State of Oregon

Department of Environmental Quality Memorandum

To: Integrated Report Work Group
Date: January 18, 2018
From: Integrated Report Improvement Team
Subject: Summary of biocriteria peer review and DEQ recommendations

Biocriteria background

The need for technical peer review was established by the legislature through ORS 468.B.039 in 2015 and required DEQ to solicit scientific peer review when developing methodologies for the assessment of state waters. Consequently, as part of its Integrated Report improvement efforts, DEQ convened a technical review panel in the fall of 2017 to solicit independent scientific and technical input regarding the biocriteria impairment thresholds. These thresholds were applied to listings for the 2012 Integrated Report and are proposed for use in the 303(d) assessment of biocriteria for the 2018 Integrated Report.

The scientific peer review panel was tasked with determining whether the existing biocriteria impairment thresholds are valid, and whether the status of non-attainment represents an impairment of the beneficial use. The seven panel members consisted of experts in the aquatic ecology field and included representatives from federal agencies, academia, and professional scientists. Members of the technical advisory panel and their affiliations included:

- Benjamin Jessup, Tetra Tech;
- Camille Flinders, NCASI;
- Dr. Chuck Hawkins, Utah State University;
- Dr. Ian Waite, USGS;
- Dr. Jan Stevenson, Michigan State University;
- Dr. John Van Sickle, consultant; and
- Dr. Michael Paul, Tetra Tech.

The technical review panel is not a decision-making body. Their role is to provide technical review and input on the biological thresholds DEQ uses to assess impairment of the aquatic life use and where appropriate, make recommendations to the Integrated Report Improvement Team on revisions to the Biocriteria Assessment Methodology.

The questions that were posed to the peer review panel were:

1. Are Oregon's biocriteria thresholds valid, and do they adequately represent the cutoff where aquatic life use is considered to be impaired?
    - If they don't adequately represent the aquatic life use attainment cutoff, what are the limitations of the thresholds and how might they be improved?

2. Oregon currently has two thresholds, one for designated use support (i.e., good biological condition, equivalent to reference) and another for designated use impairment (i.e., poor biological condition, dissimilar from reference). This approach of two thresholds creates a third category of potential concern (uncertain biological condition). DEQ has received input from EPA favoring a single threshold approach, resulting in only two categories of beneficial use support (attaining or impaired). Please provide input on which approach is ultimately more technically defensible in your professional opinion.

3. Are Type I and Type II errors sufficiently balanced by the regional biocriteria thresholds?
    - If not, suggest alternatives for balancing Type I and Type II errors.

4. Are there other methods for determining biological thresholds that DEQ should consider?

After comments were compiled from the peer review panelists, DEQ reviewed the comments and identified a majority consensus on three major conclusions from the panel members:

(1) DEQ's biocriteria thresholds are valid and are similar to thresholds used in other states.
(2) Use of two impairment thresholds are more technically defensible than use of a single threshold and may more accurately inform management decisions.
(3) Moving forward, DEQ should seek to relate impairment thresholds to ecological condition.

Validity of biocriteria thresholds

The consensus of the peer review panel was that DEQ's biocriteria thresholds are valid, are derived from standard and acceptable methods, are soundly based on a statistical distribution and are similar to methods employed by other states. One panelist was unable to comment on the validity of the thresholds since they had underlying concerns about PREDATOR model validation. Concern was also expressed about the thresholds established for the Northern Basin Region (NBR) and suggested they should be employed with caution.

It is common practice among states to tie the definition of use support to the concept of reference conditions. EPA guidance documents equate "use support" with the technical definition of "reference conditions" (i.e. the foundation on which the PREDATOR model is predicated). The definition of reference conditions is an integral part of a bioassessment model because it establishes one end of the spectrum of biological condition. The conditions for determining whether or not a stream is considered "impaired" represents the other end of the spectrum. As one of the reviewers pointed out, "As with almost anything, thresholds have limitations for protecting aquatic life use, but that does not mean the thresholds are not adequate."

Number of impairment thresholds

The majority of reviewers (five of seven reviewers) concluded that two impairment thresholds are more scientifically defensible than a single threshold. Biocriteria measurements are multivariate in nature and the assessment method simplifies the ecological complexity in the community into a single metric. The use of two thresholds better reflects a gradient of ecological

condition and is supported by EPA guidance documents[1] (CALM, 2002). Use of a single threshold approach is difficult to justify on statistical grounds given the uncertainty and variability associated with estimating O/E values (or any other index of biological condition). As one reviewer pointed out, "… the technical literature all indicates that most biological responses to stressors in streams are gradual. Therefore, there is no clear technical line of "detriment on this side, not on this side"…This distinction is only a policy one." As several reviewers pointed out, the use of multiple thresholds allows for more of a refined management response. Those sites that fall in the "gray zone" could be targeted for follow-up monitoring and likely are the sites that could be the easiest to reverse impairments through restoration and best management practices.

Linkage of thresholds to ecological condition

The third major point reiterated by panel members was the advantages of linking impairment thresholds with associated ecological function. One of the reviewers suggested that "the use of ecological information embodied by what taxa were typically protected (and or lost) under the proposed thresholds would be valuable in evaluating thresholds." An analysis on what ecological functions are lost or degraded at a loss of 10% of taxa from the reference conditions would help determine whether a detrimental change has or has not occurred. Similarly, it was recommended by multiple reviewers that DEQ consider an alternative approach in which thresholds are set based on considerations of ecological function – e.g., how much taxa loss constitutes unacceptable ecological harm. Revising thresholds based on changes to ecological function should then be subsequently supported by appropriate statistical analyses, and DEQ is proposing to include this task in the next round of methodology improvements.

Area of concern:

The most notable points of concern in DEQ's biocriteria thresholds presented by the review panel were: 1) lack of reference validation data sets to independently assess model accuracy, and 2) lack of estimates of error rates or repeatability. DEQ agrees that these concerns are valid and we are committed to addressing these concerns in future Assessment Methodology updates.

However, there are some important points to consider in why DEQ did not reserve data to validate the model and estimate error rates. First, as it was pointed out by one reviewer, DEQ does have a large enough sample size in the WCCP model to set aside a validation dataset, however, in reality there are an unbalanced number of sites in each ecoregion. The predictive functions built into the models are designed to deal effectively with a spatially unbalanced reference population. Despite this predictive function, a few of the ecoregions have small sample sizes. Reducing the sample size further and by pulling aside even a small number of sites from these ecoregions would potentially reduce representation of these sites enough to reduce the accuracy of predictions in these regions. Given these limitations, it is acceptable practice to

[1] EPA, 2002. Consolidated Assessment and Listing Methodology (CALM). Toward a Compendium of Best Practices, First Edition. United States Environmental Protection Agency. July 2002.

forego model validation. With this consideration in mind, DEQ did not utilize reserve validation datasets tor validation of either model.

Second, DEQ does not have appropriate estimates of the error rates in PREDATOR assessments because adequate repeat replication of data samples from reference sites was not available. The main reason for this sampling deficiency is the drastic reduction in funding for ODEQ's Biomonitoring Program, beginning in the mid-2000's. This has reduced the amount of monitoring in general, but reference site monitoring specifically. With re-allocation of modest funds to Biomonitoring, we have anticipated this data need. In 2015 DEQ instituted a Reference Trend network of 12 sites across the state, spread equally among ecoregions, and sampled annually. We anticipate being able to more effectively characterize the variability in O/E prior to the 2020 Integrated Report.

DEQ Recommendations:

Moving forward, DEQ intends to address the points identified by the peer reviewers in multiple steps. The first step involves minor revisions to the biocriteria assessment methodology for the 2018 Integrated Report, while the second step includes more in depth revisions to the methodology moving forward.

2018 Integrated Report

In an effort to address the comments received from panelists about the inherent variability and repeatability of macroinvertebrate sampling, DEQ is proposing a change in thresholds for Assessment Units with only one sample. Until an assessment of variability in O/E at reference and non-reference populations is completed, DEQ will require multiple samples to determine impairment at the current thresholds. For Assessment Units with one sample, the threshold for biological impairment will be moved from the 10th percentile to the 5th percentile of reference O/E scores. Assessment Units with two or more samples will retain the current biocriteria thresholds (Tables 1 and 2).

DEQ is also proposing the creation of an additional reporting category, Category 3C (formerly Category 3B in 2012 methodology) that would differentiate a minimally disturbed biological condition from those units that are on the cusp of impairment, but lack sufficient data (i.e., a single sample) to confirm the impairment conclusion. Those waterbodies that lack sufficient data to make an impairment conclusion will remain in Category 3B. DEQ is proposing to reallocate current biomonitoring resources to provide follow-up monitoring for those Assessment Units identified as Category 3B.

Table 1. Biocriteria Assessment Benchmarks for One Sample

PREDATOR Model Region	Assessment Category			
	Category 5: Water Quality Limited	**Category 3B: Insufficient Data; Potential Concern**	**Category 3C: Marginal Biological Condition**	**Category 2: Attaining**
Marine Western Coastal Forest	≥ 20% taxa loss[1]	15% to 20% taxa loss	9% to 14% taxa loss	0% to 8% taxa loss
	PREDATOR score ≤ 0.80	PREDATOR score 0.80 to 0.85	PREDATOR score 0.86 to 0.91	PREDATOR score 0.92
Western Cordillera and Columbia Plateau	≥ 27% taxa loss[1]	22% to 27% taxa loss	8% to 21% taxa loss	0% to 7% taxa loss
	PREDATOR score ≤ 0.73	PREDATOR score 0.73 to 0.78	PREDATOR score 0.79 to 0.92	PREDATOR score 0.93
Northern Basin and Range[2]	---	25% to ≥ 50% taxa loss	---	< 25% taxa loss
	---	PREDATOR score ≤ 0.75	---	PREDATOR score > 0.75

[1] Taxa loss rounded to nearest whole number
[2] Applies to both single and multiple samples

Table 2. Biocriteria Assessment Benchmarks for Multiple Samples

PREDATOR Model Region	Assessment Category		
	Category 5: Water Quality Limited	**Category 3C: Marginal Biological Condition**	**Category 2: Attaining**
Marine Western Coastal Forest	≥ 15% taxa loss[1]	9% to 14% taxa loss	0% to 8% taxa loss
	PREDATOR score ≤ 0.85	PREDATOR score 0.86 to 0.91	PREDATOR score 0.92
Western Cordillera and Columbia Plateau	≥ 22% taxa loss[1]	8% to 21% taxa loss	0% to 7% taxa loss
	PREDATOR score ≤ 0.78	PREDATOR score 0.79 to 0.92	PREDATOR score 0.93

[1] Taxa loss rounded to nearest whole number

Long-term updates

The goal of biocriteria assessment is to facilitate ~~a~~ characterization of biological conditions demonstrating support of the aquatic life use. One of the challenges inherent to developing biological thresholds is that measurements are multivariate in nature and require a method for simplifying the complexity in assemblage/community data for ecological interpretation. Defining reference conditions is an integral part of bioassessment because it establishes one end of the spectrum for biological condition while the thresholds that are set define the other end of the spectrum. Performing an ecological evaluation of what is happening at these thresholds (e.g. what taxa or functions have been lost, what taxa or functions might be vulnerable) increases confidence in model predictions. In order to address the linkage of ecological condition with biocriteria thresholds, DEQ is proposing to pursue the following tasks prior to completion of the 2020 IR:

1) Update reference screening protocols east of the Cascades and build an updated east-side model. Once DEQ has updated models covering the entire state, DEQ will:
2) Retire PREDATOR and replace with new updated O/E models.
3) Complete an assessment of variability in O/E at reference and non-reference populations.
4) Model additional biological metrics (e.g. % EPT, functional feeding groups, dominant groups, replacement/missing taxa, etc.) to be used as supporting information for biocriteria condition designations.

Further Clarification on Biocriteria

DEQ is using PREDATOR O/E to assess biocriteria status, where possible. That means DEQ will use PREDATOR O/E as the *primary* means of assessing macroinvertebrate community data. However, as pointed out to us by the Oregon Department of Justice (Larry Knudsen, pers. Comm.), a narrative implementation of a water quality standard does not allow us to require a specific type of analysis or assessment; rather, we must allow for various other forms of data to also be used to assess the narrative standard. With this in mind, it is entirely appropriate for a different bioassessment tool to be used to validate or refute a biocriteria listing. That said, DEQ reserves the right to review the assessment tool for methodological and statistical rigor and may or may not approve of its use.

In addition, this also means that DEQ is authorized to use other methods of evaluation to assess types of assessments of macroinvertebrate community condition. Examples of this would be an upstream/downstream approach to assessing point-source discharges, use of other indexes appropriate for assessing larger rivers (which are not integrated into PREDATOR models), or the use of professional judgement by qualified aquatic ecologists.