Response to Peer Review Questions from Chris Ingersoll Dec 18, 2013;
cingersoll@usgs.gov, USGS, Columbia Environmental Research Center, Columbia, MO 65201; 573-876-1819)

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

General comments from CGI:

1. A tremendous amount of thought and effort has been put into the framework. However, the application of the framework to calculate BSQCs for bifenthrin is seriously limited due to a lack of:
   a. Robust estimates of Koc and Kdoc
   b. Uncertainty estimates for the geometric mean of the Koc and Kdoc values. A 16 to 18 fold difference is evident in acceptable values listed in Table 20. The variance in acceptable Koc or Kdoc values far out weights the variation in toxicity thresholds reported Table 3 of Xu et al. 2007 (only about 2-fold differences in toxicity thresholds across four sediments based lethality of midge and on dry weight or organic carbon normalized values in Table 3 of Xu et al. 2007).
   c. Chronic toxicity spiked sediment toxicity data for any species
   d. Acute toxicity spiked sediment toxicity data for sublethal endpoints (e.g., growth of *Hyalella azteca*)
   e. Acute or chronic toxicity spiked sediment toxicity data for a minimum of five species (e.g., amphipods and midge acute available). Ideally, there would be additional chronic data for both amphipods and midge and acute and chronic data needed for mayflies, mussels, and oligochaetes.
   f. Guidance provide on how to directly sample chemicals of concern in pore water isolated from whole sediment
   g. Guidance provided on how to estimate chemicals of concern in pore water (SPME, Tenax)
   h. Guidance provided on how to normalize pore water concentrations to dissolved organic carbon
   i. Guidance on how site-specific Koc or Kdoc values would be used to adjust the criteria (Section 2.1.3.5)
   j. Specific comments and edits have been provided in the Word file for the report and in the Word file for Appendix A. Please contact me if you have any questions or if you would like to discuss my comments future.

CGI responses to specific review questions:

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

CGI: Yes, in theory, but not in practice. A clear framework is provided for calculating BSQGs based on pore water concentrations or based on organic carbon normalized concentrations. However, insufficient chemical partitioning data and limited spiked sediment toxicity data are available to apply the proposed framework for any chemicals. Only one study has been cited that evaluates toxicity thresholds based on chemical concentrations expressed as dry weight, OC normalize, pore water, or free chemical, and unfortunately, only survival (of midge) is the toxicity endpoint reported (Xu et al. 2007). More examples need to be summarized (not just cited) in the report. Moreover, lethality of midge is typically a very insensitive endpoint, so even the toxicity data cited from Xu et al. (2007) fails to provide information on more meaningful endpoints for midge (weight or biomass).

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?

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

    CGI: Yes, these approaches are scientifically valid, but insufficient information is provided in the report illustrating the validity of using these measures to reduce uncertainty in bioavailability across different sediment types.

    A standalone section should be added to the main body of the report that provides a summary of all published studies with spiked-sediment toxicity data where the investigators have expressed toxicity thresholds as: dry weight, OC normalized, DOC normalized and as freely dissolved pore water concentrations. I underlined the word "toxicity" in this previous sentence because only one study has been briefly cited in the report by Xu et al. (2007) that actually provides toxicity data normalized using these various approaches (and again the Xu et al. 2007 study is limited because no growth endpoint was evaluate for the midge, only lethality).

    A standalone section should also be added main body of the report that provides clear guidance as to how one would DOC normalize pore water data, how one would estimate freely dissolved pore water concentrations using SPMEs or Tenax and how one could isolate pore water for direct measurement of chemicals of interest in pore water.

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?

    CGI: The framework can be used to evaluate toxicity data, but until there are new techniques for conducting chronic exposures and until new data are generated using

these methods to directly establish chronic effects, there will be substantial uncertainty associated with the BSQGs.

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

   CGI: It is clear as to how the rating guidelines in Tables 7 to 13 are applied based on information provided in Appendix A. However, I do not agree with the scoring system. There should be "critical" test acceptability criteria (TAC) established that would result in a study being categorized as unacceptable. For example, a study could fail control survival requirements, but be only reduced by a rating of 7.5 (see Table 8 for example).

   The minimum critical TAC should include: (1) control survival, (2) appropriate use of a solvent control, (3) equilibration of spiked sediment for at least 1 month before the start of an exposure, and (4) measured concentrations of chemicals in whole sediment (and in pore water?) at the beginning and end of an exposure. See specific comments on Tables 7 to 12.

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

   CGI: No, see response to Question 4.

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.

   CGI: Sections 2.4 and 2.5 are clear, but the definitions for LOEC and NOEC need to be updated, see comments provided in the Word file on the definitions of LOEC and NOEC).

   Section 8.7 data prioritization and exclusion for the bifenthrin criterion are not reasonable. See specific comments on Tables 20 and 21 of the report.

   Importantly, the ranking system for toxicity data does not identify "critical" test acceptability criteria (TAC) that should be used to judge if a study would pass or fail. Some of the data in Table 21 needs to be moved to Table 22 based on failure to meet critical TAC (e.g., 1 month equilibration period after spiking). Additionally, some of the data in Table 22 might be moved to Table 21. For example, the midge weight toxicity threshold should not be automatically placed in Table 22. The Instantaneous Growth Rate (IGR) in Table 21 is not an endpoint that is required endpoint in either USEPA or ASTM standard methods. Moreover, it has not been demonstrated that IGR is a superior to toxicity thresholds compared to absolute weight. Hence, IGR should not be used to exclude what appears to be robust data in

Table 22. Note that IGR is simply ending weight divided by starting weight, not a robust approach for expressing weight, particularly in studies were the starting weight is the same for all test organisms (as was the case it appears in the one study cited in Table 21 that reported IGR).

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

   CGI: Test acceptability criteria need to be added to Table 14 (e.g., control survival, control weight).

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

   CGI: Section 3.4 dealing with Species Sensitivity Distributions—Not my area of expertise. That said, an explanation of how 50% and 95% confidence limits are estimated is needed in Section 3.4.2.

   Section 3.5 dealing with Assessment Factors (AF)—I was totally lost by this explanation. An example for bifenthrin should be presented illustrating how the AFs were calculated using the toxicity data provided in Table 15. See specific comments on the text and tables in Section 3.5.

   Section 3.6 dealing with Acute-to-chronic ratios—Additional detail needed regarding the study by Giddings et al. 2006 (gray literature for midge: not sure what the endpoints were: 10-d larval survival vs. 60-d adult emergence and reproduction?).

   Section 3.7 dealing with Chronic criteria for herbicides—Not my area of expertise.

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?

   CGI: Providing the criteria for an averaging period does not make sense to me, given that field data are rarely (never) generated for periods for less than 10 day or for less than 28 days. Moreover, the studies cited in Section 3.8.2 do not appear separate out variance that might be associated with spatial heterogeneity in sediment contamination from variance that might be associated with temporal heterogeneity in sediment contamination.

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

    CGI: Yes, given the current state of the science.

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

    CGI: No, not my area of expertise.

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

    CGI: Insufficient information is available to determine if the bifenthrin criteria are over protective or under protective. In order to determine if the criteria are over or under protective a minimum of five chronic species-specific toxicity values are needed for at least three different sediment types generated with well equilibrated and well characterized sediments (including measures of reproduction of midge and amphipods, and growth of mayflies, oligochaetes, and mussels). Additional information is also needed regarding how quantifying uncertainties in the Koc values and Kdoc values in the derivation of the criteria (see comments on Table 20). Because of these uncertainties, the values should be reported as "sediment quality guidelines" or "sediment quality benchmarks" rather than as "sediment quality criteria". Specifically, there is currently too much uncertainty in chemistry or in biology to use these values for "regulatory compliance" in an analogous way that USEPA ambient national water quality criteria are used in a legally defensible manner.