Peer Review Comments – Steven Bay, Southern California Coastal Water Research Project

Peer Review Questions

Methodology for Derivation of Pesticide Sediment Quality Criteria for the Protection of Aquatic Life. Phase II. Method Development and Derivation of Bifenthrin Sediment Quality Criteria

1. Is the way the method addresses bioavailability in accordance with the current state of research on this topic?

Yes, the consideration of relationships among sediment particles, organic carbon, and dissolved phase reflects current knowledge. However, the method considers all organic carbon to be equivalent in binding capacity, which is a simplification that does not reflect current understanding that black carbon and other forms have different partition coefficients. A recent USEPA report provides an update to approaches for carbon normalization, which should at least be cited in the document (EPA/600/R-02/012).

I have concerns regarding the specifics of the approach, which expresses all of the effects data in terms of freely dissolved contaminant. Conceptually, this approach is valid but there is high risk of developing inaccurate criteria due to current limitation in our understanding of the relationships among the binding phases. This process requires normalization and application of an equilibrium model to arrive at C_d: first, normalization to sediment OC, and then, calculation of the freely dissolved concentration. Each step introduces an unknown amount of error and K_{oc} values are stated to be uncertain in many cases. The value of converting all data to C_d is not clear, especially given the potentially high potential for error introduced by the use of inaccurate K_{oc} values. It seems more prudent to express the criteria on an OC normalized basis and thus avoid potential errors introduced by applying K_{oc}. Expression of the criteria on a sediment concentration basis is also consistent with other criteria/guidelines developed in the US and Canada and compatible with all current monitoring programs; it will also allow easier understanding and application of the new criteria. If calculation of C_d is essential to the accuracy or reliability of this method, and it outweighs the impact of additional error, then a justification for this approach should be included.

However, the recommended approach is feasible to apply, considering the uncertainties and data gaps that currently exist.

2. Are all of the ways of accounting for bioavailability included in the method (and listed below) scientifically valid? Are there additional technically valid ways to account for bioavailability that could be used?

Yes, the methods are all scientifically valid. The only other (and perhaps best) method to assess bioavailability would be to measure accumulation in benthic organisms. However, the method is not practical for this application and guidance on how to use the information

is still developing. There has been a recent SETAC-sponsored workshop on the use of passive sampling, which provides a synthesis of the state of the science and provides information on management applications. The information is presented in a series of papers in press in the journal IEAM and should be incorporated into the present document (see attached papers by Parkerton and Maruya 2013, Greenberg et. al. 2013, and Lydy et al. 2013).

The Tenax method measures the bioaccesible concentration, but not necessarily the freely dissolved concentration. This method does not seem to be appropriate for the UCDSM.

a. OC-normalized sediment concentrations
b. DOC-normalized porewater concentrations
c. Directly measured freely dissolve porewater concentrations (via SPME or Tenax).

3. Will environmental regulators and researchers be able to use existing toxicity and monitoring data included in the method to check compliance or does the method require that new techniques be used to generate new data?

Existing monitoring data can be used for compliance checking; assuming that sediment TOC was measured. The method does not describe a role for toxicity data in checking compliance. Toxicity data is only used to determine the criteria.

4. Is it clear how to evaluate studies by reading section 2.3 and appendix A (rating guides) and looking at tables 7-13?

No, this section is not complete and guidance is not clear in some cases. For example, it is emphasized that physiochemical data should be systematically evaluated for reliability, but this document does not include the methods/criteria needed to make such evaluations (another document is cited). Complete information to accomplish this task should be included in the document.

Methods for evaluation of the ecotoxicity data are clear and sufficient information is provided to assign scores.

5. Do the categories and point values assigned in tables 8-12 reflect the importance of the parameters to performing valid sediment toxicity testing?

Table 8: "Relevance" of the test is not clearly defined, so it is difficult to determine if the categories/point values are appropriate. Several of the categories are redundant with other tables (e.g., controls described and meet requirements), creating uncertainty about the need for this table. Point values are reasonable, but the criteria for classifying as relevant/less relevant/not relevant are more stringent than for the reliability evaluation; some justification for the difference in criteria classification ranges should be provided.

Tables 9-12 have reasonable categories. However, the documentation and acceptability tables are highly redundant. I don't see the purpose of having both; if the test conditions are not documented, then low scores will result for acceptability due to lack of data. The point values seem arbitrary and out of balance for some categories, probably a result of having so

many items to evaluate. For example, use of an acceptable method and proper duration are important overarching parameters, yet they are given point values similar to minor test conditions (e.g., photoperiod). If a standard method was used, then it seems that, by definition, many of the other parameters will be acceptable (e.g., replication, feeding, water quality etc.); it is redundant to evaluate such parameters twice in the same table.

6. Is it clear how to prioritize and organize data by reading sections 2.4 and 2.5? Do the data prioritization and exclusion in the bifenthrin criteria derivation seem reasonable (section 8.7)? This step plays a large role in determining which data are used to derive the criteria, and thus the magnitude of the criteria.

The prioritization criteria are reasonable for this purpose. However, the process is poorly organized in section 2.4. This section includes many instructions that have nothing to do with data prioritization, but rather provide guidance on endpoint calculation (e.g., items a-j). These items should be placed in a separate section so that Section 2.4 can focus on actual prioritization rules and thus be more effective.

7. Is it clear what information should be input in the toxicity data summary Table 14?

The information is generally clear, but sometimes excessive and of marginal value. Why does the number of reps and individuals/rep need to be documented for each concentration tested? Assuming this information is in the publication, there is no need to record the raw data a second time.

8. Are instructions in sections 3.4-3.7, describing how criteria are derived, clear and easy to follow?

Absolutely not. If this document is meant to provide easy to follow guidance and encourage use of the method, then major revisions are needed. I found the entire document to be difficult to read and even more difficult to use for calculating criteria. The major issues are: 1)the method is never described and illustrated in a concise and stepwise fashion (Fig. 2 provides a flow chart, but it is never described clearly), 2) one has to read through most of the report and a lot of technical detail before the method is even presented (reader will likely not know what is important to remember in the text), 3) examples are not given to illustrate most complex parts of the method (e.g., SSD distribution calculation, distribution fit evaluation, 4) incomplete instructions are provided for some calculations (e.g., ACR), and 5) the description for some aspects often includes a lengthy discussion of background or the results of the method development work, often without explicitly stating what is to be done (e.g., assessment factor use); this makes it difficult to understand the method as a whole.

I recommend including a section near the front of the document that describes the method in a concise step-wise manner, with details and rationale for selection referenced to later sections. If no one is likely to use the SSD procedures for the foreseeable future, why should the reader be forced to try to understand it in detail?

9. Does it make sense to derive two criteria for a given pesticide, one with a 10-d averaging period and one with a 28-d averaging period (section 3.8.2)? Should only one criterion be derived? Please comment on the thoroughness, validity, and

completeness of the review and discussion in section 3.8.2. Are there are any other considerations that should be included for determining criteria averaging periods?

The review and discussion in section 3.8.2 is thorough and provides adequate justification to support selection of the averaging periods. Derivation of both acute and chronic criteria are consistent with the approach used for water quality criteria and are appropriate for the sake of harmonization. However, so little data are available for chronic effects resulting from sediment exposure that there no way to demonstrate that the approach for developing the chronic criteria using default values from water exposures of different chemicals is accurate. My recommendation is not to develop chronic criteria until there are sufficient chronic sediment exposure data available to establish ACR values without relying on water exposures.

The frequency of exceedance recommendation of no more than once every three years seems to be overly conservative, given the dearth of data available. Most benthic species are relatively short-lived and have reproduce annually or several times per year. An annual frequency of exceedance seems to be more reasonable. Alternatively, given the reliance of this criteria method on water exposure date, the frequency of exceedance should match that used for water column monitoring.

10. Is the assumption of concentration addition reasonable for mixtures of pesticides in the same class (section 4.2)?

The concentration addition approach described is reasonable and described well.

11. Do you know of QSARs that could be used to estimate toxicity to other species, including threatened/endangered species?

No.

12. Are the bifenthrin criteria generated in section 8 protective of aquatic life, more specifically, are they neither unreasonably overprotective nor underprotective?

The bifenthrin criteria appear to be substantially overprotective of aquatic life, based on the comparison of the calculated criteria with water quality criteria and mesocosm studies. Considering the high reliance of the UCDSM on water quality criteria methods, one would expect the BSQC to be similar to WQG. Since the resulting values are more than an order of magnitude below WQC or EPA ESBs, it calls into question the reliability of method to generate reasonable values. The bifenthrin BSQC values are likely to protect aquatic life, but they may be too low to be useful for evaluating causal relationships or guiding management decisions.

Other Comments

This document provides a thorough presentation of the recommended SQC framework and the authors should be commended for providing such an in depth presentation of many considerations needed to develop sediment criteria for regulatory purposes. Justification and references for many of the recommendations are included, which makes the document a valuable resource. However, this document has weaknesses in scope, focus and structure that diminish its effectiveness and clarity. Some of these issues have been mentioned in the responses to the peer review questions, but others do not correspond to the previous questions. These items are summarized here for consideration by the authors.

1. **Disconnect between Phase I and II reports.** The Phase II report states that it describes the SQC approach recommended in Phase I, yet no such recommendation appears to have been made in Phase I (based on the report conclusions). In fact the original goal of Phase II seems to have been to evaluate several methods and arrive at a recommended method. Neither report contains a clear description of how the approach recommended for the UCDSM was selected, including a justification as to why it is superior to the other approaches based on equilibrium partitioning. Specifically, a discussion as to why the SSTT methodology as recommended is superior to the USEPA ESB approach is needed. Both methods use very similar approaches and the advantages of the UCDSM should be clearly stated.

2. **It is premature to describe SSD analysis as the approach, as data are not available for application.** The preferred approach of selecting criteria based on statistical analysis of SSDs has a strong conceptual basis, but is impractical for application due to data limitations. A method for criteria development must be both technically sound and feasible. The SSD method is merit from a statistical basis, but cannot be demonstrated and evaluated, much less applied, with data currently available. This document should focus on a method that can be applied consistently with available data sets. I recommend revising the document to focus on the AF approach as the primary method for current use, with the SSD approach described as a preferred approach should data be available.

3. **Document needs to be reorganized.** The Phase II report provides a wealth of information that is presented in a methodical format. While the structure is logical, it is not user friendly and ineffective as a teaching tool. If this document is intended to provide guidance to others wishing to apply the method, then it should be organized with this purpose in mind. I suggest reorganizing the document into three main sections: 1) step-by step description of the approach, 2) detailed background/justification for the key steps, and 3) example calculation for bifenthrin, with additional samples calculations for each step. For example, I could not calculate the bifenthrin C_d values used in the example from the values given; apparently there are some changes in units needed for the calculations that are not clearly described.

4. **Conclusions section is not effective.** The conclusions section of the report provides a summary of what is contained in the report, but very few conclusions. Either retitle this section as a "Summary" or provide more information on the results of the evaluation of this

methodology to WQC, pros/cons related to other SQC methods, and research needs for future development.

5. **Too much reliance on previous documents**. The Phase II report relies heavily on previous work done on the UCDM for WQC. In some cases the reader is referred to previous reports for descriptions of methods or background on an issue. This document should be self-sufficient and include all of the information needed to justify or conduct the analyses.

6. **Do not calculate BSQC for small sample sizes.** The strength of the UCDSM is that it is based on statistical approaches used in other programs, adapted to sediments and pesticides. Calculation of criteria values for a sample size of 1 or 2 seems to violate the strong statistical foundation of the approach. The SMAV and AFs are very uncertain for such small sample sizes and will likely result in criteria that are highly overprotective. So many extrapolations or correction factors of uncertain accuracy need to be used with small samples that it seems the approach violates the statistical rigor that is intended. It would be better to establish stronger minimum data set requirements for criteria calculation and consider as provisional criteria any values calculated with very limited data.