State of Oregon Department of Environmental Quality

Response to Peer Review Comment

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 peer review panel was convened in November of 2017. Members of the panel were solicited from a list based on input from a cross-section of stakeholders involved in a stakeholder advisory group for improvement of the Integrated Report methodology. The panelists completed review of the methodology on Dec. 29, 2017. An overview of the comments and the proposed assessment methodology based on the input of the peer review panel was made available for public review and comment in January 2017, pursuant to the requirements of the Oregon Senate Bill 829.

Persons or organizations that submitted written comments are listed in Section 1.2. Each panelist is identified by a number. All remarks, observations or recommendations were extracted from the review form for each panelist. Comment numbers are assigned by relevance to each of the questions in the charge to review panelists. Comments that addressed the same issue were grouped and a common response was given to address the comment. Unique comments were answered individually. A summary of all comments submitted and DEQ's response is presented in Section 2. The original panel member responses are available on the Integrated Report Improvement's advisory committee website.

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.

Section 1.2 List of Panelists

1. Benjamin Jessup, M.S.
   Ecological Analyst

Tetra Tech

2. Camille Flinders. M.S.
   Aquatic Biology Program Manager
   National Council for Air and Stream Improvements

3. Dr. Charles Hawkins, Ph.D.
   Director
   Western Center for Monitoring and Assessment of Freshwater Ecosystems
   Utah State University

4. Dr. Ian Waite, Ph.D.
   Research Biologist
   USGS Oregon Water Science Center

5. Dr. Jan Stevenson, Ph.D.
   Professor, Department of Integrative Biology, Michigan State University
   Co-Director, Center for Water Sciences

6. Dr. John Van Sickle, Ph.D.
   Environmental Statistics Consultant

7. Dr. Michael Paul, Ph.D.
   Senior Scientist
   Tetra Tech

The technical review panel is not a decision-making body. Their role was 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

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

potentially reduce representation of these sites enough to reduce the accuracy of predictions in these regions. Given these limitations, it is acceptable practice to 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.

Section 2 Detailed Summary of Specific Comments and Responses

Section 2.1 Key to Detail Comment Summaries

Comments and responses are compiled in tabular form in Section 2.2.

Column 1: Comment number. Consist of two numbers separated by a period. The first number corresponds to the peer review charge question. The second number is the sequential order in which the comment was recorded.

Column 2: Panelist number. A number identifying the panelist from the list of panelists (Section 1.2).

Column 3: Summary of comment. This column presents a summary of the comment extracted from peer review response. When commenters are grouped, the comment represents the comment was common to each commenter in the group.

Column 4: Response to comment

Column 5: Revision. This column states whether the methodology and/or whitepaper were revised based on the comment.

Question 1

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?

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
1.1	1,3, 5, 6,	DEQ's thresholds are valid because they were derived from standard and acceptable analysis methods, similar to those used by other states and represent impairment relative to reference condition.	Comment noted	None required.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
1.2	4	The various cutoff or breakpoints in the PREDATOR scores seem reasonable for the MWCF = Marine Western Coastal Forest and WC+CP = Western Cordillera + Columbia Plateau.	Comment noted	None required
1.3	7	The state asks whether the 10th percentile represents a departure from reference or expected condition, but Oregon's standard is not written this way – it is written to prevent "detriment". If Oregon was to define detriment as a change from reference, then the threshold could be reviewed as such and I would have to conclude that for the MWCF and WCCP, 15 and 22% loss are likely departures from reference since they are more than 1 standard deviation from the average reference score, but that 50% loss in the NBR is likely too noisy to be useful as a measure of difference from reference.	DEQ does not intend to redefine its narrative standard to reflect the reference condition approach at this time. Although it is not explicitly stated in the narrative criteria, definitions of reference and detrimental change are provided in OAR 340-041-002. Oregon's OAR 340-041-002 (75) defines "Without detrimental changes in the resident biological community" as no loss of ecological integrity when compared to natural conditions at an appropriate reference site or region. "Appropriate reference site or region" is defined in OAR 340-041-002 (5) as a site on the same water body or within the same basin or ecoregion that has similar habitat conditions and represents the water quality and biological community within the areas of concern.	None currently
1.4	1	Site classification was dependent on reference sites available per region – so the specificity might be adequate in areas with low disturbance pressures and might be uncertain in areas with higher general pressure and greater variety of stream settings.	Comment noted. DEQ intends to update its O/E model with additional reference site information for future Integrated Reports.	None required

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
	4	The current cutoffs in the ODEQ report that provide a range for moderately disturbed sites should be targeted for further evaluation, but they are not the same as the sites that are showing full impairment. These are likely to be the sites that would be the easiest to reverse the impairment through restoration and best management practices in the watershed. These may be the first sites that should be selected for such restoration and further evaluation efforts.	DEQ agrees that it is important to differentiate those sites that are likely impaired versus those sites that may have a minimal to moderate level of disturbance and may be partially supporting their aquatic life use.	DEQ is proposing to add an additional reporting category that would differentiate those sites that may be minimally disturbed.
1.5	6	At present, there are no "gold standard" methods of assessing biological impairment of Oregon streams that are currently available, independently of the PREDATOR model and its supporting data. Thus, it does not appear to be possible to determine, with high confidence, whether or not the 2012 ORDEQ thresholds "… adequately represent the cutoff where aquatic life use is considered to be impaired." This reality, along with the inherent vagueness of the terminology of "aquatic life use" and "considered to be impaired", precludes any purely technical challenge to the current thresholds.	DEQ agrees that this is a policy interpretation and requires further investigation into the loss of ecological function.	DEQ intends to pursue the linkage of current thresholds to ecological function for future assessments.
1.5	6	Such thresholds can be valuable as approximate benchmarks for evaluating changes over time and space that are due to resource usage and management activities. Estimates of change, such as % attaining, accurately quantify upward or downward trends in the overall health of Oregon's streams.	DEQ agrees that the status of Oregon's waters would be best explored in its 305(b) report.	None required.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
1.6	3,7	This percentile of reference approach has both strengths and weaknesses. Thresholds are quantitatively based, but they do not appear to be based on ecological considerations.	DEQ agrees that further investigation into the loss of ecological function would inform the validity of the biocriteria thresholds. Given more time and resources, DEQ intends to explore the linkage between O/E scores and ecological function.	DEQ intends to pursue the linkage of current thresholds to ecological function for future assessments.
1.7	2	The process by which OR DEQ came to have three regional models is reasonable (i.e. examining model performance at different regional scales). However, the resulting model frameworks have not been validated using a test dataset.	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 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 forego model validation. With this consideration in mind, DEQ did not reserve validation datasets for validation of either model.	DEQ revised its methodology to require multiple samples for a Category 5 listing unless there is overwhelming evidence to list as impaired.
1.8	2, 4, 7	The threshold for the NBR (Northern Basin and Range) seems low and may not adequately represent impairment. Certainty in model predictions is currently insufficient to make assessment and management decisions with a high degree of confidence.	DEQ agrees that more data in the Northern Basin Range is required to provide confidence in assessments. DEQ plans to target improvements to the PREDATOR model east of the Cascades for future Integrated Reports.	DEQ revised NBR Category 5 threshold to include BPJ for impairment determination.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
1.9	5	DEQ's thresholds are not without limitations, but that does not mean they are not adequate. Refinements to the thresholds would improve their validity.	DEQ recognizes these limitations and intends to gather more data to improve models, test different statistical methods for modeling, and to investigate including multiple biological assemblages for future assessments of biological condition.	None currently.
1.10	2	The datasets used for model development did not measure or account for temporal variation and intra- and inter-annual variation in O/E scores is not known.	DEQ recognizes that temporal variation of O/E scores is an important component of assessment methods. 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 revised its methodology to require multiple samples for a Category 5 listing unless there is overwhelming evidence to list as impaired.

Question 2

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.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
2.1	1, 2, 3, 4, 6	A two threshold approach is reasonable and technically defensible and ultimately recognizes the uncertainties associated with biological assessment.	DEQ agrees	DEQ added a third threshold for single samples with overwhelming evidence of impairment.
2.2	5	There are no technical issues with defensibility of having either one or two thresholds, there are just conceptual issues about what the goal of the assessments are.	DEQ's goals for assessment of biocriteria are to 1) assess the biological integrity of macroinvertebrate communities and determine where criteria are met, and 2) balance Type I and Type II errors to protect resource while not misidentifying water bodies as impaired.	None required.
2.3	7	Since Oregon's narrative standard speaks to only two conditions (detriment or not detriment) than one threshold seems more defensible.	The standard states "without detrimental changes in the resident biological communities". Clear identification of what constitutes "detrimental changes" is important and there is a line beyond which that occurs. It does not, however, preclude setting up multiple classes of support (full or partial) or impairment (degraded or highly degraded).	None currently

Question 3

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.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
3.1	1, 2, 3	Without a formal error analysis , it is difficult to determine whether Type I and Type II errors are balanced.	DEQ agrees that a formal error analysis is important, however, DEQ does not currently have the data necessary to complete such an analysis. DEQ currently lacks the resources to maintain the monitoring program for this analysis.	None currently.
3.2	4	Error rates have been sufficiently balanced in the MWCF and the WC + CP regions, but effort should be made to improve the NBR model.	DEQ agrees that more data in the Northern Basin Range is required to provide confidence in assessments. DEQ plans to target improvements to the PREDATOR model east of the Cascades for future Integrated Reports.	None required.
3.3	5	The thresholds established by DEQ reasonably balance Type I and Type II errors. A single threshold would have a higher Type I error rate.	Comment noted.	None required.
3.4	6	To accurately estimate the actual rate of Type 2 errors, one would need to apply PREDATOR and the assessment thresholds to a collection of sites that are independently known to be impaired. The percentage of such sites declared to be “reference” would then be a good estimate of the Type 2 rate. Such an independent estimate is not available. Thus, it is not possible to know	DEQ agrees that this type of analysis would inform the question of balanced error rates, however, it does not at this time have the resources to undertake such an analysis.	None currently.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
		whether Type 1 and Type 2 rates are being balanced.		
3.5	7	Since DEQ is defining attainment or impairment, then a statistical test is not applicable. With independent measures of attainment (other WQ parameters), DEQ could look at decision agreement among criteria, but type I and II error may not be relevant.	Comment noted. DEQ will look at other WQ parameters for multiple lines of evidence of impairment.	None required.
3.6	5	Alternative approaches may be to 1) set a lower threshold for distinguishing support and impairment; 2) set a boundary around the sites for requiring more information; 3) use repeated sampling for borderline conditions; or 4) use other biological metrics and assemblages to assess biological condition.	For the current IR, DEQ is proposing to apply a boundary for sites requiring more information and multiple samples for those water bodies identified as borderline.	DEQ is requiring more information and multiple samples for samples identified near the cusp of impairment thresholds.
3.7	6	It would be useful to determine which of the two types of error are more costly, both economically and politically. If one error is significantly more costly than the other, then a distinct imbalance in the error rates may be more desirable.	Resource economics is an extremely complicated topic. DEQ does not possess the bandwidth at this time perform a thorough analysis of the economic and political costs of impairment.	None currently

Question 4

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

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
4.1	1, 3, 7	DEQ should consider using a more conceptual framework such as the Biological Condition Gradient.	DEQ is participating in a BCG approach to assessing biological integrity for wadeable streams in the Willamette Valley ecoregion. This ecoregion has extensive human alterations across the landscape, effectively limiting the number of even marginally appropriate reference sites for assessments. The BCG conceptual framework can be used to either validate existing O/E thresholds for the WV, or to offer an additional line of evidence for those sites with O/E near the threshold of impairment.	None currently.
4.2	2	DEQ should consider using receiver operating characteristics. This approach uses a standard set of calculations to derive several quantitative measures of the performance of a classification model involving a threshold that divides measured and predicted data into two groups (one having (or predicted to have) an undesired condition and one without the condition), and provides a means of assessing the nature and extent of agreement between the true or measured condition and the model-predicted condition. I	DEQ intends to examine a variety of alternative methods for identifying appropriate thresholds.	None currently

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
4.3	4	DEQ should use multiple samples or years to evaluate sites that are on the cusp of the thresholds in addition to other biotic assemblages (e.g. algae, invertebrates, fish).	DEQ agrees to include multiple samples for sites on the cusp of impairment threshold. DEQ will explore the possibility of using multiple assemblages in future WQ assessments.	Include multiple sample requirement for sites on the cusp of impairment threshold.
4.4	5	DEQ should consider the use of tiered aquatic life uses where the intermediate O/E range is still acceptable but of lower quality.	DEQ will examine the application of tiered aquatic life uses in other states and evaluate the applicability of their use in Oregon.	None currently
4.5	6	Recommends retaining a statistical approach for setting thresholds. Although the uncertainties with the approach can be substantial, most of them can be quantified.	DEQ agrees that its biocriteria thresholds will be based on statistical properties, but will be validated with additional ecological information.	None required
4.6	6	DEQ could increase robustness of its assessment by adopting an average or consensus from multiple indicators of macroinvertebrate assemblage condition such as an MMI, EPT richness, etc.	DEQ intends to pursue this analysis for future iterations of the Integrated Report assessment methodology.	None currently.

COMMENT	PANELIST	COMMENT SUMMARY	RESPONSE	REVISION
4.7	6	DEQ should consider the BC Index as an alternative to O/E. It measures the compositional dissimilarity between the Observed and Expected assemblages, however it can use reference taxa without losing any discriminatory power.	DEQ will explore the applicability of the BC Index in future WQ assessments.	None currently
4.8	7	DEQ could refocus the narrative standard on reference condition and make the argument for why the 10th% is different from reference.	DEQ does not intend to redefine its narrative standard to reflect the reference condition approach at this time. Although it is not explicitly stated in the narrative criteria, definitions of reference and detrimental change are provided in OAR 340-041-002. Oregon's OAR 340-041-002 (75) defines "Without detrimental changes in the resident biological community" as no loss of ecological integrity when compared to natural conditions at an appropriate reference site or region. "Appropriate reference site or region" is defined in OAR 340-041-002 (5) as a site on the same water body or within the same basin or ecoregion that has similar habitat conditions and represents the water quality and biological community within the areas of concern.	None currently