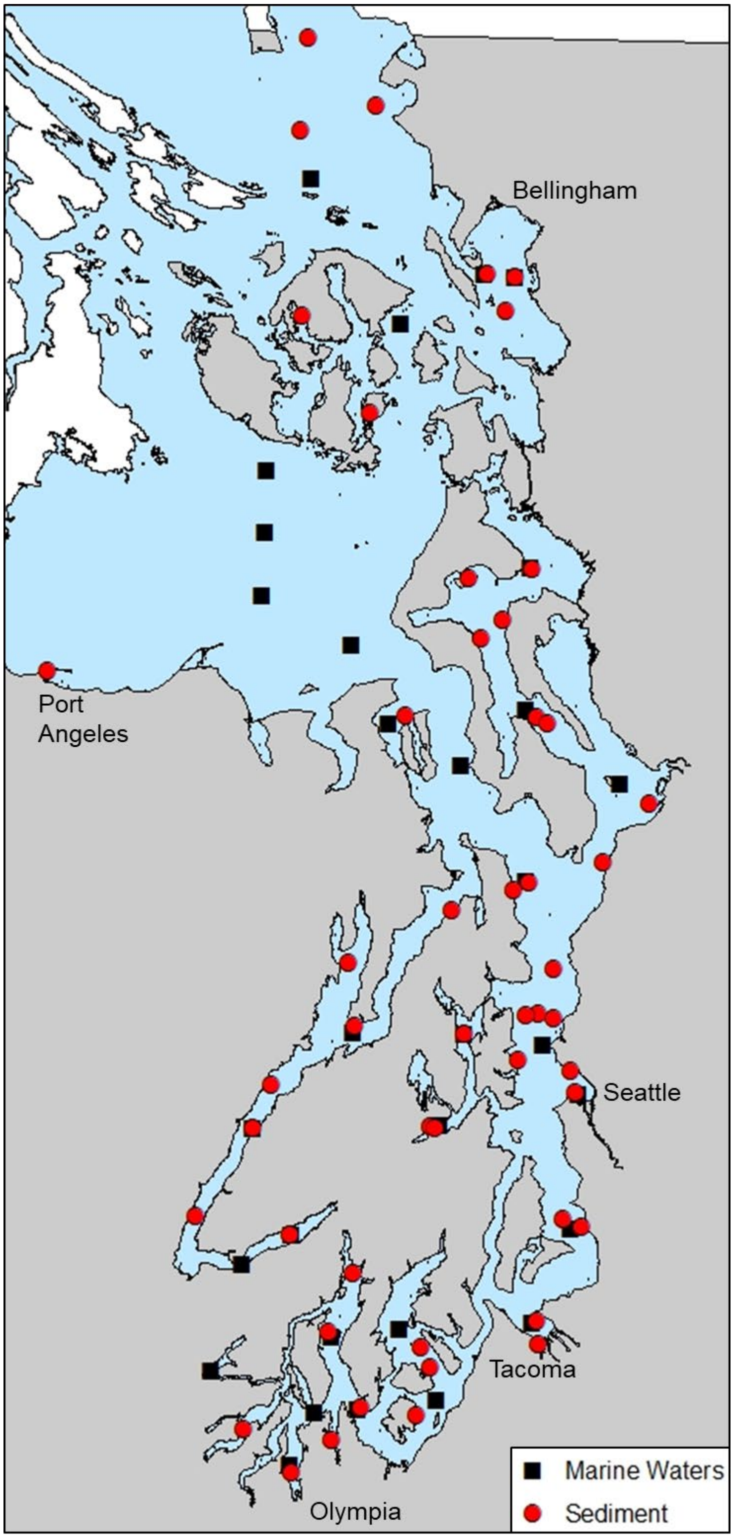


Figure B1. Long-term sediment monitoring stations and core marine waters monitoring stations in the southern Salish Sea.

Appendix C. Sources of boundary condition data

Publicly-available data on oceanographic indices, large river flows, and weather, already used by Ecology's Marine Waters Monitoring Program, include (links current as of February 2022):

Pacific Decadal Oscillation (PDO) Index, from the University of Washington
description: http://research.jisao.washington.edu/pdo/
data: http://research.jisao.washington.edu/pdo/PDO.latest.txt

North Pacific Gyre Oscillation (NPGO) Index, from the University of Washington
description: http://www.o3d.org/npgo/
data: http://www.o3d.org/npgo/data/NPGO.txt

El Niño/Southern Oscillation (ENSO) Multivariate ENSO Index (MEI), from NOAA
description: https://origin.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml
data: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

Bakun Upwelling Index Values, from NOAA
description: https://oceanview.pfeg.noaa.gov/products/upwelling/bakun
data: http://www.pfeg.noaa.gov/products/PFELData/upwell/monthly/upanoms.mon

Fraser River flow, from Environment Canada
description: https://wateroffice.ec.gc.ca/index_e.html
data:
https://wateroffice.ec.gc.ca/report/data_availability_e.html?type=historical&station=08MF005&parameter_type=Flow+and+Level

Flow data for Washington Rivers, from USGS
description and data: https://waterdata.usgs.gov/wa/nwis/qw

Flow data for Washington Rivers, from the Washington State Department of Ecology
description and data: https://ecology.wa.gov/Research-Data/Monitoring-assessment/River-stream-monitoring/Flow-monitoring

Weather data, from NOAA
description and data:
https://www.weather.gov/wrh/Climate?wfo=sew

Appendix D. Pilot study of proposed index approach

Simple model: presence/absence, no interactions

- 2017 data (50 stations spanning Puget Sound, no replicates): wide variability, but practicable size
- Best-fit habitat variables: depth, fines, penetration, salinity, temperature
- Simplifying assumptions:
  - No statistical interactions between disturbance and environment
  - Species independent
- Predict health of site based on species: “Let species vote with their feet”

Disturbance Index results

- Least (0) to Most (1) disturbed
- 95% confidence intervals

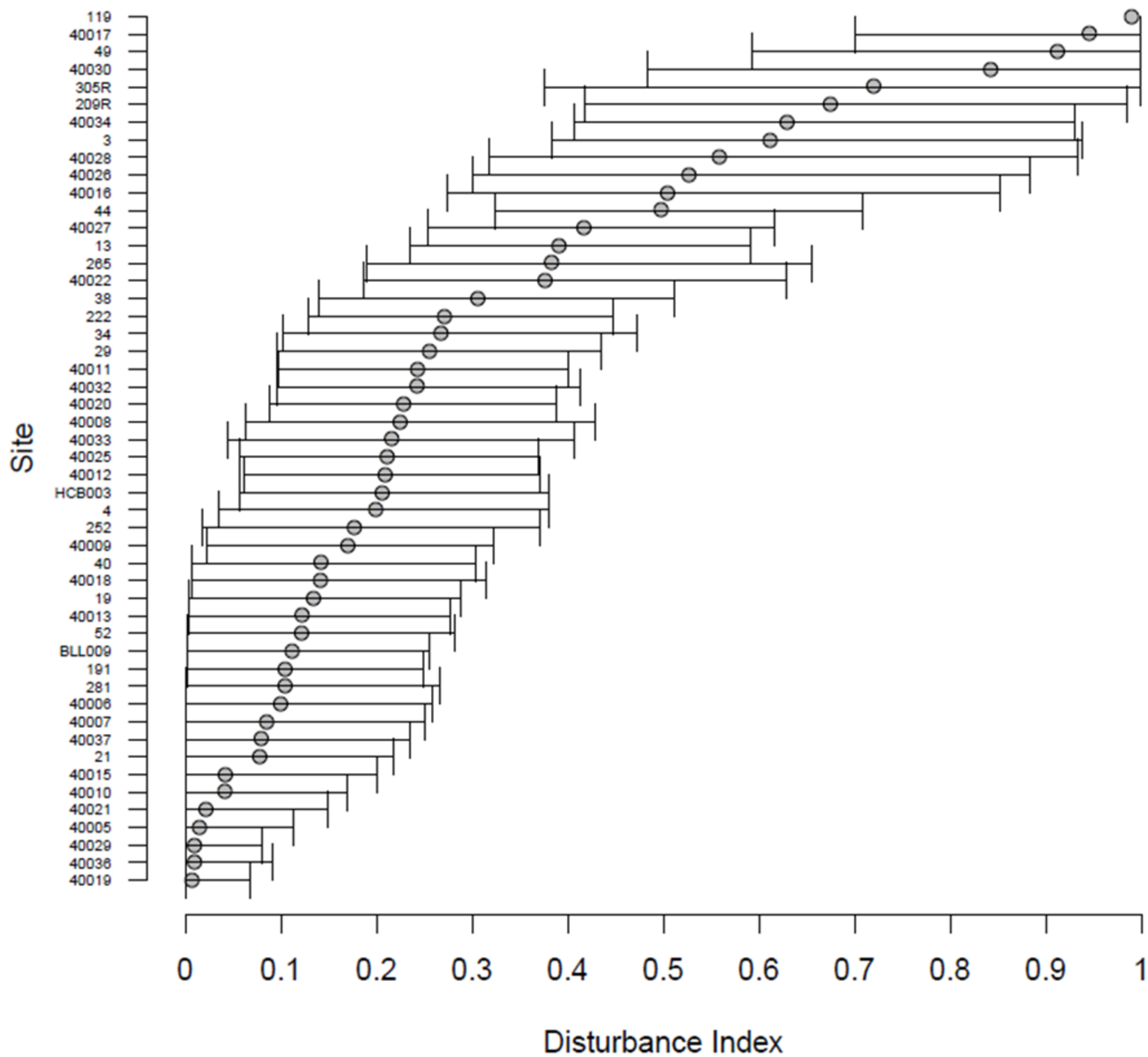


Figure D1. Disturbance Index results from pilot study of proposed index approach.

Disturbance Index results (continued)

- Compared to binary Sediment Benthic Index:
  - Upper half of sites mostly *adversely affected*; lower half mostly *unaffected*
  - Batting 80% in extremes (*unaffected* / *adversely affected*)
  - **With just simple presence/absence model!**

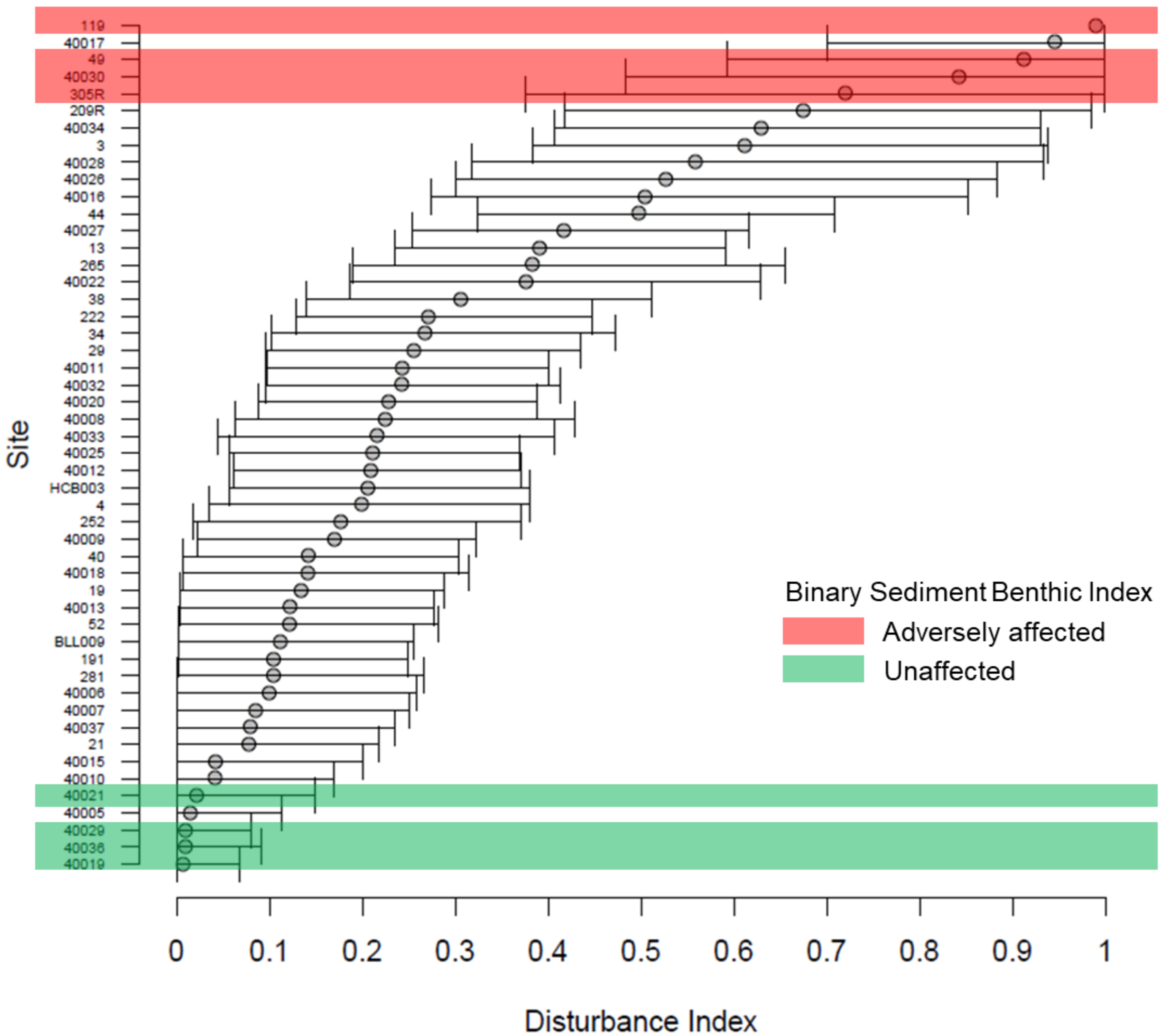


Figure D2. Comparison of Disturbance Index results to binary Sediment Benthic Index shows agreement for 4 of 5 most-disturbed sites and for 4 of 5 least-disturbed sites.

Testing a hypothesis

- Example: Sedimentary Total N
- Largely matches observations

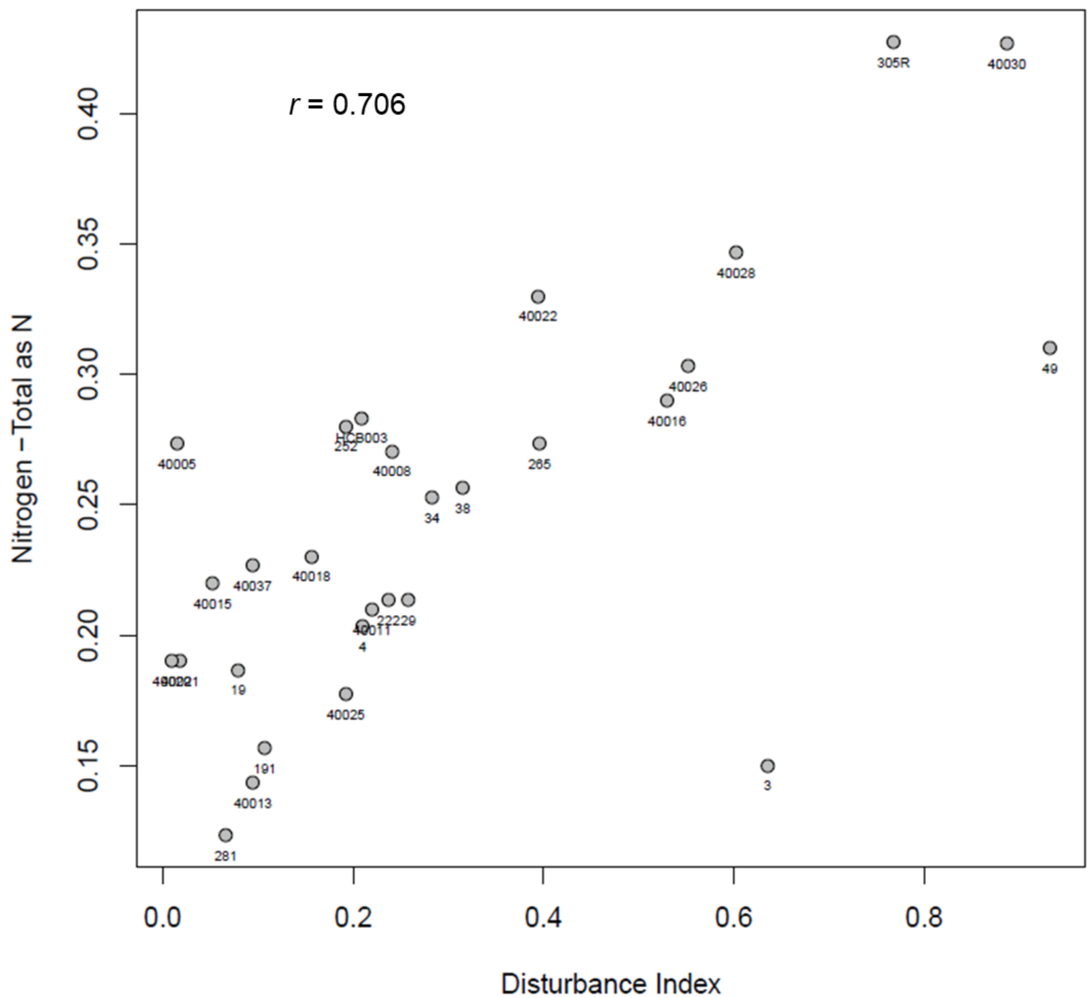


Figure D3. Comparison of Disturbance Index and total nitrogen content of sediments.
The numbers on the graph are station identifications.

Appendix E. Glossaries, acronyms, and abbreviations

Glossary of General Terms

Ambient: Background or away from point sources of contamination. Surrounding environmental condition.

Anthropogenic: Human-caused.

Azoic: Containing no living macrobenthic organisms.

Clean Water Act: A federal act passed in 1972 that contains provisions to restore and maintain the quality of the nation's waters. Section 303(d) of the Clean Water Act establishes the TMDL program.

Conductivity: A measure of water's ability to conduct an electrical current. Conductivity is related to the concentration and charge of dissolved ions in water.

Counterfactual: A probabilistic answer to a "what would have happened if" question.

Delphi procedure: An iterative procedure of ranking and feedback that can be used to improve agreement.

Dissolved oxygen (DO): A measure of the amount of oxygen dissolved in water.

Eutrophic: Nutrient rich and high in productivity resulting from human activities such as fertilizer runoff and leaky septic systems.

Eutrophication: Process by which waterbodies become eutrophic.

Graphical causal model: A set of nodes and edges in which the nodes represent random variables and the edges are depicted as arrows and represent the directional flow of casual information. These are also called Structural Causal Models (SCM) in the causal inference literature.

Interventional distribution: The population-level probability distribution calculated from a parameterized graphical causal model that results from simulating a causal intervention in the system. For example, "what would be the predicted distribution of variable Y if we set the value of X to x' ?"

Kappa analysis: A statistical procedure to measure reliability of agreement of multiple persons making qualitative judgments.

Leave-one-out cross-validation: A method to determine the predictive capacity of a model that consists of sequentially removing one sample from the training set, fitting the model, and using the resulting model to predict the response variable of removed sample.

Markov Chain Monte Carlo: A numerical method for estimating integrals of complicated integrands. It is used most often for estimation in Bayesian statistics to estimate posterior distributions.

Naïve Bayes: A categorization procedure which makes use of the simplifying assumption that samples can be categorized by a set of predictors and are independent of the assignment of any other sample.

Nonpoint source: Pollution that enters any waters of the state from any dispersed land-based or water-based activities, including but not limited to atmospheric deposition, surface-water runoff from agricultural lands, urban areas, or forest lands, subsurface or underground sources, or discharges from boats or marine vessels not otherwise regulated under the NPDES program. Generally, any unconfined and diffuse source of contamination. Legally, any source of water pollution that does not meet the legal definition of "point source" in section 502(14) of the Clean Water Act.

Nutrient: Substance such as carbon, nitrogen, and phosphorus used by organisms to live and grow. Too many nutrients in the water can promote algal blooms and rob the water of oxygen vital to aquatic organisms.

Pielou's evenness: A measure of equitability of species abundances in a sample of organisms.

pH: A measure of the acidity or alkalinity of water. A low pH value (0 to 7) indicates that an acidic condition is present, while a high pH (7 to 14) indicates a basic or alkaline condition. A pH of 7 is considered to be neutral. Since the pH scale is logarithmic, a water sample with a pH of 8 is ten times more basic than one with a pH of 7.

Point source: Source of pollution that discharges at a specific location from pipes, outfalls, and conveyance channels to a surface water. Examples of point source discharges include municipal wastewater treatment plants, municipal stormwater systems, industrial waste treatment facilities, and construction sites where more than 5 acres of land have been cleared.

Pollution: Contamination or other alteration of the physical, chemical, or biological properties of any waters of the state. This includes change in temperature, taste, color, turbidity, or odor of the waters. It also includes discharge of any liquid, gaseous, solid, radioactive, or other substance into any waters of the state. This definition assumes that these changes will, or are likely to, create a nuisance or render such waters harmful, detrimental, or injurious to (1) public health, safety, or welfare, or (2) domestic, commercial, industrial, agricultural, recreational, or other legitimate beneficial uses, or (3) livestock, wild animals, birds, fish, or other aquatic life.

Sediment: Soil and organic matter that is covered with water (for example, river or lake bottom).

Structural Equation Model: A set of parameter estimation techniques that takes as input a graphical causal model and data on observed variables and uses constraints on the correlations structure determined by the causal topology to estimate parameters.

Swartz dominance index: Minimum number of taxa that account for 75% of the abundance in a sample of organisms.

Taxa richness: Number of taxa in a sample of organisms.

Turbidity: A measure of water clarity. High levels of turbidity can have a negative impact on aquatic life.

Vital Sign: Measures of ecosystem health that guide the assessment of progress toward Puget Sound recovery goals.

303(d) list: Section 303(d) of the federal Clean Water Act, requiring Washington State to periodically prepare a list of all surface waters in the state for which beneficial uses of the water – such as for drinking, recreation, aquatic habitat, and industrial use – are impaired by pollutants. These are water quality-limited estuaries, lakes, and streams that fall short of state surface water quality standards and are not expected to improve within the next two years.

Acronyms and Abbreviations

AMBI	AZTI Marine Biotic Index
ANOVA	Analysis of variance
BQI	Benthic Quality Index
BRI	Benthic Response Index
C:N	Carbon-to-nitrogen
CBRAT	Coastal Biodiversity Risk Assessment Tool
CSL	Cleanup Screening Levels
CTD	Conductivity-Temperature-Depth sensor
DO	Dissolved oxygen
DPSIR	Driver-Pressure-State-Impact-Response model
EAP	Environmental Assessment Program
EAP-MW	Environmental Assessment Program Marine Waters database
Ecology	Washington State Department of Ecology
EIM	Environmental Information Management database
ENSO	El Niño/Southern Oscillation
EPA	U.S. Environmental Protection Agency
e.g.	For example
et al.	And others
GCM	Graphical causal model
i.e.	In other words
M-AMBI	Multivariate AMBI
MCMC	Markov Chain Monte Carlo
MMU	Marine Monitoring Unit
MQO	Measurement quality objective
MSMP	Marine Sediment Monitoring Program
MSMT	Marine Sediment Monitoring Team
NA	Not applicable
NOAA	National Oceanic and Atmospheric Administration
NPGO	North Pacific Gyre Oscillation
NWIFC	Northwest Indian Fisheries Commission
NTU	Nephelometric turbidity unit
O/E	Observed over expected
PAH	Polycyclic aromatic hydrocarbon
PAR	Photosynthetically active radiation
PDO	Pacific Decadal Oscillation
PSEMP	Puget Sound Ecosystem Monitoring Program
PSP	Puget Sound Partnership
PSWQA	Puget Sound Water Quality Authority
QA	Quality assurance

QAMP	Quality assurance monitoring plan
QAPP	Quality assurance project plan
QC	Quality control
RBI	Relative Benthic Index
RIVPACS	River InVertebrate Prediction And Classification System
RPD	Relative percent difference
RSD	Relative standard deviation
SBI	Sediment Benthic Index
SCCWRP	Southern California Coastal Water Research Project
SDI	Swartz Dominance Index
SEDQUAL	Sediment Quality database
SEM	Structural Equation Model
SMS	Sediment Management Standards
SQS	Sediment Quality Standards
SUNA	Submersible Ultraviolet Nitrate Analyzer
TAG	Technical Advisory Group
TOC	Total organic carbon
USGS	United States Geological Survey
WAC	Washington Administrative Code
WOS	Western Operations Section

Quality Assurance Glossary

Accreditation: A certification process for laboratories, designed to evaluate and document a lab's ability to perform analytical methods and produce acceptable data. For Ecology, it is "Formal recognition by (Ecology)…that an environmental laboratory is capable of producing accurate analytical data." [WAC 173-50-040] (Kammin, 2010)

Accuracy: The degree to which a measured value agrees with the true value of the measured property. USEPA recommends that this term not be used, and that the terms *precision* and *bias* be used to convey the information associated with the term *accuracy* (USGS, 1998).

Analyte: An element, ion, compound, or chemical moiety (pH, alkalinity) which is to be determined. The definition can be expanded to include organisms, e.g., fecal coliform, Klebsiella (Kammin, 2010).

Bias: The difference between the sample mean and the true value. Bias usually describes a systematic difference reproducible over time and is characteristic of both the measurement system and the analyte(s) being measured. Bias is a commonly used data quality indicator (DQI) (Kammin, 2010; Ecology, 2004).

Blank: A synthetic sample, free of the analyte(s) of interest. For example, in water analysis, pure water is used for the blank. In chemical analysis, a blank is used to estimate the analytical response to all factors other than the analyte in the sample. In general, blanks are used to assess possible contamination or inadvertent introduction of analyte during various stages of the sampling and analytical process (USGS, 1998).

Calibration: The process of establishing the relationship between the response of a measurement system and the concentration of the parameter being measured (Ecology, 2004).

Check standard: A substance or reference material obtained from a source independent from the source of the calibration standard; used to assess bias for an analytical method. This is an obsolete term, and its use is highly discouraged. See Calibration Verification Standards, Lab Control Samples (LCS), Certified Reference Materials (CRM), and/or spiked blanks. These are all check standards but should be referred to by their actual designator, e.g., CRM, LCS (Kammin, 2010; Ecology, 2004).

Comparability: The degree to which different methods, data sets and/or decisions agree or can be represented as similar; a data quality indicator (USEPA, 1997).

Completeness: The amount of valid data obtained from a project compared to the planned amount. Usually expressed as a percentage. A data quality indicator (USEPA, 1997).

Continuing Calibration Verification Standard (CCV): A quality control (QC) sample analyzed with samples to check for acceptable bias in the measurement system. The CCV is usually a midpoint calibration standard that is re-run at an established frequency during the course of an analytical run (Kammin, 2010).

Control chart: A graphical representation of quality control results demonstrating the performance of an aspect of a measurement system (Kammin, 2010; Ecology 2004).

Control limits: Statistical warning and action limits calculated based on control charts. Warning limits are generally set at +/- 2 standard deviations from the mean, action limits at +/- 3 standard deviations from the mean (Kammin, 2010).

Data integrity: A qualitative DQI that evaluates the extent to which a data set contains data that is misrepresented, falsified, or deliberately misleading (Kammin, 2010).

Data quality indicators (DQI): Commonly used measures of acceptability for environmental data. The principal DQIs are precision, bias, representativeness, comparability, completeness, sensitivity, and integrity (USEPA, 2006).

Data quality objectives (DQO): Qualitative and quantitative statements derived from systematic planning processes that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions (USEPA, 2006).

Data set: A grouping of samples organized by date, time, analyte, etc. (Kammin, 2010).

Data validation: An analyte-specific and sample-specific process that extends the evaluation of data beyond data verification to determine the usability of a specific data set. It involves a detailed examination of the data package, using both professional judgment and objective criteria, to determine whether the MQOs for precision, bias, and sensitivity have been met. It may also include an assessment of completeness, representativeness, comparability, and integrity, as these criteria relate to the usability of the data set. Ecology considers four key criteria to determine if data validation has actually occurred. These are:

- Use of raw or instrument data for evaluation.
- Use of third-party assessors.
- Data set is complex.

- Use of EPA Functional Guidelines or equivalent for review.

Examples of data types commonly validated would be**:**

- Gas Chromatography (GC).
- Gas Chromatography-Mass Spectrometry (GC-MS).
- Inductively Coupled Plasma (ICP).

The end result of a formal validation process is a determination of usability that assigns qualifiers to indicate usability status for every measurement result. These qualifiers include**:**

- No qualifier – data are usable for intended purposes.
- J (or a J variant) – data are estimated, may be usable, may be biased high or low.
- REJ – data are rejected, cannot be used for intended purposes.
  (Kammin, 2010; Ecology, 2004).

Data verification: Examination of a data set for errors or omissions, and assessment of the Data Quality Indicators related to that data set for compliance with acceptance criteria (MQOs). Verification is a detailed quality review of a data set (Ecology, 2004).

Detection limit (limit of detection)**:** The concentration or amount of an analyte which can be determined to a specified level of certainty to be greater than zero (Ecology, 2004).

Duplicate samples: Two samples taken from and representative of the same population, and carried through and steps of the sampling and analytical procedures in an identical manner. Duplicate samples are used to assess variability of all method activities including sampling and analysis (USEPA, 1997).

Field blank: A blank used to obtain information on contamination introduced during sample collection, storage, and transport (Ecology, 2004).

Initial Calibration Verification Standard (ICV): A QC sample prepared independently of calibration standards and analyzed along with the samples to check for acceptable bias in the measurement system. The ICV is analyzed prior to the analysis of any samples (Kammin, 2010).

Laboratory Control Sample (LCS): A sample of known composition prepared using contaminant-free water or an inert solid that is spiked with analytes of interest at the midpoint of the calibration curve or at the level of concern. It is prepared and analyzed in the same batch of regular samples using the same sample preparation method, reagents, and analytical methods employed for regular samples (USEPA, 1997).

Matrix spike: A QC sample prepared by adding a known amount of the target analyte(s) to an aliquot of a sample to check for bias due to interference or matrix effects (Ecology, 2004).

Measurement Quality Objectives (MQOs)**:** Performance or acceptance criteria for individual data quality indicators, usually including precision, bias, sensitivity, completeness, comparability, and representativeness (USEPA, 2006).

Measurement result: A value obtained by performing the procedure described in a method (Ecology, 2004).

Method: A formalized group of procedures and techniques for performing an activity (e.g., sampling, chemical analysis, data analysis), systematically presented in the order in which they are to be executed (EPA, 1997).

Method blank: A blank prepared to represent the sample matrix, prepared and analyzed with a batch of samples. A method blank will contain all reagents used in the preparation of a sample, and the same preparation process is used for the method blank and samples (Ecology, 2004; Kammin, 2010).

Method Detection Limit (MDL): This definition for detection was first formally advanced in 40CFR 136, October 26, 1984 edition. MDL is defined there as the minimum concentration of an analyte that, in a given matrix and with a specific method, has a 99% probability of being identified, and reported to be greater than zero (Federal Register, October 26, 1984).

Percent Relative Standard Deviation (%RSD): A statistic used to evaluate precision in environmental analysis. It is determined in the following manner**:**

$$\%RSD = (100 * s)/x$$

where s is the sample standard deviation and x is the mean of results from more than two replicate samples (Kammin, 2010).

Parameter: A specified characteristic of a population or sample. Also, an analyte or grouping of analytes. Benzene and nitrate + nitrite are all parameters (Kammin, 2010; Ecology, 2004).

Population: The hypothetical set of all possible observations of the type being investigated (Ecology, 2004).

Precision: The extent of random variability among replicate measurements of the same property; a data quality indicator (USGS, 1998).

Quality assurance (QA): A set of activities designed to establish and document the reliability and usability of measurement data (Kammin, 2010).

Quality Assurance Project Plan (QAPP): A document that describes the objectives of a project, and the processes and activities necessary to develop data that will support those objectives (Kammin, 2010; Ecology, 2004).

Quality control (QC): The routine application of measurement and statistical procedures to assess the accuracy of measurement data (Ecology, 2004).

Relative Percent Difference (RPD): RPD is commonly used to evaluate precision. The following formula is used**:**

$$[Abs(a-b)/((a + b)/2)] * 100$$

where "Abs()" is absolute value and a and b are results for the two replicate samples. RPD can be used only with 2 values. Percent Relative Standard Deviation is (%RSD) is used if there are results for more than 2 replicate samples (Ecology, 2004).

Replicate samples: Two or more samples taken from the environment at the same time and place, using the same protocols. Replicates are used to estimate the random variability of the material sampled (USGS, 1998).

Representativeness: The degree to which a sample reflects the population from which it is taken; a data quality indicator (USGS, 1998).

Sample (field): A portion of a population (environmental entity) that is measured and assumed to represent the entire population (USGS, 1998).

Sample (statistical): A finite part or subset of a statistical population (USEPA, 1997).

Sensitivity: In general, denotes the rate at which the analytical response (e.g., absorbance, volume, meter reading) varies with the concentration of the parameter being determined. In a specialized sense, it has the same meaning as the detection limit (Ecology, 2004).

Spiked blank: A specified amount of reagent blank fortified with a known mass of the target analyte(s); usually used to assess the recovery efficiency of the method (USEPA, 1997).

Spiked sample: A sample prepared by adding a known mass of target analyte(s) to a specified amount of matrix sample for which an independent estimate of target analyte(s) concentration is available. Spiked samples can be used to determine the effect of the matrix on a method's recovery efficiency (USEPA, 1997).

Split sample: A discrete sample subdivided into portions, usually duplicates (Kammin, 2010).

Standard Operating Procedure (SOP): A document which describes in detail a reproducible and repeatable organized activity (Kammin, 2010).

Surrogate: For environmental chemistry, a surrogate is a substance with properties similar to those of the target analyte(s). Surrogates are unlikely to be native to environmental samples. They are added to environmental samples for quality control purposes, to track extraction efficiency and/or measure analyte recovery. Deuterated organic compounds are examples of surrogates commonly used in organic compound analysis (Kammin, 2010).

Systematic planning: A step-wise process which develops a clear description of the goals and objectives of a project, and produces decisions on the type, quantity, and quality of data that will be needed to meet those goals and objectives. The DQO process is a specialized type of systematic planning (USEPA, 2006).

References for QA Glossary

Ecology, 2004. Guidance for the Preparation of Quality Assurance Project Plans for Environmental Studies. Washington State Department of Ecology, Olympia, WA. https://apps.ecology.wa.gov/publications/SummaryPages/0403030.html.

Kammin, B., 2010. Definition developed or extensively edited by William Kammin, 2010. Washington State Department of Ecology, Olympia, WA.

USEPA, 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process EPA QA/G-4. http://www.epa.gov/quality/qs-docs/g4-final.pdf.

USGS, 1998. Principles and Practices for Quality Assurance and Quality Control. Open-File Report 98-636. U.S. Geological Survey. http://ma.water.usgs.gov/fhwa/products/ofr98-636.pdf.