Independent Multidisciplinary Science Team (IMST)

State of Oregon

Michael J. Harte
Robert M. Hughes
Victor Kaczynski
Nancy Molina
Carl Schreck
Clint Shock
J. Alan Yeakley

c/o
Oregon State University
Department of Forest Ecosystems & Society
321 Richardson Hall
Corvallis OR 97331-5752

January 15, 2010

Neil Mullane
Department of Environmental Quality
811 SW Sixth Avenue
Portland, OR 97204-1390

Dr. Mr. Mullane,

In response to your October 20, 2009 letter, the Independent Multidisciplinary Science Team (IMST) reviewed the Department of Environmental Quality's (DEQ) November 2009 draft document titled *Development and Selection of Bedded Sediment Benchmarks in Oregon*. In your letter, you asked IMST to address two main questions:

1. Is the approach, described in DEQ (2009), a credible and scientifically defensible method to select indicators and values to use to assess and identify streams impaired by excess sediment? Elements of the approach include:
   a. Use of statewide reference site data,
   b. Stream classification,
   c. Sediment indicator selection, and
   d. Process to select benchmark value.
2. Are the results of the technical evaluation of relevant and available data and information, described in DEQ (2009) and contained in Jessup (2009), valid and appropriate as the basis for selection of indicators and benchmark values?

The IMST commends DEQ for its efforts to develop bedded sediment benchmarks for Oregon wadeable streams. IMST recognizes the scientific and technical difficulties associated with the task of developing indicators for a narrative water quality criterion. We recognize the substantial challenge that DEQ has undertaken in this regard. Clearly, the agency is at the forefront of US states regarding benchmarks for bedded sediment. Within the attached review, the IMST provides specific answers to your questions as well as more general comments about DEQ's draft document.

If you have any questions regarding the IMST's review please feel free to contact either of us. If you or your staff would like to schedule time at an upcoming meeting to discuss the review, please contact Kathy Maas-Hebner (Kathleen.MaasHebner@oregonstate.edu or 541-737-6105).

Sincerely,

Carl Schreck
IMST Co-Chair

Nancy Molina
IMST Co-Chair

cc:

Jennifer Wigal, DEQ
Karla Urbanowicz, DEQ
Eugene P. Foster, DEQ
Doug Drake, DEQ
Greg Pettit, DEQ
Sue Knapp, GNRO
IMST

IMST Review of Oregon Department of Environmental Quality's
Development and Selection of Bedded Sediment Benchmarks in Oregon
(*November 2009 draft*)

Released on January 15, 2010

Independent Multidisciplinary Science Team
Oregon Plan for Salmon and Watersheds
http://www.fsl.orst.edu/imst

Members:
Michael Harte
Robert M. Hughes
Vic Kaczynski
Nancy Molina
Carl Schreck
Clint Shock
J. Alan Yeakley

Citation: Independent Multidisciplinary Science Team. 2010. IMST Review of Oregon Department of Environmental Quality's Development and Selection of Bedded Sediment Benchmarks in Oregon (November 2009 draft). Oregon Watershed Enhancement Board, Salem, Oregon.

Review Preparation: This review was prepared by the IMST based on an initial draft by an IMST subcommittee (Bob Hughes, Vic Kaczynski, and Clint Shock with Kathy Maas-Hebner providing technical support). Susie Dunham also provided technical assistance to the subcommittee during the review. Jennifer Wigal (DEQ) and Karla Urbanowicz (DEQ) briefed the Team on the process and data analyses that DEQ used to develop and select bedded sediment benchmarks at the IMST's public meeting on October 1, 2009. Karla Urbanowicz, Doug Drake (DEQ), and Ryan Michie (DEQ) were present at the IMST's January 8, 2010 public meeting to discuss the IMST's draft comments and to provide clarification of DEQ's process. The IMST discussed initial review comments at its January 8, 2010 meeting and unanimously adopted (Kaczynski absent) at its January 15, 2009 meeting.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS i

INTRODUCTION 1

GENERAL COMMENTS 3

- Reference Sites and Natural Variability 3
- Sites With Human Disturbance 4
- Sampling Issues 5
- Classifications 5
- Analytical Assumptions 6
- Indicators 7
- Overall Document 8

SPECIFIC COMMENTS 8

EDITORIAL COMMENTS 14

REFERENCES 15

ACRONYMS AND ABBREVIATIONS

DEQ	Department of Environmental Quality, Oregon
EMAP	Environmental Monitoring and Assessment Program
FSP	fine sediment potential
GIS	Geographic Information System
IBI	index of biological integrity
IMST	Independent Multidisciplinary Science Team
LRBS	log relative bed stability
PREDATOR	Predictive Assessment Tool for Oregon
RBS	relative bed stability
TMDL	Total Maximum Daily Load
USEPA	US Environmental Protection Agency

This page left blank intentionally

INTRODUCTION

The Independent Multidisciplinary Science Team (IMST) reviewed the document titled *Development and Selection of Bedded Sediment Benchmarks in Oregon, November 2009 Draft* at the request of the Oregon Department of Environmental Quality (DEQ; letter from Neil Mullane dated October 20, 2009). In particular, DEQ asked the IMST to review the draft document (DEQ 2009) and a supporting document (Jessup 2009) and to evaluate them with respect to two questions:

1. *Is the approach, described in DEQ (2009), a credible and scientifically defensible method to select indicators and values to use to assess and identify streams impaired by excess sediment? Elements of the approach include:*
   a. *Use of statewide reference site data,*
   b. *Stream classification,*
   c. *Sediment indicator selection, and*
   d. *Process to select benchmark value.*
2. *Are the results of the technical evaluation of relevant and available data and information, described in DEQ (2009) and contained in Jessup (2009), valid and appropriate as the basis for selection of indicators and benchmark values?*

The IMST commends the DEQ for its efforts to develop benchmarks for bedded sediment in Oregon wadeable streams. IMST recognizes the scientific and technical difficulties associated with the task of developing indicators for a narrative water quality criterion. We want to emphasize the inherent difficulty of the challenge that DEQ has undertaken in this regard. The agency is at the cutting edge in terms of considering a benchmark for bedded sediment. Other than Kaufmann et *al.* (2008, 2009), we know of no clear precedents in this regards that can be followed. In addition, other than USEPA (2006), Kaufmann & Hughes (2006), Bryce et *al.* (2008; *in press*) and Jessup (2009), there is little scientific literature directed specifically toward a biologically-based benchmark for this variable. As indicated in DEQ (2009) the narrative water quality standard for bedded sediments is:

> *The formation of appreciable bottom or sludge deposits or the formation of any organic or inorganic deposits deleterious to fish or other aquatic life or injurious to public health, recreation, or industry may not be allowed* (Oregon Administrative Rule 340-041-0007 (12)).

The document prepared by DEQ is a draft and will be used in a pilot application in 2010 reporting of water quality to the US Environmental Protection Agency (USEPA) and in the Total Maximum Daily Load (TMDL) process. The IMST offers the following comments with the purpose of improving the final document by increasing the clarity of information presented and by better explaining the strengths and weaknesses of the proposed bedded sediment indicators and benchmarks. In the review below, the IMST provides brief responses to the two questions. These are followed by IMST's general comments on the documents as a whole. We then provide comments that address specific sections or topics within the document, editorial comments, a formal recommendation to the Oregon Plan Monitoring Team, and finally references cited in our review.

1. **Is the approach, described in DEQ (2009), a credible and scientifically defensible method to select indicators and values to use to assess and identify streams impaired by excess sediment?**

IMST finds that, given the state of the science, DEQ has developed a generally credible and scientifically defensible first step to select indicators for assessing and identifying stream sites that are likely impaired by excess fine sediments. However, DEQ's use of reference conditions, stream classification, and selection of benchmark values need much greater elaboration and defense as further explained in our detailed comments below. Briefly, the process used for selecting reference sites with natural catchment disturbances and in ecoregions like the Willamette Valley that are extensively and intensively disturbed by anthropogenic action needs detailed explanation. The classification of catchment erodibility and lithology would benefit from some clear examples of when the process is easy to apply (i.e., consistent slopes and bedrock type) and complex (i.e., heterogeneous slopes, lithology, soil, and alluvium). IMST also suggests that DEQ provide a clear rationale for classifying continuously varying slopes and lithology (which increases, rather than decreases, variance and interpretability of natural phenomena). The selection of biological and sediment benchmarks appears arbitrary. Because such benchmarks have the potential for becoming legal criteria, IMST suggests that the selection process be much more transparent, repeatable by others outside of DEQ, and understandable to both technical and non-technical audiences.

2. **Are the results of the technical evaluation of relevant and available data and information, described in DEQ (2009) and contained in Jessup (2009), valid and appropriate as the basis for selection of indicators and benchmark values?**

The general approach proposed in DEQ (2009) seems appropriate, but its validity can be called into question because the variability between stream power and the observed values of log relative bed stability (LRBS) is insufficiently explained as are the possible uses of this relationship. Are the figures relating stream power to LRBS actually meant to demonstrate the weak relationship between the two variables, leaving the remaining variability potentially explainable by anthropogenic disturbances or other variables? Do the figures indicate that stream power is largely factored out in the calculations of LRBS? If so, this needs to be clearly stated. If not, DEQ will need to better explain how it will deal with variability between stream power and observed values of LRBS on an operational basis. For example, for a stream power $\log_2$ of 3, LRBS can range across 5 orders of magnitude; likewise, for an LRBS of -1, the stream power $\log_2$ can range from 1 to 5. Although DEQ (2009) does not report any R^2 values with regressions, based on the R values in Table 1 (Appendix 1, page 38 of the report) the polynomial and exponential regressions have low R^2 values (0.16 and 0.165) meaning that only 16.0–16.5% of the observed variability in LRBS is associated with stream power. The apparently arbitrary selections of the 80th percentile for sediment and the 5th percentile for biology substantially affect the resulting benchmarks. IMST believes that such choices need further support and explanation. Also, landscape erodibility is complex and highly variable, incorporating a greater range of erodibility than is captured in the landscape classification proposed in the draft. Likewise, local in-stream sediment deposits are inherently variable because of local variability in land use, stream slope, historical sediment deposits, local soils, beaver activity, and large wood (as indicated by the scatter in the points relating LRBS to stream power). Some discussion of those issues would help the reader understand the observed variability in the indicators.

GENERAL COMMENTS

Reference Sites and Natural Variability

IMST suggests DEQ revisit the way the reference sites were chosen, given the variability in natural erosion across the spectrum of soil and rock types in Oregon. By their nature, reference sites may be quite disturbed in regions with highly erodible soils (e.g., Northern Basin Range) and where intensive land uses are extensive (e.g., Willamette Valley). In Appendix H (Jessup 2009), there are many reference sites with disturbance index scores exceeding 25 yet there are also many sites with disturbance index scores less than 30 that are not classified as reference sites. Providing an explanation for the lack of consistency in the use of disturbance index scores for selecting reference sites would increase the transparency of the process. The scatter-plots indicate quite high *% fines* and *% fines and sand* values and low RBS values for some reference sites. Given the importance of reference sites in determining the proposed sediment criteria, IMST suggests that the reference sites be carefully re-examined and explanations provided in Appendix H indicating why they were included or excluded at some index score (say 25 or 30). Those borderline reference and other sites should be flagged in scatter-plots and possibly considered as candidate outliers, especially if they have considerable weight in curve fitting and box plots. The candidate outliers could then be examined for site and catchment characteristics that may explain their position in scatter-plots or box plots. If explanations for outliers are found (e.g., land use, landslides), the other sites would, ideally, be re-examined for those characteristics. The process of examining candidate outliers can help explain alternative causes of variability in LRBS than stream power alone, which may be one use of the LRBS-stream power plots.

At this stage of sediment criteria development, it is useful to consider and develop alternative hypotheses relating to excess fine sediments. However, the formation and delivery of sediments are complex processes, and even reference sites should be expected to have variable responses; they should be retained as reference unless there are clear ecological reasons for rejecting them and not because they fail to correspond to perceptions or appear to be outliers. An additional possibility for outliers (more or fewer sediments than predicted from stream power) are the historical or legacy effects of past land uses (e.g., past logging and forest re-growth) or stream uses (e.g., splash dams). See Harding et *al.* (1998) and Walter & Merritts (2008) for examples of legacy effects in eastern US streams and McCormick et *al.* (2001) for the effects of reference sites of differing quality on index of biological integrity (IBI) expectations.

DEQ (2009) states that the data from Environmental Monitoring and Assessment Program’s (EMAP) probabilistic sampling studies provide good geographic coverage and are sufficient to capture the range of expected conditions in Oregon. However, the basis for determining good geographic coverage and sufficient range of conditions is not specified. It would help the reader if “good” and “sufficient” range of conditions is specified and supported. Also, there are no reference sites shown for the Columbia Plateau or Snake River Plain ecoregions and very few shown, relative to their geographic extent, for the Eastern Cascades, Northern Basin Range, and Willamette Valley ecoregions. By not using available data aggregation techniques to incorporate data from other sources (see IMST 2009) DEQ may not have represented these areas as well in the analyses as suggested by Jessup (2009) and DEQ (2009). We also suggest that the data from other agencies could be used as an independent test of analytical methods used by Jessup (2009). If other agencies’ data are inappropriate, it would be useful for DEQ to state why. If DEQ can specify what data are needed to run the models, then other agencies could possibly modify their sampling protocols so the data they collect in the future are similar enough to be incorporated into future analyses and regulatory activities.

If reference sites with a high amount of natural disturbance were included (i.e., natural disturbance events on stream reaches such as fires and storms, associated sediment deliveries, and high storm flows), then the text needs to be more specific. If these more disturbed sites were included, were enough of them included so as to be representative of ecoregions or were there so few that they may be considered outliers? It does appear that there is a lack of adequate representation of natural variability across ecosystems when Jessup (2009) stated:

> *Effects of erodible soils, fires, and landslides were not important in structuring the sediment data. The lack of relationships between these variables and the sediment indicators were confirmed using scatter plots. What relationships emerged were weak and contradictory to expectations. Most sites had minimal effects of fires, landslide, or highly erodible soils, though the broadest range of sediment indicator values was seen in those sites with minimal effects. At the higher extremes of effect levels, there were few sites and they were not always predominated by fine or unstable sediments. We regarded fire, landslide, and soil effects as too infrequent or inconsequential to be considered in a statewide analysis.* (unpaginated, Appendix D, Jessup 2009)

The conclusion that the effects of naturally erodible soils, fires, and landslides were not important in structuring the sediment data suggests that the range of variation of naturally-disturbed sites were not adequately represented by the reference sites, and time since natural disturbance may not have been incorporated into the analysis. Although the low frequency of erodible soils could mean that they are not significant overall, erodible soils usually are considered major contributors to local sedimentation because of the scale of their distribution. If naturally disturbed reference sites were incorporated in the analyses, the manner in which this was done needs to be explained. If they were included, perhaps the naturally disturbed reference sites could be flagged in the figures.

Sites With Human Disturbance

It would benefit the transparency of the process used by Jessup (2009) and DEQ (2009) if the implications of using non-reference sites in the regressions of stream power versus LRBS was described and discussed in more detail. In particular, a description of how non-reference sites affect stream class and TMDLs would be helpful. Certainly, human modifications to landscapes can alter the variables (e.g., slope and discharge) being used to quantify stream power. It would also seem useful to use just the reference site data for developing regressions, as well as both reference and non-reference data, and to compare the differences.

Additionally, on page 11, DEQ states: "*Future work is also needed to determine if the use of non-reference sites with varying human disturbance in the regression biases the model and gives a false and inaccurate representation of natural variability which could result in errors that do not identify impairments.*" This seems like a substantial unknown and warrants further discussion by DEQ (2009) including its potential implications for at-risk aquatic ecosystems. This concern also does not appear to be adequately covered under DEQ's planned evaluations or future work (pages 26-27).

Sampling Issues

IMST suggests that DEQ briefly describe its field protocol for sampling stream sediments, and also indicate the within-site variability of its sediment variables. It would be useful here to include a text box or table describing how sediment was sampled and how slope was estimated. Sediment data can be very "messy" if too few measurements are taken. Slope data can be very noisy at low slopes if the proper instruments were not used. Field measurement error is likely one source of the scatter seen in the plots. Reflecting on that issue here (or elsewhere) and how it might be corrected by simple improvement in the field measuring device (e.g., water level for slope, sieving for sediment; Kaufmann et *al.* 2008) could lead to improved relationships between channel slope/power and LRBS/sediment size. This is especially important as slopes near or <1%, where subtle slope differences mean the differences between fine gravel, sand, or fines. A discussion on the repeatability of the sampling results and how that could affect variability between LRBS and stream power would also be useful. See Kaufmann et *al.* (1999), Faustini & Kauffmann (2007), and Roper et *al.* (*in press*) for examples.

The methodology presented on pages 8-9 in DEQ (2009) pertains to wadeable streams at summer low flow conditions. It is not clear to what extent temporal variation (within and between streams) has been factored into the analysis. It may not be applicable to larger streams and other seasonal flows. If this is the case, DEQ should state explicitly which seasonal conditions are more important biologically to salmonids, why it focuses on the summer low flow sampling season, and the applicability of the methods to non-wadeable streams and other seasonal conditions. Is DEQ concerned with sediments that are not problematic in the summer, such as fine gravel, that might fill pools or riffles during subsequent winter high flows?

Streams have various reaches with greater and lesser sloping segments. Explain how is it possible to get an unbiased sample selection to characterize an entire stream or river reach, or entire channel network. Explain the rationale of a 40 channel width site. Would a site length 10 or 100 times the channel wetted width (rather than 40) be equally appropriate? How was the choice of segment with a representative slope assured?

It also does not appear as if the scale of sampling in the EMAP protocol fits the scale at which the benchmarks will be applied. It seems inappropriate for a single site to be used to characterize a whole stream network. It would be helpful if DEQ would describe the monitoring and regulatory implications if they are not at the same scale, and how DEQ might resolve such scale and resolution differences.

Classifications

IMST briefly summarizes the site classification process in this paragraph for readers who have not studied DEQ (2009) and Jessup (2009) in detail. Both DEQ (2009) and Jessup (2009) recognize that natural streams vary in their physical and biological characteristics, and Jessup captured two aspects of natural variability by using four stream classes based on lithology and stream power. Jessup (2009) used the stream class natural sediment values in the development of his protocols and resultant potential benchmarks. The authors of DEQ (2009) followed Jessup (2009). Jessup (2009) was able to associate expected sediment variables (LRBS, percent fine sediments [< 0.06 mm], and percent fine and sand sediments [< 2 mm]) to those four stream classes. Ten potential natural benchmarks resulted when the four stream classes were combined with the three sediment characteristics. The percent sediment values were on a percent area basis (not percent by volume or weight). This was not always stated, but would improve clarity if it was better explained in DEQ (2009) and so noted in figures and tables.

IMST is concerned with DEQ using two erodibility and two stream power classes (for a total of four classes) when each class is continuous and not a discrete phenomenon. Classes in nature are most valid at their centroids; their margins or transitions are typically extensive. Thus, classes create distinctions, even when the distinction between classes is weak; therefore classes may be misleading. See Hawkins et *al.* (2000) for an overview of this issue, Dolph et *al.* (2010) for an example of how classes increase variability in IBI scoring and impairment decisions, and Kauffman & Hughes (2006) for examples of using multiple continuous and class variables. IMST suggests using continuous variables whenever possible, certainly when a variable like stream power begins as a continuous variable.

In Appendix 2 (an Excel spreadsheet; DEQ 2009), describe the rationale and approach for determining the fine sediment potential (FSP) score for each geologic unit. Providing the scientific basis for having three classes versus another number such as four or five would increase clarity. Were professional geologists or soil scientist specialized in erosion processes consulted for the classifications? Was a subset of the sites ground-truthed? How were the three classes converted to two? Also the erodibility issue might be most important in nearstream buffers versus uplands far from the stream. Perhaps some of the noise in the LRBS could be reduced by classifying the nearstream buffer area of the entire channel network as well as the entire catchment. For example, Van Sickle et *al.* (2004) found the nearstream buffer a better predictor than the catchment for Willamette Valley streams. IMST suggests using both catchment and channel network buffer to assess which relates best to in-stream sediment variables.

Analytical Assumptions

The analyses presented in Jessup (2009) and summarized in DEQ (2009) are very complex and can be difficult to follow. The assumptions used in Jessup (2009) and the implications of those assumptions to the selection and implementation of benchmarks and indicators could be much more explicit to increase clarity of the science involved. For example, Jessup (2009) used the assumption that "*optimal sediment conditions were 0% fines or sand and fines*" (pages 17-18, Jessup 2009; reiterated on page 17 of DEQ 2009). This does not appear to be a realistic assumption for streams in Oregon or elsewhere. This assumption has implications for aquatic biota and regulations. Later, Jessup states:

> *Our assumptions were based on precedent examples that assumed the zero fine sediments were optimal. Our results showed that in low power, erodible lithology situations, using these same assumptions resulted in identification of very low benchmarks. In hindsight, we probably should have adjusted our assumptions of optimal conditions to reflect some characteristic of the reference conditions.*
> (page 27, Jessup 2009).

Jessup (2009) further states:

> *...the sediment indicators were not always responsive to the landscape level measures that define the disturbance gradient ...However, the biological indicators were responsive to sediment conditions, regardless of disturbance gradient. The fauna that reside in streams with high power and resistant lithologies, that naturally have coarse and stable substrates, may be more sensitive to small changes in sediment conditions. We did not always observe this, perhaps because the biological metrics were not precise enough to register small changes. However, the biological responses to the sediment indicators may become more important in determining benchmarks when the sediment indicators are less sensitive to the disturbance gradient.* (page 26, Jessup 2009)

It would strengthen the scientific credibility of the document if DEQ were to explain how it dealt with uncertainties in the process of choosing benchmarks. Part of such a discussion could address whether the addition of data from other reference sites in some ecoregions would have changed the assumptions. In addition, the rationale, discussion, and results concerning sediment and biological indicators from Kaufmann et *al.* (2008, 2009) and Bryce et *al.* (2008; *in press*) could lend credibility to the DEQ benchmark decisions.

Indicators

The sediment indicators described in DEQ (2009, pages 10–11) and the components incorporated into them are not obvious. This could be clarified by more explanation regarding what data were integrated, and how and why this was done. The lack of clarity may simply be a poor summary of Jessup (2009) but warrants revision and possibly restructuring the information. An example of possible restructuring is on page 11 (DEQ 2009), where the explanation on the stream power variable would have been more useful earlier in the discussion rather than at the end. The example on page 10 (DEQ 2009) details a complex analysis using two indices. However without a demonstration that both of the indices have any bearing in reality, there is no way for a reader to judge the scientific quality of the analysis. By analogy, these analyses are like using a stock market index to predict an unemployment index. Each variable has issues with how well it represents the whole picture and trying to relate them may not make much sense or being statistically justifiable. It would greatly help with clarity and scientific confidence if the foundation underlying each index, the specific information each index imparts, and the reasoning behind the expectation that the regression analysis is meaningful were to be fully explained. Again, material from Kauffman et *al.* (2008, 2009) might add clarity. The information provided for the lithology indicator (DEQ 2009, page 12) is a much better explanation of an index and its basis in reality than the stream power variable. For scientific rigor, a presentation of, or reference to, actual data and literature to back up the assertions would be very useful.

IMST suggests that DEQ not assume the readers of this document are as familiar with biological and sediment indicators as its authors are. The derivations of each indicator should be described so that an interested reader could repeat the work with a reasonable effort. Those indicator derivations would include sampling methods, analytical methods, indicator interactions, and the rationales for each along with pertinent literature references.

Overall Document

IMST feels that the document would benefit from a careful editing to correct grammatical, spelling, and logical errors. A few of these are listed in the specific and editorial comments below. Inconsistent use of terms also makes the document difficult to read (e.g., reference sites are referred to as undisturbed and least disturbed; Jessup (2009) also refers to minimally disturbed streams). If all three types of reference sites are being used by DEQ, each should be defined and explained and examples and rationales behind their use and application should be provided. See Stoddard et *al.* (2006) for suggested terminology.

All figures and tables need enough information, explanation of terms, definition of acronyms, etc. to be stand alone pieces within the document. This is standard scientific writing and helps to ensure clarity and understandability. Titles and legends should be made much more explicit and holistic if the document is to be more readable by a general audience. As they are now presented, figure titles are implicit, not explicit.

The quantile regression approach could also be better explained by reviewing the original literature and providing examples of how the technique is used in ecological analyses where multiple limiting factors affect a response variable. Additionally, the general basis for the selection of benchmarks is not well-explained in the document. It would greatly improve the document if more discussion was added on the subjective nature of the benchmarks and how protective various benchmarks might be versus their potential for indicating excess sediments for purely natural reasons.

IMST also believes that it is critical for DEQ to fully incorporate the science/technical issues, implementation issues, and the implications of the benchmarks chosen into DEQ's document. The document has an excellent start on the science, but the potential implementation implications of its benchmark choices could use much further examination, explanation, and examples. For example, the report could use more discussion on how it will be used in the TMDL process, and how the unique characteristics of a particular stream/watershed might affect application of that process. The process of how these benchmarks will be applied on a day-to-day basis is missing. It might help alleviate public and stakeholder concerns if much more detail were added in this area.

SPECIFIC COMMENTS

Page 2 – In the second and third paragraphs there are several inadequately explained statements. The analysis may be too complex to be adequately described in a brief executive summary. IMST suggests approaching the executive summary from the perspective of a lay person and using simple terminology. Also, linking this project to future regulatory processes may provide insight for non-scientists.

Page 3 –Quantile regression facilitates detection of relationships at the upper limits of biological condition, such as the 95th percentile, but that percentile is estimated with a large confidence interval. Because the estimated 95th percentile is at the upper limit of biological condition at the reference sites, clarifying what is meant by "biological loss" and "Attaining" used in the draft could help the reader understand what is expected as LRBS or sediment size decrease from reference values. Linking these results with the published peer-reviewed results of Kaufmann et *al.* (2008, 2009) and Bryce et *al.* (2008: *in press*) would lend credibility to the DEQ decisions.

Page 4 – Does DEQ include the egg and pre-emergent stages under "spawning stage"? Juvenile fish may also be affected by excess sediment. The document would be more well-rounded scientifically if the association of juvenile lamprey with sand habitat were to be mentioned. The effects of fines and sand on lamprey could even affect the beneficial use for salmonids (e.g., by providing a buffering from predators). Results from Bryce et *al.* (2008; *in press*) would help here also.

Page 5 – Under "Sediment Issues" DEQ lists the focus of the project on sediments ranging from silt to boulders, but in the next paragraph it states that excess fine sediment is considered a pollutant. It is not clear where the boulders, pebbles, etc. are incorporated in the analyses, or why the coarse sediments are even mentioned

Page 5 – The report indicates that the project did not incorporate toxic chemicals that may be associated with sediments. It would be useful to briefly discuss the environmental risk of not incorporating this aspect with the sediment benchmarks. If DEQ will be addressing toxics associated with sediments in other standards, how will those be integrated with these benchmarks? This seemed to be better explained in the framework produced by USEPA (2006). A reference to that document may be useful here. At a minimum, it could help the lay-reader understand the multidimensional effects of sediment if associations with toxic contaminants were to be mentioned.

Page 8 – Are boatable rivers included in the data and analyses? If not, are the benchmarks and indicators applicable to boatable rivers? Also, how are wadeable streams with unwadeable pools assessed for sediment condition? Be explicit, both in the description of field methods and data analyses.

Page 8 – There is inconsistent use of "undisturbed" and "least disturbed" for reference conditions/sites here.

Page 9 – Most of the predictive modeling IMST is aware of uses continuous variables not class variables. This is also true of Cao's publications, which were cited by DEQ (2009) and Jessup (2009).

Page 9 – Under sediment indicators, would it be possible to check the values of the estimated critical particle diameters for the reference sites? Would one way to check the procedure be to compare the observed mean particle diameter with the estimated critical diameter for the reference sites? We would expect that the average ratio would be about 1. A large deviation from 1 would suggest that revisions are in order.

Page 10 – Is equation 3 that relates stream power to average annual precipitation a valid assumption for all streams? How might this relationship change in reaches below dams where water levels and velocities are artificially controlled? Is this methodology applicable to regulated systems? Also, how is surface runoff accounted for within stream power? More precipitation can be captured by some landscapes than others and this can vary from year to year depending on landscape disturbance (e.g., fire, clear-cut logging followed by recovered natural vegetation, rain on snow events). Might such variables as these be associated with some of the scatter between predicted and observed values in various plots?

Page 10 – For step number 5, what was adjusted and why? Why use the full data set for this step rather than the reference sites only?

Page 11 – One thing that happens with moderate disturbance (e.g., sedimentation) is that species richness increases as tolerant species invade. Invertebrate productivity also increases with modest sedimentation along with the increase in species numbers. Some modest sediment may actually benefit fish productivity and compensate for reduced egg and alevin surviv*al.* This may happen in the case of tolerant or burrowing EPT (Ephemeroptera, Plecotera, Trichoptera) taxa and thereby affect PREDATOR scores. Were the taxa lists examined to omit such effects?

Page 11 – Although DEQ (2009) and Jessup (2009, Appendix A) were completed in 2009, they did not incorporate the fish IBIs of Whittier et *al.* (2007) or Pont et *al.* (2009). Both studies incorporated data from outside and inside western Oregon, and both may be more appropriate than Hughes et *al.* (2004) for Willamette Valley, and central and eastern Oregon streams. Hughes et *al.* was based on Coast Range streams. Also, Pont et *al.* used a predictive modeling approach, similar to PREDATOR, based on potential stream discharge, mean air temperature, and channel slope so it was found to be less variable than Whittier et *al.* which only separated xeric from mountain sites.

Page 12 – A text box or table thoroughly describing the biological indicators and the metrics they include would be very helpful to the reader. Another option would be to place complete descriptions of the sampling methods and indicator development processes for both biota and physical habitat in appendices.

Page 12 – A discussion of the degree to which sediment characteristics (e.g., density, shape) are captured in the models is missing. If some characteristics are not included, a brief discussion of why and whether they are important enough to be addressed in future research would be appropriate here.

Pages 12-13 – The GIS data show substantial areas where "bedrock" is clay, silt, or sand. Would areas dominated by clay not have more risk of erosion than areas with rock made of sandstone or silt stone? Would areas dominated by silt and/or sand not be more erodible than clay? Why would these parts of the landscape not function as different erodibility classes? Is it plausible to aggregate areas with deep clay, silt, or sands with areas of consolidated sandstone and siltstone in terms of their erodibility?

Although erodibility is a non-linear relationship between slope and surface materials, should erodibility be based not only on underlying lithology, but on soils, vegetation cover, and topography as well? For example, basalt and andesite rock are rated as being highly resistant to erosion in DEQ (2009) and Jessup (2009). Yet basalt and andesite rock substrate in various parts of eastern Oregon have sparse vegetation and their overlying soils may be classified as erodible entisols or inceptisols. The basalt and andesite tend to degrade to silt and major climatic events tend to move that silt off the landscape. The nature of these shallow soils is that they have not developed over many thousands of years. Could the reasons for this poor soil development be related to the climate, the rocks, sparse vegetation, and the tendency of large natural events to move the developing soil particles off the landscape? Would these processes deliver large volumes of silt to the bedded sediment of the streams associated with them? Does the aggregation of fundamentally dissimilar lithology into highly simplified erodibility categories risk aggregation of highly dissimilar ecological sites with highly dissimilar performance? Does this process create expectations of low amounts of fine material in bedded sediments under circumstances where low amounts of fine material and bedded sediments are infeasible? Such issues as these deserve more discussion, explanation, and perhaps examples from descriptive data analyses

Page 13 – A figure showing the 25% breakpoint would help illustrate the discussion in the first paragraph. How does the number of categories in the ranking system affect the breakpoint?

Page 13 – Under stream classification, catchment area is a variable in the stream power equation so the two are probably correlated.

Page 14 – "The data sets used to develop the stream classification system were robust and valid across the state." How were the data sets determined to be robust and how were they validated? Which variables were considered and why were some not used?

Page 15 – "reference site data represent the best or least disturbed sites" which is it? Best or least disturbed or both? And how valid is the assumption that the reference site data are representative? The reference sites do not capture all natural variability. It would help the report if this were discussed and the presumably minor effect of not doing so explained.

Page 15 – Explanation of the rationale for using the 75th percentile of the reference site data versus the 50th or 25th percentiles would strengthen the scientific credibility of the document. The 75th percentile means that 75% of reference sites fail to meet an expectation.

Pages 16- 17 – This entire discussion is unclear. The selection of an appropriate quantile regression line appears arbitrary in the DEQ methodology and the selection is important in determining potential LRBS benchmark values. Cade et *al.* (1999) appeared to indicate that maximal quantile regressions (90%, 95%) would be most helpful in illuminating potential relationships. Is the analysis technique appropriate for the data as the variances do not look unequal? Jessup (2009) used 90% and 95% regression lines. Why did DEQ not follow suit? DEQ (Figure 8) showed multiple quantile regression lines and stated that the 80th percentile line is the highest parallel line. Parallel to what? Parallel to each other? Inspection of Figure 8 appears to indicate that there is a higher parallel line (90%). Was 80% used to obtain a more conservative benchmark value? If 90 or 95 had been used, the benchmark may have been more negative and less conservative. Also, the continuous/discrete distinction is not well explained, and it makes Table 4 (Table 16) nearly incomprehensible.

Pages 18-19 –Natural conditions vary in space and time across the landscape. Therefore, there would probably be few "optimal" sets of conditions but many tolerable sets of conditions. Is this process based on reaching ideal optimal biological conditions? If benchmarks are to be based on naturally occurring conditions, explain why it is reasonable to assume that any part of the landscape would necessarily have optimal biological conditions for any assemblage? Also explain why near-optimal assemblages or conditions for those assemblages should be expected, and support the assertions with references (e.g., Bryce et *al.* 2008; *in press*). Again, do not assume the reader understands the assumptions and expectations underlying IBIs or expected species richness. Also, what happened to Categories 1 and 4?

Page 19 – Elaboration on the choice of percentiles would greatly benefit the report. For example, how did DEQ determine that 5% biological loss is acceptable and 20% is too severe? It would also be useful to explain why DEQ is not using standard deviations from reference site values. Is this a function of small sample size? Also the 50th percentile would be less subject than the 75th percentile to edge effects with small sample sizes. It would also be beneficial to explain why quantile regression yields a more protective criterion than reference sites sometimes but not other times. Which specific sediment indicators increase with disturbance?

Page 21 – With reference to potential benchmarks presented in Table 5, it is difficult to obtain the LRBS benchmark values from the information presented here and in the document. As presented the materials in the report do not fully support the benchmark values presented.

Page 22 – Box plots showing the reference site quality distributions for each class would be useful. It is stated here that three biological indicators were used to develop criteria, but five are listed in Table 1. Which ones were used and why?

Pages 24-26 –Some discussion of the context in which the benchmarks could be used to DEQ could be useful to the reader. The IMST also believes it could be beneficial to outline the steps involved in the pilot application to determine if the proposed sediment benchmarks are appropriate for further use or need to be revised.

Page 25 – DEQ indicates that "the pilot application will use one site reach sample to assign a status for the assessment." However, Fausch et *al.* (2002) and Smith & Jones (2005) suggested continuous sampling or 15–119 random samples per catchment, respectively. IMST suggests that DEQ provide thorough justification of why one sample can be assumed to accurately represent an entire segment (or channel network?). This may be common practice with regulatory agencies given their limited budgets relative to their missions, but it is contrary to recommended stream ecology sampling strategies.

Figure 3, page 30 – What are the R^2 and the P-values? What are the confidence and prediction intervals? Why bother to log transform data when using a non-linear model to fit the data? If the issue is unequal variance why not use quantile regression here as well?

Stream power was related to LRBS via polynomial (curved) regression relationships. Figure 3 shows the log/log plot. The general relationship (more stable stream beds with higher stream power) is logical as stronger flows have larger bed sediments that are harder to pluck and move downstream. But there is high variability in the log/log plot. Are the figures relating stream power to LRBS actually meant to demonstrate the weak relationship between the two variables, leaving the remaining variability potentially explainable by anthropogenic disturbances or other variables? Do the figures indicate that stream power is largely factored out of the calculations as described by Kaufmann et *al.* (2008, 2009)? If so, this needs to be clearly stated. Such high variability means that predicted LRBS values have wide confidence intervals and low R^2 values. DEQ's (2009) Appendix 1 provides more information. In Figure 1 (page 37, Appendix 1) the relationship is weak and needs to be better explained. Figure 2 (page 38, Appendix 1) shows a transformed log/log relationship and so does Figure 3 (page 39; 2 polynomial regression relationships). The log/log plots appear as slightly drifting shotgun patterns but a polynomial and an exponential relationship were extracted. Although the R^2 values are not reported with Figure 3, based on R values in Table 1 (Appendix 1, page 38), the R^2 values are 0.16 and 0.165; this means that the polynomial and exponential "models" can account for about 16% of the variability in the log stream power/log RBS relationship. Also 84% of the variance could not be explained at a

95% confidence level. This means that confidence in the models is low. How do other factors (e.g., large wood, upstream conditions, land use, lithology) substantially affect this relationship? Again, examples of how the relationships (or lack of relationships) may be used by DEQ to assess concerns with excess fine sediments or fines and sand would clarify the issues and reduce possible misunderstandings.

Figure 7, Page 33– There are no parallel slopes in Figure 7; there are in Figure 8 (recognizing that the figures are from Cade et *al.* 1999). Only a slight adjustment of the quantile regression line would produce parallel slopes for all lines. Were the lines drawn by eye or by binning? Why not use the 90th percentiles as in Figure 9? Figure 7 shows multiple quantile regression lines but it is not obvious how Figure 7 came from Figure 6. Are Figures 6 and 7 necessary or do they just confuse people? More explanation of how these figures are applicable to DEQ's task is warranted.

Figure 8, page 34 – What are the confidence and prediction intervals? The variances do not appear to be unequal. Were unequal variances tested for (Cade et *al.* 1999)?

Figure 9, Page 35 – Figure 9 plots multiple quantile regression lines and the vertebrate IBI reference distribution. The vertebrate IBI distribution is revealing: the mean is about 66, the 95% confidence interval is about 55 to 75, and the 90% confidence interval is about 38 to 96. Explain how this distribution and variability affect the analysis that follows (as presented in Figure 10). Explain why the 75th percentile reference vertebrate IBI was used in Figure 9. Is it just a coincidence that it is the upper 95% confidence level? Explain why the mean vertebrate IBI (about 66) was not used. If the mean was used, the corresponding residual LRBS would be about -1.5. This is a less conservative potential benchmark. The selection of the vertebrate IBI value (and other biological indicators) significantly affects resultant potential benchmarks. It would be wise for DEQ to support those choices with rationale, thorough discussion, and examples, and to support those statements with references (e.g., Bryce et *al.* 2008; *in press*) so that they appear less arbitrary.

Figure 10, Page 36 – Figure 10 shows the derivation of a potential LRBS benchmark (which appears to be about -0.25) for an erodible stream class. An 80th quantile regression was used. Why? The use of 80% results in a more conservative LRBS potential benchmark. If 90% was used, the potential benchmark would have been more negative, less conservative. And if 95% had been used, it would have been even more negative. The selection of the quantile regression line used in the analysis significantly affects benchmark outcomes. Next, what does a 5% reduction mean? It appears to relate to a 5% reduction in the vertebrate IBI but what is presented is less than a 5% reduction. Mathematically a 5% reduction in the vertebrate IBI score would be a 71.25 value. The corresponding ResLBRSpow2 intercept value (and potential benchmark) would be about -0.6. The rational behind the choices made by DEQ needs further discussion and supported with references (e.g., Bryce et *al.* 2008; *in press*).

EDITORIAL COMMENTS

Throughout document –

- Numbers in various paragraphs and tables lack units. Could these be more clearly presented?
- We suggest that R^2 values be added to the regression equations within the document to provide a measure of how much variability is being explained by the equation?

Page 2, paragraph 2, line 2 “recommendations”

Page 4, Line 2 – “Oregon’s”

Page 4 – Why is dissolved oxygen considered a pollutant? Is there a better pollutant example to use here?

Page 4, last paragraph – Misspelling in “One focuses of this...”

Page 5, paragraph 5, line 2 – Suggest omitting “in headwater and other areas”

Page 7, line 7 – Problem with tense – use “with” not “to be”

Page 12 – Provide a citation for the IBI used here.

Page 13 – Explain what “physical habitat and other information” means and which list of factors were examined and which ones were significant.

Page 13 – Misspelling in “...related to a set of variable(s) that included”

Page 14, Table 3 – The numbers of reference sites are not equal in the two major classification categories. Although it is stated here that a reference score needed to be <25, Appendix H lists many with scores >25. Table 3 and Appendix H do not appear to be in agreement.

Page 25 – “catchment area and watershed” is a triple redundancy in this sentence. Also, “one site reach sample” seems to be a contradiction of stream classification. Definitions and consistent usage of site and reach would reduce possible confusion.

Page 25, paragraph 3, Line 6 – Omit “T”.

Page 26 – “Peer” review not “Per” review.

Page 37 – No reference is provided for Cao (2008).

Page 21, Table 5 – Indicate what the averages are based on. Ideally all tables should have adequate information to stand alone from the text.

Page 26, Table 7 – The table is difficult to comprehend. Consider revising.

Appendix 1 – Consider giving figures in appendix unique reference numbers to avoid confusion with figure numbers in the body of the report.

REFERENCES

Bryce SA, Lomnicky GA, Kaufmann PR, McAllister LS, Ernst TL (2008) Development of biologically based sediment criteria in mountain streams of the western United States. *North American Journal of Fisheries Management* 28(6): 1714–1724.

Bryce SA, Lomnicky GA, Kauffman PR (*in press*) Protecting sediment-sensitive aquatic species in mountain streams through the application of biologically-based streambed sediment criteria. *Journal of the North American Benthological Society*.

Cade BS, Terrell JW, Schroeder RL (1999) Estimating effects of limiting factors with regression quantiles. *Ecology* 80(1): 311–323.

DEQ (Department of Environmental Quality) (2009) *Development and Selection of Bedded Sediment Benchmarks in Oregon (November 2009 Draft)*. Oregon Department of Environmental Quality, Water Quality Division. Portland, Oregon.

Dolph CL, Sheshukov AY, Chizinski CJ, Vondracek B, Wilson B (2010) The index of biological integrity bootstrap: Can random sampling error affect stream impairment decisions? *Ecological Indicators* 10(2): 527–537.

Fausch KD, Torgersen CE, Baxter CV, Li HW (2002) Landscapes to riverscapes: Bridging the gap between research and conservation of stream fishes. *BioScience* 52(6): 483–498.

Faustini JM, Kaufmann PR (2007) Adequacy of visually classified particle count statistics from regional stream habitat surveys. *Journal of the North American Water Resources Association* 43(5): 1293–1315.

Harding JS, Benfield EF, Bolstad PV, Helfman GS, Jones EBD III (1998) Stream biodiversity: The ghost of land use past. *Proceeding of the National Academy of Science of the United States of America* 95(25): 14843–14847.

Hawkins CP, Norris RH, Gerristen J, Hughes RM, Jackson SK, Johnson RK, Stevenson RJ (2000) Evaluation of the use of landscape classification for the prediction of freshwater biota: Synthesis and recommendations. *Journal of American Benthological Society* 19(3): 541–556.

Hughes RM, Howlin S, Kaufmann PR (2004) A biointegrity index (IBI) for coldwater streams of western Oregon and Washington. *Transactions of the American Fisheries Society* 133(6): 1497–1515.

IMST (Independent Multidisciplinary Science Team) (2009) *Issues in the Aggregation of Data to Assess Environmental Conditions*. Technical Report 2009-1. Oregon Watershed Enhancement Board. Salem, OR.

Jessup B (2009) *Development of Bedded Sediment Benchmarks for Oregon Streams, Final Report*. Prepared by Tetra Tech, Inc. for Oregon Department of Environmental Quality and US EPA Region 10, EPA contract number EPA-C-08-004. Tetra Tech, Inc. Montpelier, Vermont.

Kaufmann PR, Hughes RM (2006) Geomorphic and anthropogenic influences on fish and amphibians in Pacific Northwest coastal streams. Pages 429–455 in RM Hughes, L Wang, PW Seelbach (eds.). *Landscape Influences on Stream Habitat and Biological Assemblages*. Symposium 48, American Fisheries Society, Bethesda, Maryland.

Kaufmann PR, Levine P, Robison EG, Seeliger C, Peck DV (1999) *Quantifying physical habitat in wadeable streams*. EPA/620/R-99/003. US Environmental Protection Agency. Washington, D.C.

Kaufmann PR, Faustini JM, Larsen DP, Shirazi MA (2008) A roughness-corrected index of relative bed stability for regional stream surveys. *Geomorphology* 99(1–4): 150–170.

Kaufmann PR, Larsen DP, Faustini JM (2009) Bed stability and sedimentation associated with human disturbances in Pacific Northwest streams. *Journal of the American Water Resources Association* 45(2): 434–459.

McCormick FH, Hughes RM, Kaufmann PR, Herlihy AT, Peck DV (2001) Development of an index of biotic integrity for the Mid-Atlantic Highlands Region. *Transactions of the American Fisheries Society* 130(5): 857–877.

Pont D, Hughes RM, Whittier TR, Schmutz S (2009) A predictive index of biotic integrity model for aquatic-vertebrate assemblages of western U.S. streams. *Transactions of the American Fisheries Society* 136(2): 292–305.

Roper BB, Buffington JM, Bennett, Lanigan SH, Archer E, Downie ST, Faustini J, Hillman TW, Hubler S, Jones K, Jordan C, Kaufmann PR, Merritt G, Moyer C, Pleus A (*in press*) A comparison of the performance and compatibility of protocols used by seven monitoring groups to measure stream habitat in the Pacific Northwest. *North American Journal of Fisheries Management*.

Smith KL, Jones ML (2005) Watershed-level sampling effort requirements for determining riverine fish species composition. *Canadian Journal of Fisheries and Aquatic Sciences* 62(7): 1580–1588.

Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH (2006) Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecological Applications* 16(4): 1267–1276.

USEPA (US Environmental Protection Agency) (2006) *Framework for Developing Suspended and Bedded Sediments (SABS) Water Quality Criteria*. EPA-822-R-06-001. Office of Water and Office of Research and Development, US Environmental Protection Agency. Washington, DC.

Van Sickle J, Baker J, Herlihy A, Bayley P, Gregory S, Haggerty P, Ashkenas L, Li J (2004) Projecting the biological condition of streams under alternative scenarios of human land use. *Ecological Applications* 14(2): 368–380.

Walter RC, Merritts DJ (2008) Natural streams and the legacy of water-powered mills. *Science* 319(5861): 299–304.

Whittier TR, Hughes RM, Stoddard JL, Lomnicky GA, Peck DV, Herlihy AT (2007) A structured approach for developing indices of biotic integrity: Three examples from streams and rivers in the western USA. *Transactions of the American Fisheries Society* 136(3): 718–735.