Results of Validation Exercise for Marine Benthic Index

Environmental Assessment Program

Publication 23-03-009

Authors:

Valerie Partridge, WA State Department of Ecology,
Donald Schoolmaster, U.S. Geological Survey

March 2023

Abstract

Marine benthic invertebrates (benthos) are key components of the Puget Sound ecosystem. Because of their direct association living in, and sometimes consuming, sediments, benthos can be valuable sentinels of ecosystem health. Therefore, indicators of benthic invertebrate community health can serve as direct measures of sediment and water quality.

In 2021, the Puget Sound Partnership funded development of a *Marine Benthic Index*. The *Marine Benthic Index* thus developed uses a novel approach that accounts for habitat preferences of the benthic invertebrate species. This report describes the design and results of the exercise conducted to validate the *Marine Benthic Index*.

The goals of the validation exercise were to determine (a) how well the *Marine Benthic Index* matches more standard ways of assessing community health and (b) how finely it is possible to distinguish between levels of disturbance. A controlled experiment was devised in which simulated benthic communities were generated to correspond to predetermined levels of disturbance, and experts in benthic ecology determined which communities reflected the more-disturbed conditions. In this way, the index was directly compared to traditional methods of assessing benthic communities.

The results provide strong evidence that the "latent disturbance" model used to derive the *Marine Benthic Index* is identifying effects that benthic experts recognize as disturbance. Not only did the model agree with the experts overall, but also the probability of agreement strongly increased with increasing difference in disturbance level.

The validation exercise results indicate that the *Marine Benthic Index* is a reliable method of determining disturbance without the necessity of assuming *a priori* knowledge of the disturbance. Furthermore, the numerical approach embodied in the *Marine Benthic Index* has the advantage of being able to find patterns beyond the capability of individual experts to know the effects of human disturbances for all species under all environmental conditions.

Publication Information

This report is available on the Department of Ecology's website at: https://apps.ecology.wa.gov/publications/SummaryPages/2303009.html.

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

Suggested Citation:
Partridge, V. and D. Schoolmaster. 2023. Results of Validation Exercise for Marine Benthic Index. Publication 23-03-009. Washington State Department of Ecology, Olympia. https://apps.ecology.wa.gov/publications/SummaryPages/2303009.html.

Contact Information

Publication 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

Background

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.

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, despite multiple attempts over 30 years (Partridge and Schoolmaster, 2022).

In 2021, the Puget Sound Partnership (PSP) funded development of a *Marine Benthic Index* to fill this gap, as part of the PSP's *Monitoring to Accelerate Recovery* initiative. The *Marine Benthic Index* will be used to assess and communicate status and trends of benthic invertebrate community health, as an indicator of the Puget Sound Marine Water Vital Sign.

The *Marine Benthic Index* has been developed using an approach that accounts for habitat preferences of the benthic invertebrate species (Partridge and Schoolmaster, 2022). The index development made use of a rich dataset from over 30 years of Puget Sound ambient marine sediment monitoring.

A critical step in the acceptance of the *Marine Benthic Index* is validation of the index. Because the approach of this index differs from traditional methods of assessing benthic invertebrate community health, it is important to demonstrate that the index will perform at least as well as other methods. This report describes the validation exercise undertaken and its results.

Methods

Index approach

The approach for the *Marine Benthic Index* is to model species occurrence as a function of environment (E) and human disturbance (D). The environmental variables in E (measured) are those unlikely to be affected by human activity on the temporal and spatial scales of regular Puget Sound sediment monitoring—those we call the "habitat" variables. Human disturbance, D, is unobserved, but is estimated from variation within sites across species.

The *Marine Benthic Index* is based on individual-taxon models of responses of 128 selected taxa to 8 field-collected habitat parameters to predict what species we would expect, or not expect, given the habitat conditions (Appendix A). The models were "trained" (i.e., model coefficients were estimated) with 168 samples collected at 50 sites throughout the southern Salish Sea (Puget Sound proper and embayments of the San Juan Islands, eastern Strait of Juan de Fuca, and southern Strait of Georgia) in April 2016 and April 2017.

The 128 taxa (Appendix B) were chosen based on how much information they provide about their habitat preferences, plus selected taxa of importance based on consultation with local experts in marine benthic invertebrate taxonomy and ecology. All chosen taxa have been found in Puget Sound samples since 1997, in both spring (April) and early summer (June), the timing of sampling.

The habitat variables (Appendix C) were chosen from available field-collected data. Environmental variables thought not to be influenced by human activity on the scale of ordinary ambient sediment monitoring were included in E. Variables such as total organic carbon were excluded from the calibration and validation, but these variables are included in plans for future work as potential indicators of disturbance (Partridge and Schoolmaster, 2022).

Validation exercise

The goal of this validation exercise was to determine how well the *Marine Benthic Index* matches standard methods of assessing community health. We also sought to determine how finely experts could distinguish between levels of disturbance. For that reason, we used simulated disturbance, habitat, and community data, so we would know exactly what the level of disturbance was for each sample.

The questions we wanted to be able to answer were:

- How does the concordance between expert opinion and model output vary as a function of disturbance?
- Given the variation inherent in expert opinion and the variation inherent in model structures, what is the smallest difference in disturbance that can be reliably discerned by expert evaluation?

Because human disturbance is not something we can measure directly, **trying to validate an index using natural samples necessarily involves assumptions about what is and is not disturbance**. In the validation study, we are testing concordance in disturbance levels between model output and expert opinion for a set of samples with known levels of disturbance.

Using the parameter estimates from the training data, we simulated samples for the validation exercise by (1) specifying habitat characteristics, (2) specifying a disturbance level, (3) adding residual variation, and (4) probabilistically determining the presence/abundance of each taxon of 128 taxa in the simulated sample. The habitat parameters were selected randomly from a multivariate model of the 8 habitat variables based on the 2016-2017 dataset, in which the correlations among variables were preserved.

From each of 108 sets of randomly selected habitat variables, two simulated communities were generated from D values symmetrically about zero, as described above. $D=0$ represents the reference condition of the mean of the baseline Puget Sound-wide dataset used to calibrate the index model. Values of $D > 0$ indicate conditions less disturbed than the mean; values of $D < 0$ indicate conditions more disturbed than the mean.

From the 108 matched pairs of communities, 30 pairs were chosen at approximately equal intervals of |D| for the exercise. All taxa and all habitat variables were examined for all selected simulated samples to ensure that the values were realistic, by comparison to distributions of taxa and habitat variables for 840 samples collected from 1997 to 2019, split into those with positive and negative values of D. The final |D| ranged from 0.0218 to 4.968 (Figure 1; Appendix D).

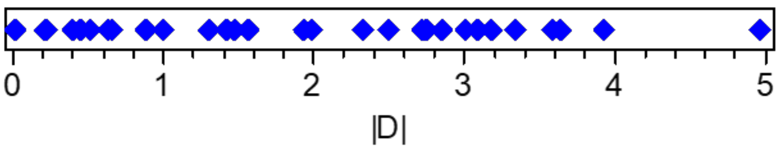

Figure 1. Absolute values of Disturbance level used to generate simulated communities. D=0 represents the mean of the Puget Sound-wide baseline used to calibrate the index model. For each of 30 random habitats, two communities were generated: one corresponding to the positive value of D (i.e., healthier than the Puget Sound average) and the other corresponding to the negative value of D (i.e., more disturbed than the Puget Sound average).

To create a worksheet to present to the experts, the pair and community labels were randomized, so that there was no relationship between pair number (1, 2, ..., 30) or community letter (A, B) and value of D. The validation exercise consisted of choosing which community in each pair was the more disturbed, i.e., which had the negative value of D (Figure 2). Reference in this report to correctness of assessments by the benthic experts is narrowly constrained to mean only whether their assessments agreed with the predetermined community conditions in this controlled study.

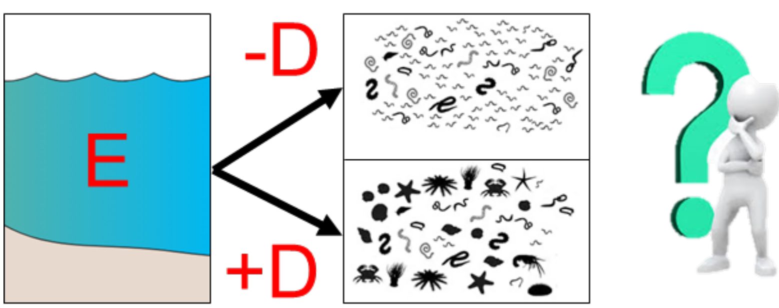

Figure 2. Schematic representation of validation exercise for one simulated benthic community pair.

Eight benthic experts completed the validation exercise (Appendix E). Their responses are anonymized in this report, so there is no relationship between name order and Expert #.

The benthic experts were asked to work independently to (a) determine which of the two communities in each pair was more disturbed and (b) provide their rationale and level of confidence in their answer (Appendix F). They were provided an Excel workbook with the habitat data and simulated species abundances for all 30 pairs (Appendix G), as well as an answer sheet (Appendix H).

Results

The assessments by the benthic experts (Appendix I) agreed with the known disturbance index much of the time (65.5% on average; Table 1). Of the 30 simulated relatively disturbed communities, the number correctly identified by an individual expert as being more disturbed ranged from 17 to 22 (i.e., 56.7% to 73.3%) correct. (We reiterate that reference to correctness of assessments by the benthic experts means only whether their assessments agreed with the predetermined values of D.) More than half of the experts correctly identified the more disturbed community in 20 of the 30 pairs. There were two pairs for which only one benthic expert agreed with the disturbance index and four pairs for which none of the benthic experts agreed with the disturbance index.

The benthic experts' answers agreed with each other completely for 12 of the 30 pairs. Fleiss' Kappa, a measure of the reliability of agreement of independent assessors amongst themselves, was 0.55, a moderate level of agreement. For only 8 pairs, however, were all the benthic experts in complete agreement with each other and the benthic index (Fleiss' Kappa = 0.3).

Table 1. Results of benthic expert assessments.

Pair	More disturbed community (specified)	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Percent correct
1	B	B	B	A	B	A	B	B	B	75.0%
2	B	B	B	B	B	B	B	B	B	100.0%
3	B	A	A	A	A	B	A	A	A	12.5%
4	A	A	A	A	A	A	A	A	A	100.0%
5	A	B	B	B	B	B	B	B	B	0.0%
6	B	B	B	B	A	B	B	B	B	87.5%
7	A	A	A	B	A	A	B	A	A	75.0%
8	B	A	A	A	A	B	B	A	B	37.5%
9	B	A	A	A	A	A	A	A	A	0.0%
10	A	B	A	B	B	A	A	B	B	37.5%
11	B	B	B	B	B	B	B	B	B	100.0%
12	B	B	B	B	B	A	B	B	A	75.0%
13	B	B	B	B	A	B	B	B	B	87.5%
14	A	B	A	B	A	B	A	A	A	62.5%
15	B	B	B	B	B	B	B	B	B	100.0%
16	B	B	B	B	B	B	B	B	B	100.0%
17	A	A	A	A	A	A	A	A	A	100.0%
18	B	B	A	A	A	B	B	A	A	37.5%
19	A	A	A	A	A	A	B	B	A	75.0%
20	B	B	B	B	A	B	B	B	B	87.5%
21	A	B	B	B	B	B	B	B	B	0.0%
22	B	A	A	A	A	A	A	A	B	12.5%
23	B	B	B	B	B	B	B	B	B	100.0%
24	A	A	A	A	A	A	A	A	B	87.5%
25	A	A	A	A	A	A	A	A	A	100.0%
26	A	A	A	A	A	A	A	A	B	87.5%
27	A	A	A	A	A	A	A	B	A	87.5%
28	B	B	B	B	A	B	B	B	B	87.5%
29	A	B	B	B	A	A	A	A	B	50.0%
30	A	B	B	B	B	B	B	B	B	0.0%
Percent correct		66.7%	70.0%	56.7%	56.7%	73.3%	73.3%	63.3%	63.3%	Average 65.4%

As expected, the probability of correctly identifying which community in each pair was more disturbed was a function of the level of disturbance (Figure 3).

The value of $|\Delta D|$ that corresponds to 50% probability of correctly identifying the more disturbed community is 1.75 ± 0.72 (Figure 3). In other words, the theoretical minimum distance at which the experts can distinguish more- vs. less-disturbed communities is $|\Delta D| = 1.75$. Since D is a z-score (Appendix A), this means that experts can distinguish communities about 1 to 2.5 standard deviations apart in disturbance level.

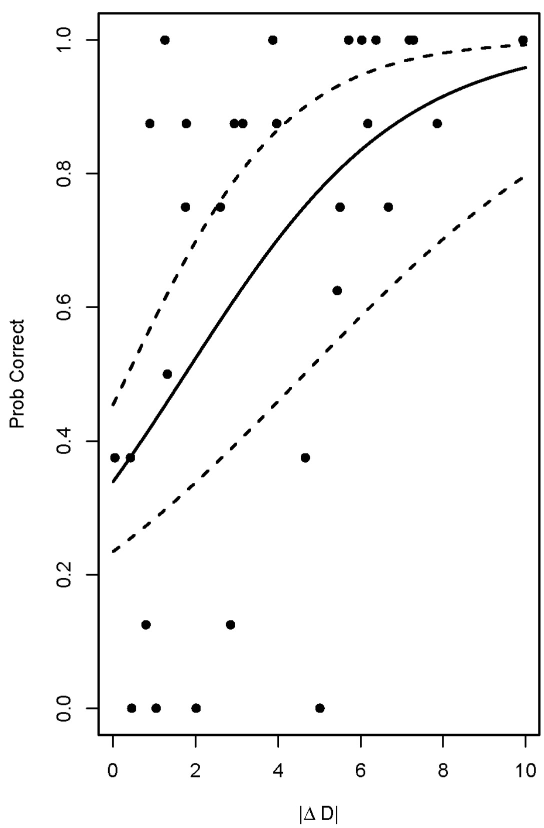

Figure 3. Probability of correctly identifying the more disturbed community as a function of the difference in disturbance level (|ΔD|).

Experts were more successful at identifying more-disturbed communities for deeper stations than for shallower stations (Figure 4a). The value of |ΔD| corresponding to 50% probability of correctly identifying the more disturbed community was 2.49 ± 0.86 at the 10^{th} percentile of station depth (i.e., at the shallowest stations), but 0.756 ± 0.83 (an interval that includes zero) at the 90^{th} percentile of station depth (i.e., at the deepest stations) (Figure 4b). In other words, the experts were much more likely to distinguish more- vs. less-disturbed communities for deep stations than for shallow stations. Further, at the deepest stations, experts were likely more discriminatory even at very small values of |ΔD|.

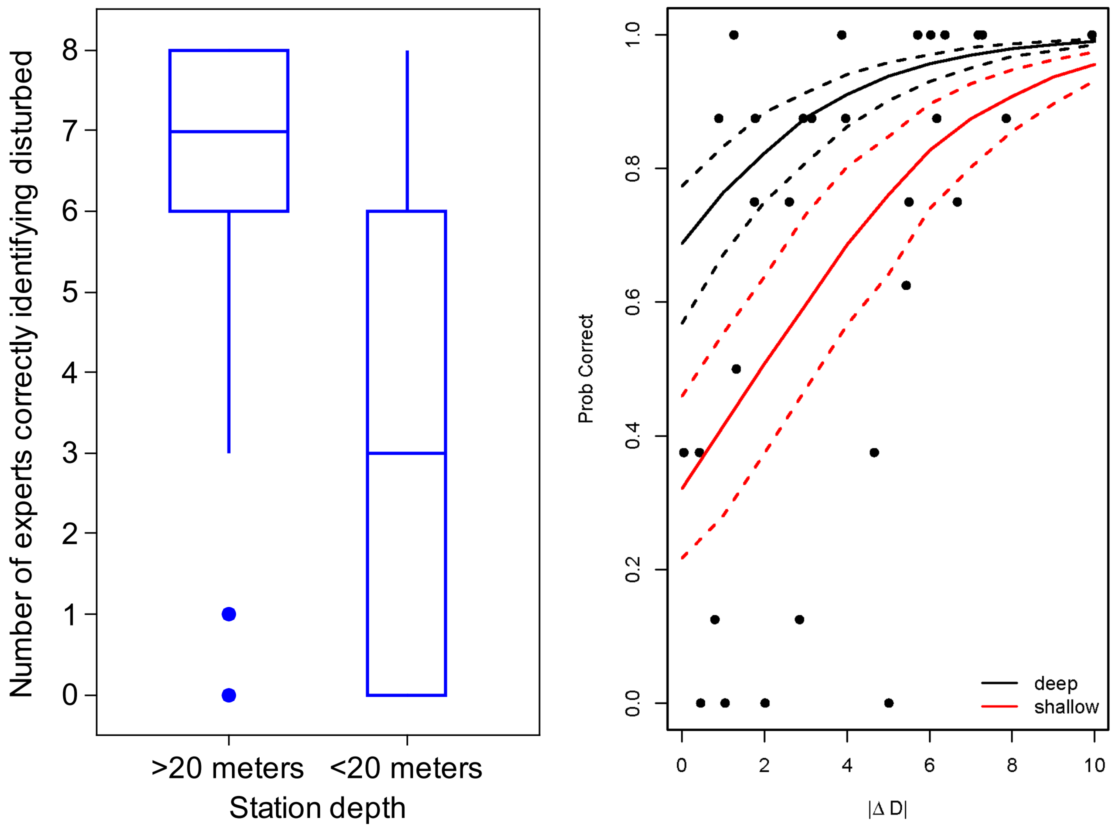

Figure 4.
a (left): Number of experts correctly identifying the more disturbed community in each pair as a function of station depth.
b (right): Probability of correctly identifying the more disturbed community as a function of both difference in disturbance level (|ΔD|) and station depth. The curves correspond to the shallowest and deepest depths.

Expanding the analysis to include the full set of environmental variables suggests that, all else being equal, the probability of experts' choosing the correct option increased with increasing salinity, percent fines (silt and clay), and, marginally, station depth. The probability decreased with increasing percent gravel and increasing grab penetration depth (Figure 5). The individual response rates over each variable individually are presented in Appendix J.

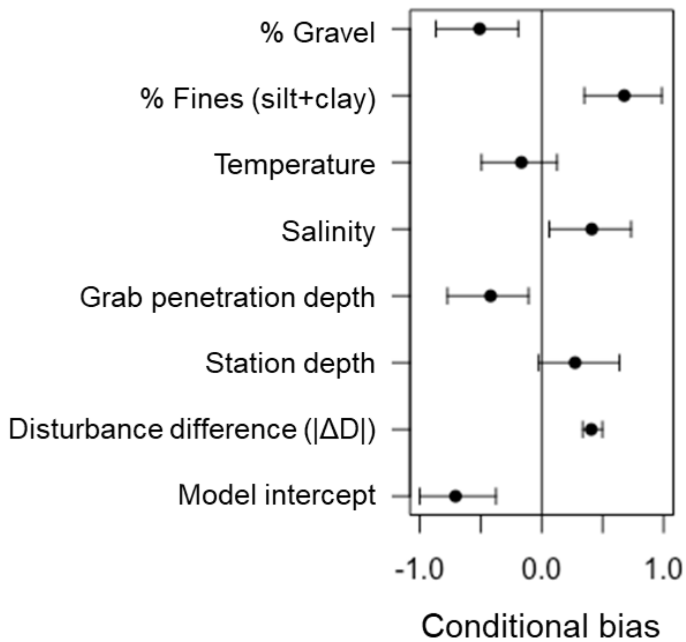

Figure 5. Bias in habitat variables and in disturbance difference (|ΔD|), conditional on the other variables.

Positive values of the bias parameter indicate that the experts were more likely to correctly determine the more-disturbed communities as the corresponding variable value increased, given the values of the other variables.

Negative bias parameter values indicate that the experts were less likely to correctly determine the more-disturbed communities as the corresponding variable value increased, given the values of the other variables.

The criteria used by the experts to support their decisions (Appendix I) were numerous. Categorizing responses into groups shows that the commonness of the different reasons ranged from rare (used by only a single expert) to unanimous (Table 2).

All the experts cited "diversity," though the definition employed was only sometimes defined (e.g., Shannon-Wiener H'), sometimes left unspecified. And all the experts mentioned occurrence, abundance, or dominance of tolerant or opportunistic taxa in their rationale.

Although both the occurrence/abundance and rarity/paucity of sensitive taxa were frequently used, only one expert used the rarity or paucity of tolerant taxa. Most of the experts used calculated community measures (e.g., richness, evenness) in their decision-making.

Only two experts mentioned characteristics of the habitat in their rationale. One expert performed ordination analyses and calculated M-AMBI for the communities.

Table 2. Criteria the experts used to decide between communities.

Reason	Number of experts
Diversity	8
Occurrence/abundance/dominance of tolerant/opportunistic taxa	8
Taxa richness	7
Total abundance	6
Specific taxa	6
Evenness	5
Dominance	5
Occurrence/abundance/dominance of sensitive/intolerant taxa	5
Absence/rarity/paucity of sensitive/intolerant taxa	5
Taxa inferring disturbance/pollution	5
Taxa inferring habitat/ecosystem complexity	3
Functional feeding guild (Macdonald et al., 2010, 2012)	3
Percentages/abundances/dominance of specific phyla	3
Habitat characteristics	2
Absence/rarity/paucity of tolerant/opportunistic taxa	1
M-AMBI (Pelletier et al., 2018), Ecological Group (Grall and Glémarec, 1997)	1
Ordination of communities	1
Habitat inferring specific taxa	1

Agreement of the experts with the predetermined disturbance level varied. Experts 2 and 7 generally were incorrect for only small differences in disturbance (Figure 6). Experts 5 and 6 tended to be able to distinguish correctly between less- and more-disturbed communities at smaller differences than the other experts did (medians of correct=Y boxplots in Figure 6).

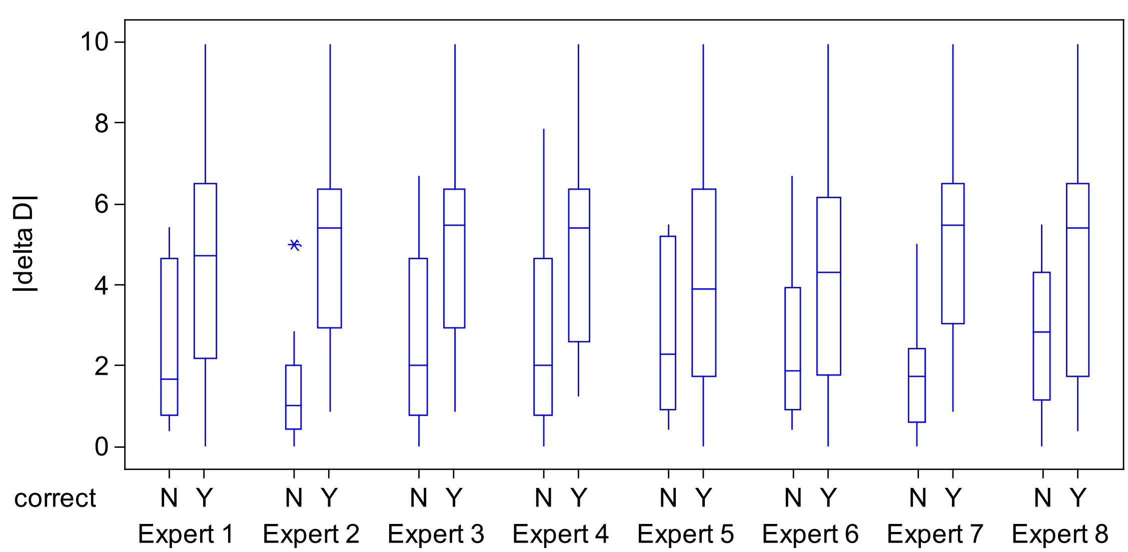

Figure 6. Difference in disturbance (|delta D|) vs. whether the more-disturbed communities were distinguished correctly. (N = no, Y = yes)

As anticipated, overall, the experts' confidence levels (self-assessed) increased with the difference in disturbance (Figure 7) and the number of correct responses (Figure 8). Except for the communities correctly distinguished by no or only a single expert, average confidence level increased with increasing agreement by the experts (Figure 8). In other words, as a group, the better able the experts were to correctly identify the more-disturbed community, the more their confidence reflected their skill. At an individual level, only Expert 5's confidence level tended to reflect their performance: higher for correct answers, lower for incorrect answers (see Figure K1 in Appendix K).

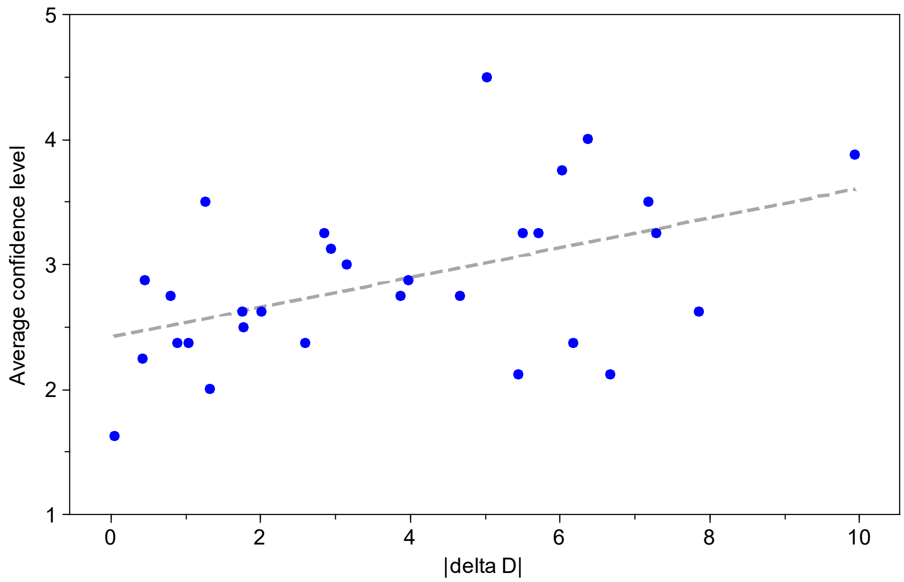

Figure 7. Average level of expert confidence (self-assessed) vs. the difference in disturbance level within the pair (|delta D|). Regression R^2 = 23.5%.

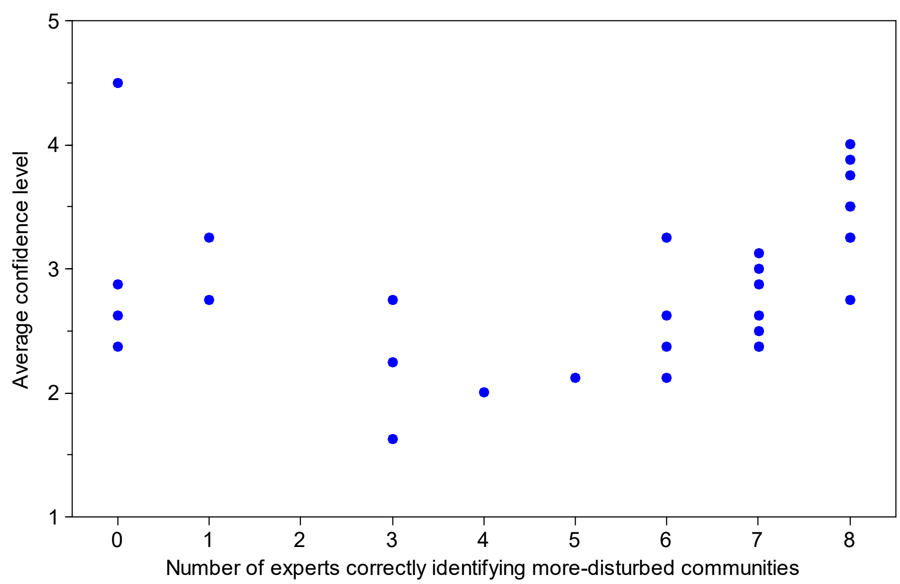

Figure 8. Average self-assessed confidence level vs. number of experts correctly identifying the more-disturbed communities.

Discussion

Overall, this validation exercise provides strong evidence that the "latent disturbance" model used to derive the *Marine Benthic Index* is identifying effects, quantified in *D,* that the benthic experts recognize as disturbance. This is evidenced both by the overall level of agreement between the model and the experts and especially by the strong increase in the probability of agreement with increasing difference in *D* between the pairs. Had the effects quantified by *D* not been recognized as disturbance by the experts, then we would expect no trend at all between the probability of agreement and the difference in *D* of the pairs.

In addition to the difference in *D*, analysis of the experts' choices found that the probability of correctly choosing the more disturbed community of the pair (i.e., assessment agreeing with predetermined value of *D*) varied also systematically with other environmental variables. Conditional on the other variables, experts were more likely to correctly determine the more-disturbed communities as percent fines or salinity increased but less likely as percent gravel or grab penetration depth increased.

This bias suggests that the experts' default schema is of deeper, higher-saline sites with high fines, low gravel, and low penetration. In other words, the combination of habitat variables most familiar to, or most easily recognized by, the experts in making their assessments is deep, salty, muddy but more compact sites. Combinations of habitat variables outside that may be harder for the experts to assess.

The relative difficulty of the experts' recognizing more-disturbed conditions in shallower locations, coarser sediments, or less-saline waters emphasizes the advantage of a numerical approach that can find patterns beyond the capability of individuals to know the effects of human disturbances for all species under all environmental conditions.

Conclusions

The positive performance of this disturbance index under controlled conditions gives us confidence that the index can be relied upon to provide reasonable assessments of benthic condition for real samples.

Acknowledgments

We greatly appreciate the time and effort of all the benthic experts who participated in the validation exercise.

References

Grall, J., and M. Glémarec, M. 1997. Using biotic indices to estimate macrobenthic community perturbations in the Bay of Brest. *Estuarine, Coastal and Shelf Science* 44(Supplement A):43-53.

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. doi: 10.3354/meps09478

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.

Pelletier, M.C., D.J. Gillett, A.T. Hamilton, T. Grayson, V. Hansen, E.W. Leppo, S.B. Weisberg, and Á. Borja. 2018. Adaptation and application of multivariate AMBI (M-AMBI) in US coastal waters. *Ecological Indicators* 89:818-827.

Appendix A: Index statistical method

The value of the index, *D*, is calculated as follows:

Let $P(S_i = 1|\, E) = \; p_i(E)$ denote the probability that taxon *i* occurs, given the habitat.

We estimate $p_i(E)$ using logistic regression: $logit(p_i(E)) = EB_i$, where B_i estimates the effect of the environmental factors *E* on the presence of taxon *i*, based on the training data.

The quantile residual, r_{ij}, for species *i* at site *j* is calculated as $\Phi((2a + b)/2)$, where $\Phi(\quad)$ is the cumulative distribution function of the standard normal,

$$a = min\{p_i(E), 1 - p_i(E)\}, \text{ and } b = \sum_{S=0}^{S=S_{ij}} P\left(S\middle|E_j\right).$$

The "sensitivity scores", α_i, and disturbance scores, D_j, are calculated to minimize the set of functions $r_{ij} = \alpha_i D_j$ for values of α_i and D_j.

D is assumed to be normally distributed with mean 0 and variance 1, which is equivalent to setting the scale of *D* to the standardized mean of *D* in the training data. Higher values of *D* indicate healthier, less disturbed environments, relative to the baseline, which has been chosen to be the mean of the data used to train the model. Lower values of *D* indicate environments which are more disturbed, less healthy than the baseline. The sensitivity parameter α_i indicates whether species *i* is more likely ($\alpha_i < 0$) or less likely ($\alpha_i > 0$) to be found in disturbed areas, *given the habitat*.

Appendix B: Index benthic invertebrate taxa

Excel file: Appendix_B_Benthic_invertebrate_taxa_used_in_index.xlsx

Contains Table B1, Taxa used in the model, with taxonomic hierarchy and assigned functional feeding guild (Macdonald et al., 2010, 2012).

Appendix C: Index habitat variables

Habitat variables used in the model:

- station depth below water surface (m)
- van Veen grab penetration depth (cm)
- salinity of overlying water in van Veen grab (ppt)
- temperature of surficial sediment (C)
- percent fines in surficial sediment (% silt + clay in top 2-3 cm)
- percent gravel in surficial sediment (% gravel in top 2-3 cm)
- presence of submerged aquatic vegetation (type unspecified)
- presence of shell hash

Appendix D: Exercise pair disturbance levels

Table D1. Disturbance levels (value of D) and difference (|ΔD|) for communities in each pair.

Pair	D value A	D value B	More disturbed	|Δ D|
1	1.3000	-1.3000	B	2.6000
2	2.8557	-2.8557	B	5.7114
3	1.4240	-1.4240	B	2.8479
4	-3.5899	3.5899	A	7.1798
5	-1.0065	1.0065	A	2.0131
6	1.4684	-1.4684	B	2.9368
7	-3.3364	3.3364	A	6.6728
8	0.2107	-0.2107	B	0.4215
9	2.5062	-2.5062	B	5.0124
10	-2.3295	2.3295	A	4.6590
11	0.6277	-0.6277	B	1.2554
12	2.7503	-2.7503	B	5.5006
13	3.0874	-3.0874	B	6.1747
14	-2.7174	2.7174	A	5.4348
15	4.9680	-4.9680	B	9.9359

Pair	D value A	D value B	More disturbed	|Δ D|
16	3.1867	-3.1867	B	6.3733
17	-3.6403	3.6403	A	7.2805
18	0.0218	-0.0218	B	0.0436
19	-0.8780	0.8780	A	1.7560
20	0.4455	-0.4455	B	0.8910
21	-0.5209	0.5209	A	1.0418
22	0.3979	-0.3979	B	0.7959
23	3.0136	-3.0136	B	6.0272
24	-1.9824	1.9824	A	3.9647
25	-1.9361	1.9361	A	3.8721
26	-1.5709	1.5709	A	3.1418
27	-0.8867	0.8867	A	1.7733
28	3.9283	-3.9283	B	7.8565
29	-0.6585	0.6585	A	1.3170
30	-0.2254	0.2254	A	0.4507

Appendix E: Exercise participants

The following benthic experts completed the validation exercise:

- Dany Burgess, Washington State Department of Ecology
- Wendy Eash-Loucks, King County Department of Natural Resources and Parks
- Steve Hulsman, SGH Group
- Michelle Knowlen, EcoAnalysts
- Oliver Miler, Northwest Indian Fisheries Commission
- Blair Paul, Skokomish Tribe
- Marguerite Pelletier, US Environmental Protection Agency
- Sandra Weakland, Washington State Department of Ecology

Appendix F: Exercise instructions

Word file: Appendix_F_Instructions_for_Marine_Benthic_Index_validation_exercise.docx

Appendix G: Exercise data

Excel file: Appendix_G_Validation_exercise_data_workbook.xlsx

Appendix H: Exercise answer sheet

Excel file: Appendix_H_Validation_exercise_response_worksheet.xlsx

Appendix I: Exercise expert responses, notes on methods

Excel file: Appendix_I_collected_expert_responses.xlsx

Appendix J: Exercise expert responses by habitat parameter

Figures J1-J6 and Table J1 display the distributions of the responses of the individual benthic experts in relation to the ranges of the habitat variables, independently of the other habitat variables.

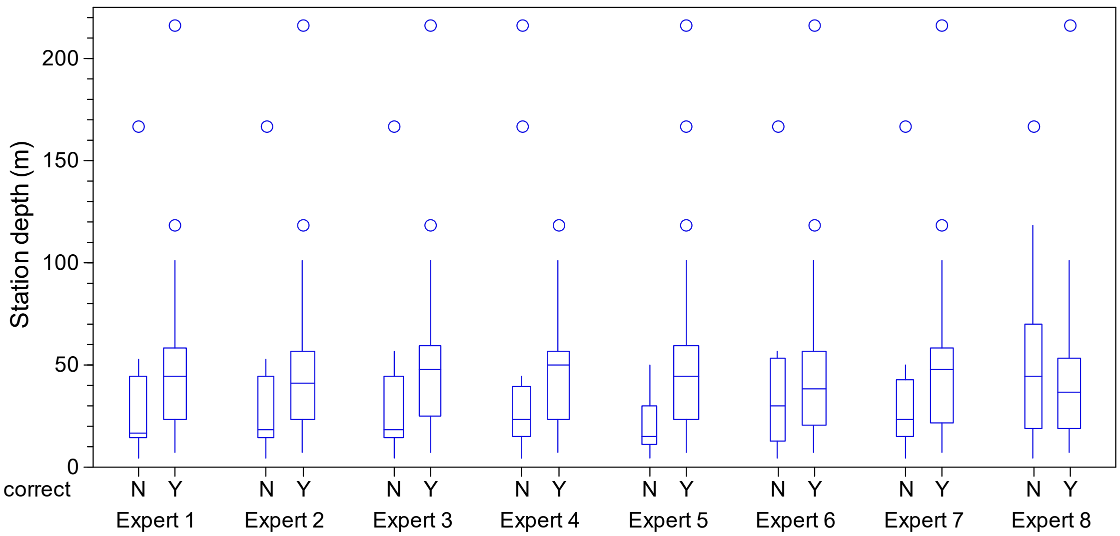

Figure J1. Station depth vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

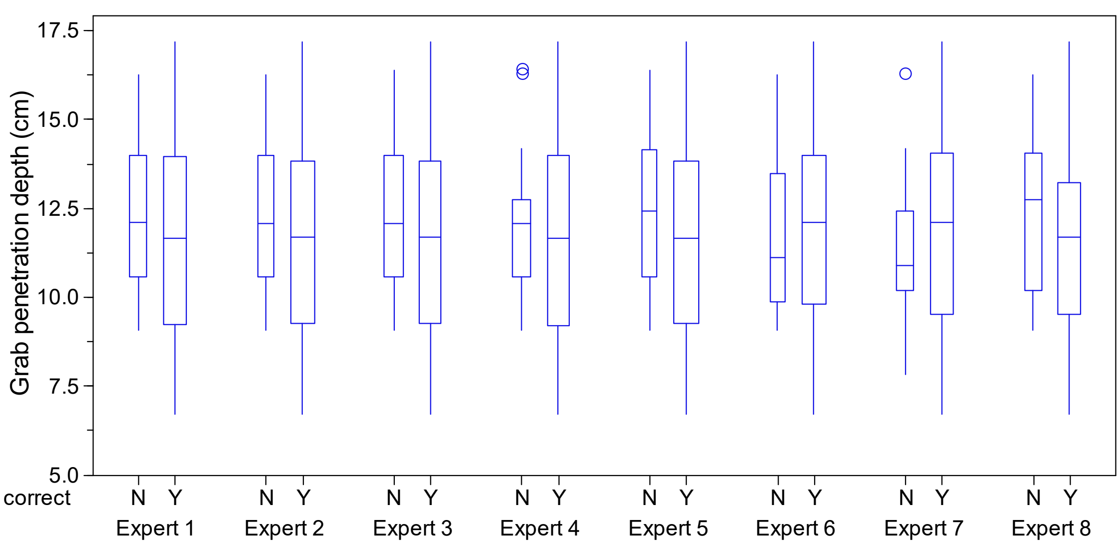

Figure J2. Grab penetration depth vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

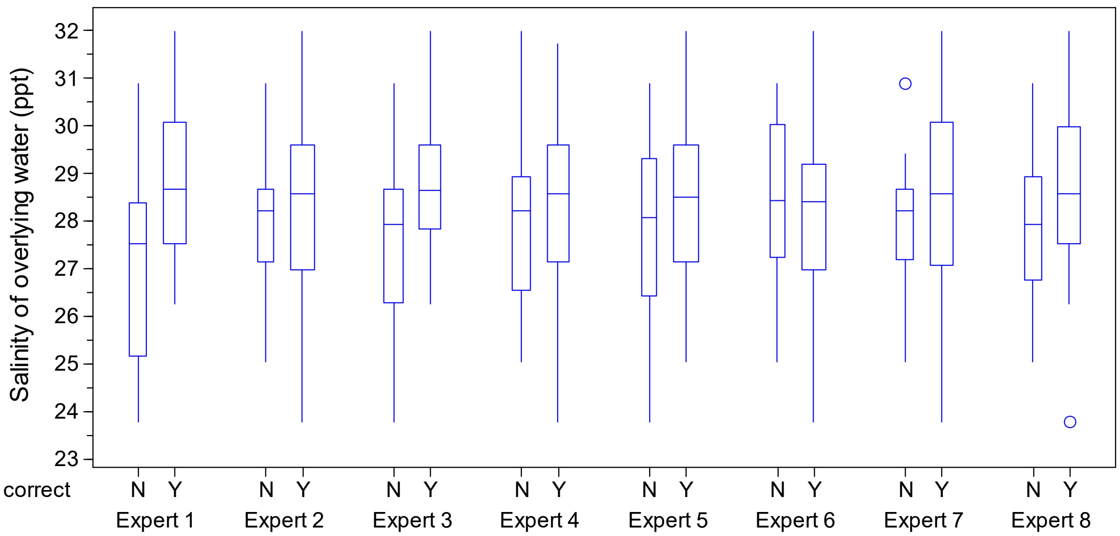

Figure J3. Salinity of overlying water vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

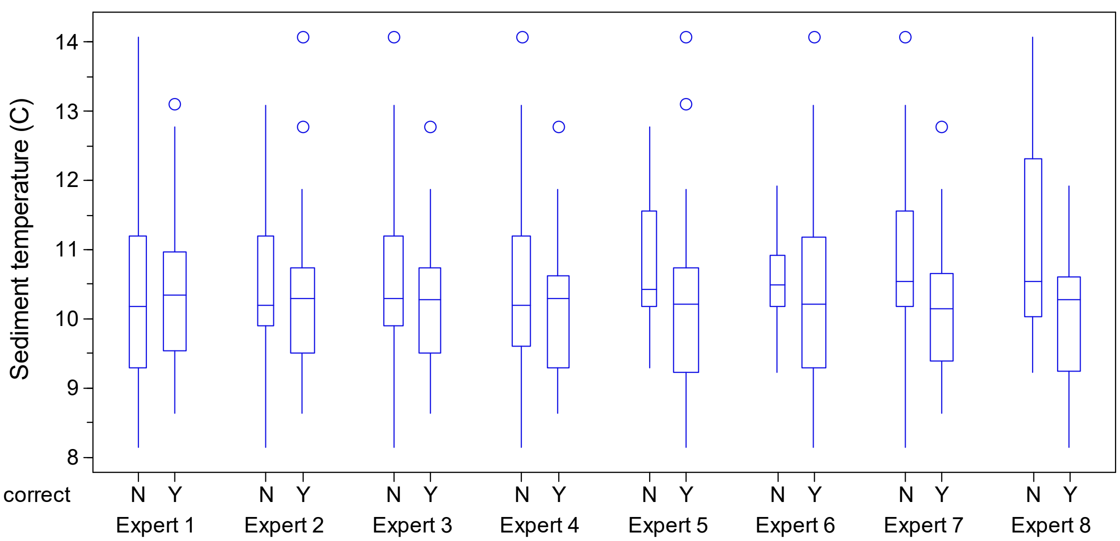

Figure J4. Sediment temperature vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

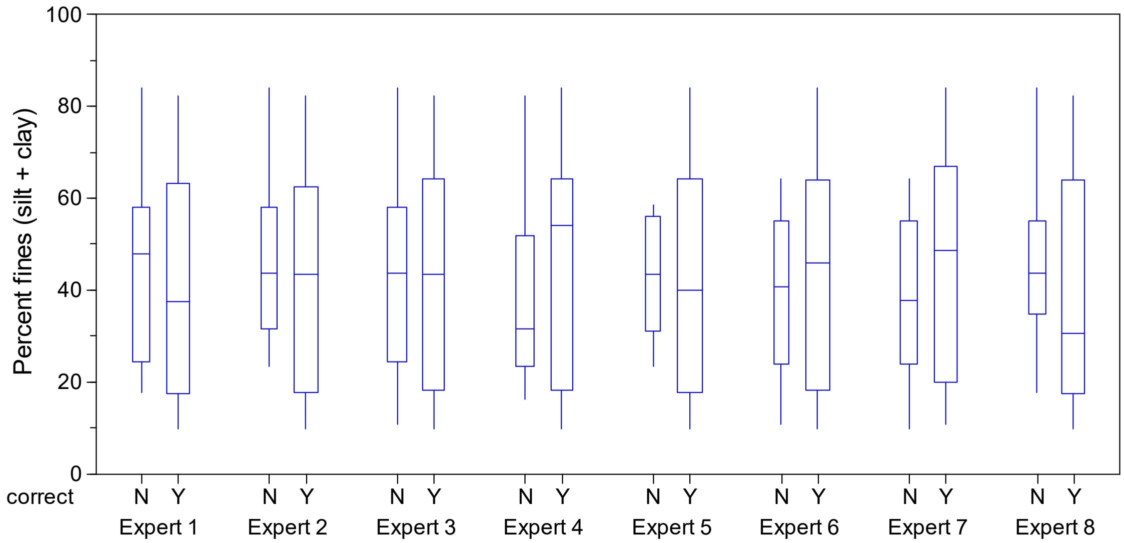

Figure J5. Sediment percent fines (silt + clay) vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

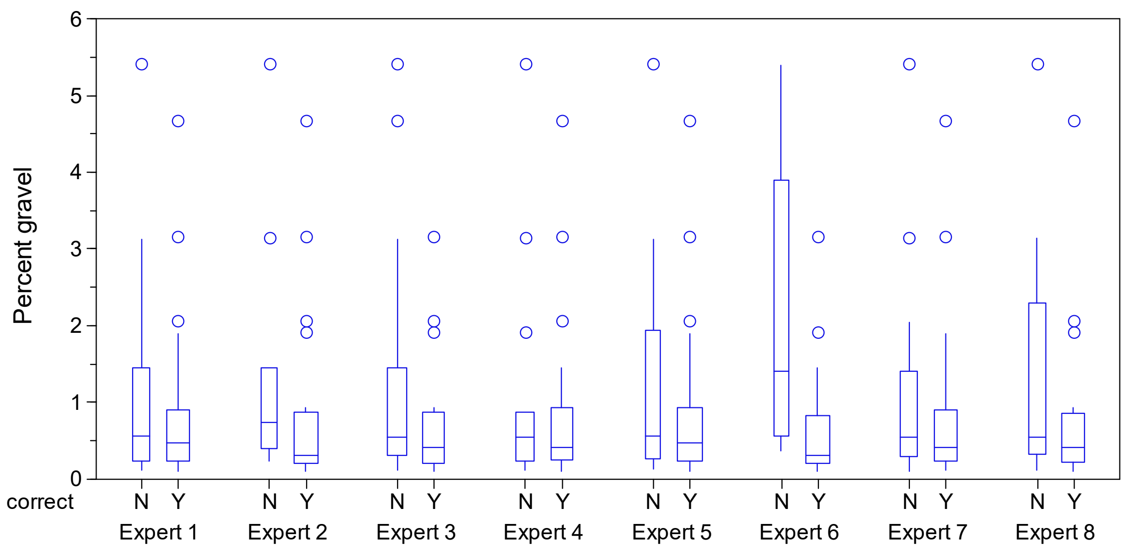

Figure J6. Sediment percent gravel vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).

Table J1. Percent of experts correctly identifying the more-disturbed community with submerged aquatic vegetation (SAV) or shell hash present.

SAV	Number of pairs	% experts correct
present	1	62.5%
absent	29	65.2%

Shell Hash	Number of pairs	% experts correct
present	1	87.5%
absent	29	66.1%

Appendix K: Exercise expert confidence level by response

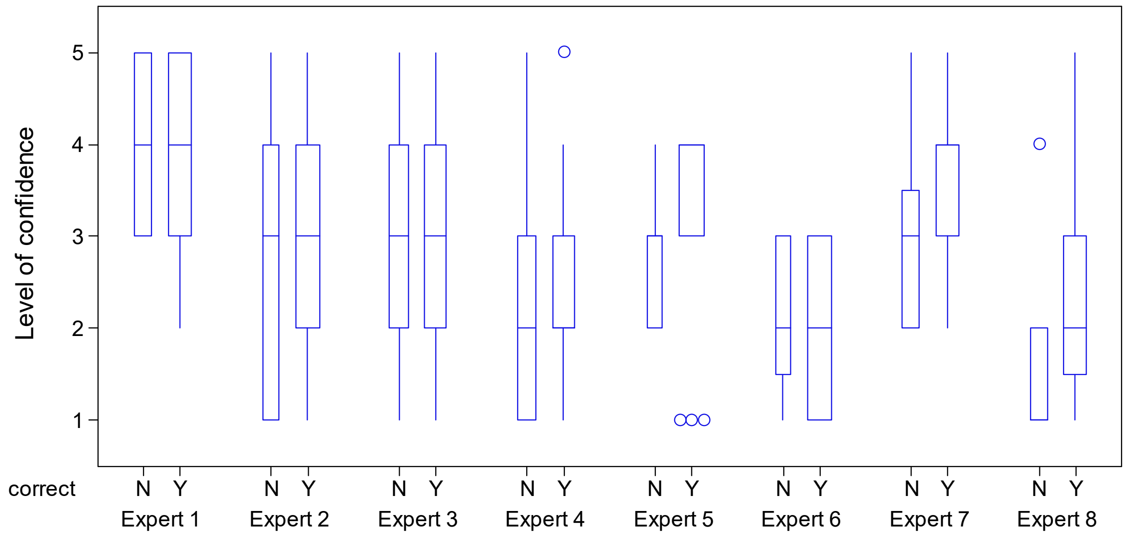

Figure K1. Self-assessed confidence level vs. whether the more-disturbed communities were distinguished correctly (N = no, Y = yes).