

US EPA ARCHIVE DOCUMENT

# Final Summary Report

## Peer Review of EPA's Draft Document

### *Recommended Field-based Method For States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity*

December 22, 2014

#### Peer Reviewers:

Yong Cao, Ph.D.  
Marian R.L. Maas, Ph.D.  
Raymond P. Morgan II, Ph.D.  
Edward T. Rankin, M.S.  
Carl E. Zipper, Ph.D.

Contract No. EP-C-13-010  
Task Order 2014-16



**Prepared for:**  
Rachael Novak  
U.S. Environmental Protection Agency  
Office of Water, Office of Science and Technology  
Health and Ecological Criteria Division  
1200 Pennsylvania Avenue NW  
Washington, DC 20460



**Prepared by:**  
Versar, Inc.  
6850 Versar Center  
Springfield, VA 22151

## Table of Contents

<b>I. INTRODUCTION .....</b>	<b>1</b>
<b>II. CHARGE TO REVIEWERS.....</b>	<b>3</b>
<b>III. SUMMARY OF PEER REVIEWER COMMENTS BY QUESTION .....</b>	<b>6</b>
GENERAL IMPRESSIONS .....	7
RESPONSE TO CHARGE QUESTIONS.....	8
CHARGE QUESTION 1 .....	8
CHARGE QUESTION 2 .....	8
CHARGE QUESTION 3 .....	9
CHARGE QUESTION 4 .....	9
CHARGE QUESTION 5 .....	10
CHARGE QUESTION 6 .....	10
CHARGE QUESTION 7 .....	10
CHARGE QUESTION 8 .....	11
CHARGE QUESTION 9 .....	11
CHARGE QUESTION 10 .....	12
CHARGE QUESTION 11 .....	12
CHARGE QUESTION 12 .....	13
CHARGE QUESTION 13 .....	13
CHARGE QUESTION 14 .....	14
<b>IV. PEER REVIEW COMMENTS TABLES.....</b>	<b>15</b>
TABLE 1. GENERAL IMPRESSIONS .....	16
TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: <i>DATA SET CONSIDERATIONS</i>	
CHARGE QUESTION 1 .....	23
CHARGE QUESTION 2 .....	29
CHARGE QUESTION 3 .....	36
TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: <i>EXAMPLE CRITERIA</i> <i>CALCULATIONS</i>	
CHARGE QUESTION 4 .....	40
CHARGE QUESTION 5 .....	44
CHARGE QUESTION 6 .....	49
CHARGE QUESTION 7 .....	51
CHARGE QUESTION 8 .....	53

TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: *GEOGRAPHIC APPLICABILITY*

CHARGE QUESTION 9 .....	58
CHARGE QUESTION 10 .....	65
CHARGE QUESTION 11 .....	68
CHARGE QUESTION 12 .....	71

TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: *FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)*

CHARGE QUESTION 13 .....	76
--------------------------	----

TABLE 6. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: *FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)*

CHARGE QUESTION 14 .....	82
--------------------------	----

TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT .....

APPENDIX A- INDIVIDUAL REVIEWER COMMENTS .....	114
--	-----

## I. INTRODUCTION

U.S. EPA, through its authority under Section 304(a) of the Clean Water Act (CWA), has developed a draft method that states can use to derive field-based ecoregional ambient aquatic life criteria for conductivity, a measurement of ionic strength or concentration. EPA has also developed three case studies to illustrate how the field-based method can be used to develop criteria in ecoregions with different background conductivity and how to assess the geographical applicability of criteria developed for one ecoregion to a different ecoregion. The case studies use field data to demonstrate how to derive example criteria for conductivity for flowing waters dominated by calcium, magnesium, sulfate, and bicarbonate ions. Elevated conductivity has been shown to impact aquatic life designated uses in a range of freshwater resources. Elevated conductivity is associated with multiple anthropogenic sources, including discharge from wastewater treatment facilities, ground water infiltration affected by climate change, surface mining, oil and gas exploration, runoff from urban areas, and discharge of agricultural irrigation return waters, among others. Dominant ions associated with these sources may differ.

Among the documents EPA relied on to develop the field-based method for conductivity are EPA's 1985 "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses" and the 2011 EPA document, "A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams" (hereafter referred to as the "EPA Benchmark Report"). In the EPA Benchmark Report, a field data set was used to estimate a numeric conductivity benchmark. The method and the benchmark were then validated using an independent data set. The method and derivation of the conductivity benchmark and validation exercise were reviewed in 2011 by internal and external reviewers, including EPA's Science Advisory Board (SAB). The current draft field-based method uses the SAB-reviewed method as well as additional methods to estimate a protective maximum exposure concentration, duration, and frequency. It also presents a method for assessing applicability of field-based conductivity criteria developed in one geographic area to another area.

The purpose of the requested letter review is for EPA to receive written comments from five experts on the draft document and recommended field-based method. The review focused on the clarity of the descriptions and validity of the new components of the draft field-based method and case studies.

### Peer Reviewers:

#### **Yong Cao, Ph.D.**

Illinois Natural History Survey  
University of Illinois at Urbana-Champaign  
Champaign, IL 61820

#### **Marian R.L. Maas, Ph.D.**

Independent Consultant  
Bellevue, NE 68123

**Raymond P. Morgan II, Ph.D.**

University of Maryland Center for Environmental Science  
Appalachian Laboratory  
Frostburg, MD 21532

**Edward T. Rankin, M.S.**

Midwest Biodiversity Institute  
Columbus, OH 43221

**Carl E. Zipper, Ph.D.**

Department of Crop and Soil Environmental Sciences  
Virginia Polytechnic Institute and State University  
Blacksburg, Virginia 24061-0404

## II. CHARGE TO REVIEWERS

### Charge Questions:

#### Questions 1-3: Data Set Considerations

**1. Ion Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).

**2. Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

**3. Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

#### Questions 4-8: Case Studies: Example Criteria Calculations

**4. Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the  $\text{XC}_{95}$  and  $\text{HC}_{05}$  (Section 3.1, Deriving a CCC).

**5. Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum

concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

**6. Duration:** Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

**7. Frequency:** Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

**8. Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

#### **Questions 9-12: Geographic Applicability**

**9. General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

**10. Geographic applicability to a new area within an ecoregion:** Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.

**11. Geographic applicability to a new area in another ecoregion:** Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.

**12. Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial



streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

**13. General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the bootstrapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

**14. Protection:** Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.

### **III. SUMMARY OF PEER REVIEWER COMMENTS BY QUESTION**

## **Summary of Draft Peer Review Comments on *Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity***

### **GENERAL IMPRESSIONS**

All reviewers thought the document was well-written, and were impressed that the authors used a large dataset to support the methods and approaches used. They also provided comments on specific concerns and suggested potential improvements of the document.

Some general comments made included that there is some redundancy of writing throughout the document, some of the equations are not in standard formats, and some terms are not defined clearly. Also, one reviewer noted that there are several, rather significant, missing publications. With respect to conductivity specifically, one reviewer finds the manner in which the term “conductivity” is used is problematic and suggests that the term “specific conductivity” (SC) be used instead. Another reviewer recommends more discussion on the definition of reference condition, which will influence the derivation of criteria.

A couple reviewers provided suggestions that would help the States use the methodology. These suggestions included providing more explanation and examples of calculation processes, and also providing a discussion on critical elements of the monitoring plan needed to accomplish the methodology.

Specific concerns of individual reviewers include the following:

- Using the case studies to justify that criteria can be extrapolated from region to region is risky.
- Although one reviewer likes the Level III approach used in the document, the reviewer is concerned that there may not be enough applicable and robust databases for each of the 85 Level III ecoregions.
- Conductivity criteria should be considered in a tiered aquatic life use framework, as benchmarks could be over-protective of baseline warmwater aquatic life use and under-protective of exceptional (EWH) uses. The reviewer feels, however, that the methodology could be adjusted to take these factors into account.
- There is inadequate biological confirmation of results in the methodology. Additional analysis should be conducted to determine if the limit-defining taxa occur throughout the resources being proposed for application.
- One reviewer is unable to conclude that the proposed method can be implemented nationwide without unanticipated problems. This reviewer has concerns about applying an approach developed in Appalachia (where issues concerning elevated conductivity are well studied) to the rest of the country because of uncertainties in the causal mechanism given that both natural processes and anthropogenic activities release ions to waters, and

given that the relationship between aquatic biota and conductivity/major ions are not well documented in other areas of the country.

- Regulatory procedures intended to enforce the maintenance of 95% of reference taxa in local streams and rivers are not in place.
- The rigor of the criterion development process has been compromised because one cannot establish any consistent and meaningful relationship between the XC<sub>95</sub> of a genus and its extirpation.

## RESPONSE TO CHARGE QUESTIONS

### CHARGE QUESTION 1

***Ion Matrix Characterization:*** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium (Ca<sup>2+</sup>) plus magnesium (Mg<sup>2+</sup>) and the anions bicarbonate (HCO<sub>3</sub><sup>-</sup>) plus sulfate (SO<sub>4</sub><sup>2-</sup>) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by [SO<sub>4</sub><sup>2-</sup>] + [HCO<sub>3</sub><sup>-</sup>] > [Cl<sup>-</sup>]. Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for [SO<sub>4</sub><sup>2-</sup>] + [HCO<sub>3</sub><sup>-</sup>] > [Cl<sup>-</sup>] on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).*

The reviewers, in general, agreed with using the [SO<sub>4</sub><sup>2-</sup>] + [HCO<sub>3</sub><sup>-</sup>] > [Cl<sup>-</sup>] approach to develop the conductivity criteria. However, one reviewer noted specifically that it would make sense to include all sites with [SO<sub>4</sub><sup>2-</sup>] + [HCO<sub>3</sub><sup>-</sup>] **naturally** greater than [Cl<sup>-</sup>] in the dataset. One reviewer recommended running the analysis with no sample exclusion and to compare the XC<sub>95</sub> and HC<sub>05</sub> for a few selected sensitive genera and some important benthic assemblages. This reviewer was also concerned about eliminating samples for lotic systems. Another reviewer brought up possible confounding by other issues such as variation in conductivity at reference sites, and the presence of natural biodiversity hotspots which could drive XC<sub>95</sub>/HC<sub>05</sub> values. One reviewer also recommended that if there is concern about the merits of keeping or excluding data, the concerns should be directly discussed in the two introductory chapters, and should include a list of the pros and cons for excluding chloride and sites with < 6.0 pH.

### CHARGE QUESTION 2

***Catchment Size:*** *All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas (>155 km<sup>2</sup>) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate*

*anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

The reviewers provided varying responses to this charge question. Two reviewers thought that the discussion and analyses were generally adequate for the decision. In contrast, one reviewer thought the discussion was not adequate and there should be a stream-size cut-off, and provided justification. This reviewer noted that most importantly, that there are numerous studies which have all found altered benthic macroinvertebrate communities in low-order streams influenced by major ions discharged by coal surface mining, but there is no comparable supporting science for higher-order, high-drainage area streams. Another reviewer thought that only data from wadeable streams should be used because this is the critical field design driver for stream assessment with EPA, and many of the eastern States and NGOs. Finally, another reviewer thought that other variables were more important than stream size. This reviewer was more concerned with other natural classification issues and some anthropogenic changes that might have occurred from human habitation and land disturbance that are not acute or readily controllable, and are within a definition of “least impacted” streams.

### **CHARGE QUESTION 3**

***Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

All but one reviewer generally agreed with not weighting to account for seasonal effects since these effects are minor. One reviewer noted that this is the case as long as sampling time is not correlated with conductivity and suggested examining this correlation. The one reviewer who did not agree with the “not to weight” decision stated that there was no justification for a procedure that would mingle data from all seasons, and in fact Case Study 2 results justify the need for consideration of season. Another reviewer noted that care should be taken if including data for extreme stream flow conditions, and that antecedent conditions within watersheds should be examined carefully. Another reviewer recommended providing a discussion of how conductivity patterns might influence monitoring for compliance with any derived criteria.

### **CHARGE QUESTION 4**

***Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the  $XC_{95}$  and  $HC_{05}$  (Section 3.1, Deriving a CCC).*

All reviewers generally found the method for deriving  $XC_{95}$  and  $HC_{05}$  clear and easy to understand. A couple of reviewers suggested that more examples or instructions be provided, such as for calculation of weighted CDF values, statistical software to use (Excel, R, etc.), and bootstrapping. One reviewer believes the method for estimating extirpation threshold ( $XC_{95}$ ) is confusing and

problematic because it considers neither the direction of response of a genus to increased conductivity nor the relative frequency of the taxon (“probability of capture” in the text), and suggested two options to address this issue. Another reviewer noted that it is not clear what the CCC is intended to be within the context of a potential regulatory program. Finally, a reviewer suggested performing a check on the CCC by using the TITAN model to examine stressor (i.e. conductivity) relationships with biota.

#### **CHARGE QUESTION 5**

***Criteria Maximum Exposure Concentration (CMEC):*** *The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?*

Most reviewers thought the approach for deriving the CMEC seemed reasonable and clear, but had some questions or concerns. One reviewer felt that the CMEC concept was not well defined in the document and the logic for the CMEC derivation was not well supported, especially with regards to biological logic. This reviewer also suggested changes to the site selection procedures for CMEC derivation. One reviewer noted that the method is only reasonable if the authors determined the frequency of distribution (e.g. normal) before using Eq. 3-2, and used an appropriate critical value. Another reviewer also questioned if field data demonstrated that conductivity will vary in time independently and as a normal distribution. One reviewer noted that further empirical analyses of the consequences of the approach would be useful, and finally, one reviewer suggested that document provide information on how the 90% was determined, a working example of Figure 3-6, and greater description and an example of LOESS.

#### **CHARGE QUESTION 6**

***Duration:*** *Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Most reviewers thought the description and justification appeared adequate, and none were aware of additional publications. However, two reviewers suggested analysis of continuous monitoring datasets to examine duration questions in more detail. One reviewer commented that a CCC should be used with caution because, as illustrated in Figure 3-7, there is the potential for large yearly variations in stream conductivity. A reviewer also suggested adding footnotes indicating which data points represent only one sample per year.

#### **CHARGE QUESTION 7**

***Frequency:*** *Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*



Most reviewers thought the description and justification supporting the estimation of frequency appeared adequate, and none provided additional publications. One reviewer stated that the recovery of macroinvertebrates strongly depends on nearest sources, therefore, three years frequency should be enough for re-colonization if exposure occurs at a local scale, but may not be sufficient if exposure happens at broad scales. Another reviewer suggested ambient analyses as another form of evidence, such as deriving a “biological stressor metric” using the XC<sub>95</sub> values. Finally, one reviewer noted surprise at the high level of conductivity before extirpation of sensitive crustaceans, and also noted that in the causal assessment methodology, more consideration should have been given to other known stressors.

#### **CHARGE QUESTION 8**

***Alternate measurement endpoint: Is the example alternate measurement endpoint ( $\text{HCO}_3^- + \text{SO}_4^{2-}$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?***

All reviewers thought the alternative measurement endpoint was clear and adequately supported, and preferred conductivity as the primary measurement. However, one reviewer would like to see additional investigations to document the appropriateness of  $[\text{HCO}_3^- + \text{SO}_4^{2-}]$  as a biotic condition indicator that would provide information other than that which is provided by conductivity. This reviewer also noted that there is no support for an argument that  $[\text{HCO}_3^- + \text{SO}_4^{2-}]$  would be a “better” endpoint than conductivity; but does not presume it to be more or less representative as an indicator of the “actual toxicant” because the actual toxicant or toxicants is/are unknown. Two reviewers noted that states will most likely focus on conductivity, and not the alternative measurement, because of cost and ease of measurement. In addition, a reviewer noted the criticisms are the same for this measurement as for conductivity, with respect to estimating XC<sub>95</sub> and HC<sub>05</sub> (i.e. one cannot establish any consistent and meaningful relationship between XC<sub>95</sub> of a genus and its extirpation).

#### **CHARGE QUESTION 9**

***General: Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?***

The reviewers generally thought the process was clearly described, however had some concerns. One reviewer felt that the concerns of the SAB were not fully addressed with regards to macroinvertebrate fauna comparability. Another reviewer interpreted the SAB’s recommendation to mean a direct, species-by-species comparison, rather than looking at a set of species/genera and how the communities in general respond to a stressor, as the authors of this document chose to do. Another reviewer commented that a statement might be made that the processes would be

applicable for any Ecoregion III level imbedded within any Ecoregion II level. This reviewer also noted that a good test might be to examine two Ecoregion II levels watersheds that are not contiguous, or two watersheds close to each other and two watersheds distant from each other. One reviewer noted that modifications would be needed to derive benchmarks under a tiered approach, and discussed this issue. The reviewer also questioned if the process addresses conditions where a subset of streams may need to be considered separately because they have uniquely lower conductivity. In addition, a reviewer stated the following concerns:

- The process is not fully supported as a reasonable process for regulatory development.
- There is no supporting evidence for the underlying assumption that if background conductivity estimates for two regions are similar, then sensitivity of regional taxa to elevated conductivity will be similar as well.
- Biological confirmation is missing and a method for evaluating biological data should be included. One specific deficiency that the reviewer notes is that the selection of which percentile to use as a background conductivity value is left to judgment, and the databases examined by the reviewer indicate that the 5<sup>th</sup> and 25<sup>th</sup> percentiles for conductivity distribution can differ by multiples of 3 to 10.

#### **CHARGE QUESTION 10**

***Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

The reviewers generally thought the process was clearly described. One reviewer, however, did not understand why samples from the whole ecoregion were not be included to develop region-wide criteria. Two other reviewers had similar comments as for Charge Question 9, with regards to using the approach in a tiered use framework, and verifying the assumption that taxa comprising benthic macroinvertebrate communities within the smaller area are similar to those that occur within the larger area (i.e. biological confirmation).

#### **CHARGE QUESTION 11**

***Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

The reviewers generally thought the geographic applicability analysis was clearly described, and many had similar comments to the previous charge comments. One reviewer questioned the results of the case study because the results may differ substantially if the new ecoregion is further away and associated with very different benthic fauna. One reviewer wanted more clarity on a few details, such as the process for defining background conductivity, the rationale used in Case Study 3 for the selection of the single background conductivity estimate, and how the 95% CIs were



derived for the background conductivity estimated. Another reviewer noted the large confidence interval difference between Ecoregion 70 and the other regions, and requested an explanation for the difference. This reviewer was also concerned about the need for ~500 samples in order to achieve consistent results with the HC<sub>05</sub> derivation. Again, one reviewer stated that the background matching process should be justified with biological confirmation, and another reviewer had concerns using the approach in a tiered use framework.

#### CHARGE QUESTION 12

***Applicability to ephemeral streams:*** *In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?*

Three of the reviewers specifically stated that temporal/ephemeral streams are important and need to be protected. One reviewer commented that these streams may be “over-protected” by a HC<sub>05</sub> established for perennial streams. Again, another reviewer commented the need for biological confirmation to apply the HC<sub>05</sub> derived from a permanent stream to an ephemeral stream, such as verifying the presence of the permanent streams’ limit-defining taxa within the ephemeral streams. One reviewer suggested the following three references for EPA to incorporate:

Lake, P.S. 2011. *Drought and aquatic ecosystems: effects and responses*. Chichester, UK. Wiley-Blackwell.

Steward, A.L., D. von Schiller, K. Tockner, J.C. Marshall, and S.E. Bunn 2012. When the river runs dry: human and ecological values of dry riverbeds. *Frontiers in Ecology and the Environment* **10**: 202–209

Williams, D.D. 2006. *The biology of temporary waters*. New York, NY: Oxford University Press.

#### CHARGE QUESTION 13

***General:*** *The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification*

***to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.***

Most of the reviewers expressed some concerns regarding the approach to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values. One reviewer stated that sample comparability of the various data sources needs to be evaluated (i.e. sampling methods between agencies are likely different, which could introduce noise into the analysis) and that numerical thresholds for stream size are actually arbitrary. Two reviewers commented that certain species of fish should be excluded. Other concern/suggestions of individual reviewers were:

- Exclude smallmouth bass because it is generally intolerant of pollutants and poor water quality, and another reviewer would like to exclude both rainbow trout and brown trout because they are exotic introduced species.
- Reference sites were not identified in the dataset, and heavily forested sites do not necessarily assure high water quality.
- How the regions are combined given biogeographically differences in fish distributions across ecoregions, and again would like to see discussion of tiered uses.
- Bootstrapping process is reasonable, but again pointed out criticisms on the XC<sub>95</sub> estimation and it's relevancy to species extirpation and SD curves.

#### **CHARGE QUESTION 14**

***Protection: Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.***

The reviewers generally agreed that the criteria would be protective of fish. One reviewer stated that no additional data and analyses are needed. One reviewer commented that sample sizes in some areas are continually growing as monitoring programs mature, and so State water quality standard programs should be continually exploring their databases to refine aquatic life uses and the criteria. The reviewer also commented that robust data is needed to conduct the analyses in this document. Again, concerns previously raised were made, such as the lack of biological relevancy of XC<sub>95</sub> to species extirpation and vague interpretation of SD curves.

#### IV. PEER REVIEW COMMENTS TABLES

**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
Reviewer 1	<p>It is great to see U.S. EPA developing a field-based method for establishing aquatic-life criteria for conductivity, an increasingly important stressor for freshwater ecosystems. The effects of abiotic (habitat, flow, water chemistry) and biotic factors (e.g., competition and predation) on the responses of a taxon to increased conductivity are complex and poorly understood. It is therefore sensible to use a field-based approach, rather than lab tests, to derive the criterion. I also believe that genera are the best choice of taxonomic units because the sensitivities of species from the same genus are often similar, and the identifications at this level are normally more accurate and less costly than at the species level. Clumping taxonomic data to the genus level also increases the number of taxon occurrences and makes more taxa available in a region for establishing a conductivity criterion. Overall, the document is well written. However, I have several major concerns, particularly on the concept and measure (XC<sub>95</sub>) of taxon extirpation and associated statistical analysis. The vague and inconsistent relationship between XC<sub>95</sub> and extirpation appears to have compromised the rigor of the process of criterion development. I am also worried about how specific case studies in the document are used to justify extrapolation of a conductivity criterion developed for one region to another region. In addition, some terms (e.g., probability of capture) need to be more clearly defined, and equations need to be constructed in a standard format.</p>	
Reviewer 2	<p>It was a pleasure to review the US EPA draft document, <i>Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity</i>. As a biologist who has worked both in the field and in managerial capacity for water quality monitoring and bioassessment, the area of specific conductivity has long been more or less overlooked. This is largely due to two reasons: 1) most city, watershed and state monitoring programs don't quite know what to do with conductivity data, and 2) the role of conductivity in its effect on aquatic and benthic organisms is not well understood. The need for a criterion is basic to the improvement in these areas. This document's clear and strong guidance in providing a method for development of conductivity criterion and for a method to make it applicable to adjoining regions is an immensely valuable new and long over-due tool for monitoring programs. The document provides considerable information on the effects of high conductivity levels on macroinvertebrates and fish, and provides strong, data-supported rationale for its approaches and methods. Very large data sets, paired analyses, and strong/reliable/widely used statistical models were used in all of the analytical processes. The biological information in all sections was especially accurate, thorough, and clearly written. It was a pleasure to read those sections and to learn new information. The statistical material was less clear</p>	

**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>for me, but that is more of a deficiency on my part than that of the document. In that regard, perhaps more explanation of several of the calculation processes could be provided, and a full, working example of each (probably placed in the Appendices), would be helpful for water quality staff who have limited statistical training. With such examples, staff could follow the step-by-step process. I realize this might be viewed by the authors as somewhat of an unnecessary effort, but I believe it would help the document's usability by a greater number of staff with varying backgrounds and knowledge base.</p> <p>Thank you for the opportunity to review this excellent document. The document is well done and its conclusions correct. I believe it will be a valuable guidance for development of much needed criterion for conductivity. It will provide an immensely important function in the improvement of water quality and ecological health for the nation's streams and rivers.</p>	
Reviewer 3	<p>I was impressed by the overall depth and breadth of this very well prepared report by EPA (in my opinion it is one of the best technical reports, within my research areas, ever prepared by EPA). Obviously, this report has already gone through a rigorous internal EPA review (by many people who I know professionally) and by EPA contract support, as well as review by several other people who I also know and respect for their work. There is no question that the internal EPA technical workgroup contributed strongly to the report quality, again many people that I know professionally and respect.</p> <p>Having worked with acid mine drainage, acid rain and numerous stream chemistry studies (as well as a few other lotic and lentic projects where conductivity was measured, besides a past life in estuarine and marine ecosystems measuring salinity), I am quite familiar with the strengths and vagaries of this very important measurement in both the field and laboratory. Also, I was pleased to see most of the key, but rather ancient, papers cited (e.g., the 1985 Hem paper), indicating that the literature review was excellent (although the key Hem 1982 paper was missing). However, there were also some recent, rather significant papers missing, but that may be due to the timing of the report preparation. No matter how hard you try, there are always supportive papers that may be missed in any literature review.</p>	

**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>One concern was the redundancy of writing throughout the report (clarity was excellent for most sections), where general concepts appear to be often restated within some sections of the main report. I don't think that you always need to restate the obvious throughout every section of the main body of the report. However, with the potential wide array of future readers, some writing redundancy may be helpful. Here, my advice would be to have an outside professional editor (e.g., someone from Academic Press, Science, etc.) review the report structure and make recommendations to streamline this effort.</p> <p>I like the Level III ecoregion approach, a method that I am using in some of my own work. My only concern here would be if there are applicable and robust data sets for each of the 85 Level III ecoregions. The report certainly uses a very robust, regional database to develop criteria and to examine the statistical techniques needed to develop conductivity criteria within an ecoregion, and adjoining ecoregions. It will be really interesting to see how the States and Tribes respond to this report, as well as Congress.</p>	
Reviewer 4	<p>Overall, this is a well-written, scientifically sound paper which lays out the technical approach and underpinnings for deriving conductivity benchmarks for aquatic life for streams using field-derived measures of water chemistry and ionic strength measures with co-currently collected measure of aquatic macroinvertebrate response at the genus level of taxonomy.</p> <p>My major issues are related to the actual application of these results to protect aquatic life uses in State water quality standards and, particularly, under tiered aquatic life use frameworks. Application of this method in states, like Ohio, that have tiered aquatic life uses, could result in benchmarks that are overprotective of the baseline warmwater aquatic life use, but could also be under-protective of exceptional (EWH) uses. Fortunately, I think the methodology presented here can take these factors into account. For example, for "EWH" streams there may be a more restrictive suite of species that occur in these waters and exclusion of more tolerant forms could drive the XC95 a bit lower. In contrast, for "WWH" streams, the most sensitive species may not occur frequently enough in those streams to "drive" the XC95 benchmarks and the more common sensitive species might result in a less stringent, but perhaps more attainable benchmark.</p>	



**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>Another key issue that may influence the derivation of criteria with this method is the definition of reference conditions. I think some more discussion of reference conditions as in Stoddard et al. (2006) would be helpful, I will delve into these comments in more detail in the charge questions below.</p> <p>I do think this paper provides a solid technical basis for deriving benchmarks using field derived SSDs and the derivation of HC<sub>05</sub> and XC<sub>95</sub> values. My comments focus on the need to deal with some of the application issues surrounding these benchmarks. The ability of a State to use this methodology will be related to the quality of their monitoring, assessment, and water quality standards programs (Yoder and Barbour 2009). Reference to the critical elements that monitoring programs should have to accomplish this methodology should be a part of this document. Again, more detail will be provided below.</p>	
Reviewer 5	<p>For the most part, the document is well written – although with some exceptions that I have noted below. I was impressed with the thoroughness of the presentation. I found most of the substantive information in the document to be accurate. I did find some errors, and I differ with some of the interpretations.</p> <p>I find language used to express essential concepts as problematic, especially the manner in which the term “conductivity” is used to express what is more widely described as “specific conductance”. If EPA persists with its current use of the term “conductivity” within the context of the program proposed, the term “conductivity” would then have two different meanings: electrical conductivity (the raw measure) and the 25°C-temperature-corrected value. This result can only breed confusion. I strongly encourage EPA to use the words “conductivity” and “specific conductance” (SC) in accord with well-recognized and widely used precedents and practice.</p> <p>Given the current status of scientific knowledge concerning major-ion effects on aquatic biota in the Appalachian coalfield, I see the primary method presented by the document and illustrated by Case Studies 1 and 2 as generally adequate, as a temporary measure, for describing specific conductance (SC) levels that will be protective of 95% of benthic macroinvertebrate taxa in Appalachian coalfield streams. I say “temporary” given the lack of scientific certainty concerning the precise nature of the stressor that is causing the effects that are being observed so widely. I expect that the stressor will</p>	

**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>be identified with greater certainty, eventually, and that will allow more precise targeting by regulatory and other management actions. I do have some technical concerns which are described in the responses. Many of my technical responses are focused on what I see as inadequate biological confirmation of results which are obtained from analysis of large datasets. While the conclusions derived from such analysis would likely be considered as adequate if expressed with appropriate caveats in the context of academic studies, these results are proposed for application to individual situations of widely varying circumstances as a regulatory program.</p> <p>Clearly, aquatic communities are depressed in waters influenced by mining throughout the Appalachian coalfield. Given that numerous studies have found close associations between elevated SC/major ions from mining and alterations of benthic macroinvertebrate community metrics, it is reasonable to expect that some effect on water by surface coal mining is playing a major role. Given the number of studies that have found negative associations between elevated SC and benthic macroinvertebrate conditions in the Appalachian coalfield, and the lack of relevant studies that have failed to find such effect, release of SC/major ions has to be considered as prime suspect. However, the direct causal agent – i.e., the precise combination of ions and/or SC-associated stressor such as, perhaps, specific ion combinations or ratios, mining-induced hydrologic alterations, or other unstudied factors – is not known. Hence, it is my view that any public policy actions taken should recognize that the science defining causation is not yet settled. Hence, I see the following statement from the document's forward as appropriate: "State and tribal decision makers would retain the discretion to adopt approaches on a case-by-case basis that differ from those described in this draft document, even if the method in this document is issued under CWA section 304(a)."</p> <p>I have been conservative in my interpretations due to recognition that the document's content has the potential to become a regulatory program. Given the consequences of the potential regulatory actions that may be based on HC<sub>05</sub> values derived as described here, ensuring those values' validity across the full range of resources targeted for application requires additional biological confirmation. The HC<sub>05</sub> values are being defined by SC values associated with small numbers of taxa ("limit-defining taxa"). Those values should be checked by conducting additional analysis to determine if the limit-defining taxa occur throughout the resources being proposed for application. For example, the ecoregions used for the examples of Chapters 5 and 6 extend over considerable distances in the</p>	



**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>north-south direction, and climatic differences can be expected to occur throughout such ecoregions. Do the limit-defining taxa occur only within one portion of the ecoregion, or throughout? Such logic can also be applied across stream orders, and across other dimensions that define the extent of water resources proposed for application. Given the potential consequences of a regulatory program, I see the additional assurances that would be provided by such confirmations as prudent and essential.</p> <p>I have concern with the document's attempt to nationalize an approach developed in Appalachia where issues concerning elevated SC are well studied. Given the lack of understanding that concerns the causal mechanism, given that ions at issue are released to waters due to both natural processes and anthropogenic activities, and given that relationships of aquatic biota to SC/major ions are not well documented in other areas of the country (at least to my knowledge), I am unable to reach the conclusion that the proposed method – and its reliance on either SC or <math>[\text{HCO}_3^- + \text{SO}_4^{2-}]</math> as a measurement endpoint – could be implemented without unanticipated problems in other areas of the country.</p> <p>I reach my conclusions concerning the method's adequacy reluctantly due to several related concerns:</p> <ul style="list-style-type: none"> <li>• Concern for the effect that a water quality criterion for (SC) would have communities throughout the Appalachian coalfield and the people who live there, given the historic and recent importance of coal mining as an economic activity that brings money into region. The economic and human effects of recent coal-mining declines in these communities are severe, and implementation of a <math>\sim 300 \mu\text{S}/\text{cm}</math> water quality criterion would continue that trend.</li> <li>• Concern for “regulatory equity.” or a lack thereof in this case. As I understand Clean Water Act implementation procedures elsewhere in the US, such as urban, agricultural, and residential areas: Regulatory procedures intended to enforce maintenance of 95% of reference taxa in local streams and rivers are not in place, as the multimetric indices that are commonly used for biomonitoring and bioassessment are developed on a different basis.</li> </ul> <p>I express these concerns with expectation that a <math>\sim 300 \mu\text{S}/\text{cm}</math> criterion, if established as a firm limit</p>	

**TABLE 1. GENERAL IMPRESSIONS**

REVIEWER	COMMENT	RESPONSE
	<p>in the Appalachian coalfield, would fail to incentivize further development and implementation of the mining and reclamation technologies that are intended to reduce mining environmental impacts and improve environmental restoration – the incentive would be to shut the mines down.</p> <p>I also have concern for environmental quality in the Appalachian coalfields; that concern is informed by recognition that regional ecosystems are among the richest (biologically) and well-preserved non-tropical ecosystems on the face of this earth; and that the scales of mining operations and mining effects are large and growing.</p> <p>With all of that said: I have reviewed the document objectively and have endeavored to provide my professional and technical opinions without bias.</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
Reviewer 1	If NaCl in stream water is from natural sources, it would be appropriate to exclude those samples dominated by $\text{Cl}^-$ . However, if it is clearly from human activities, such as road de-icing, exclusion of the samples will make the conductivity criterion derived not applicable to NaCl contamination, a major stressor in streams of the snow zone. It seems to make sense to include all sampling sites where $[\text{SO}_4^{2-}] + [\text{HCO}_3^-]$ is <b>naturally</b> greater than $[\text{Cl}^-]$ .	
Reviewer 2	<p>I believe it is appropriate to remove samples (data) from the data set which might move the results from reflecting the true condition. The more “types” of data included in a database, outliers, for example, the more general/less specific will be the results – and therefore, less accurate.</p> <p>The question for this study was “how to derive example criteria for conductivity for flowing waters dominated by <u>calcium, magnesium, sulfate and bicarbonate ions</u>” (pg. xvi), and <u>not</u> for flowing waters dominated by <u>chloride ions</u>. All of these ions are predominant throughout the study’s ecoregions because of the geology, physiography, vegetation, animal life, climate, soils, water quality, and hydrology found here. However, calcium, magnesium, sulfate and bicarbonate come from weathering of limestone and dolomite (the geological composition of this region) and are the ions which have the greatest impact on specific conductivity which is the intent of this study. Although chloride ions are also prevalent, the decision to exclude chloride anions is logical and appropriate.</p> <p>Additionally, the decision to exclude sample sites with <math>&lt; 6</math> pH is also probably wise, although this is perhaps less definitive. Acidity directly affects conductivity by causing calcium and magnesium to become more mobile with decreasing pH, thus having a clear role in</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
	<p>conductivity levels. But the level of its effect and its associated variables – such as temperature – would then also need to be considered, increasing the study’s data needs and broadening the question. On the converse, acidic conditions do exist in waters of this geographical region because of anthropogenic influences such as urban stormwater runoff, surface mining runoff, gas/oil extraction waste water, and aerial deposition. And from this standpoint only, there might be adequate justification for its inclusion. However, since <math>&lt; 6</math> pH waters in this study were not large in number, the decision to include or exclude could go either way. Would their inclusion have had much influence? A basic rule of thumb for most scientific studies is “the more specific the testing or measuring, the more specific and accurate will be the results”.</p> <p>Toxicities of ions differ, and keeping the data collection and the subsequent analyses limited to the four ions ensures data results free of the additional variables inherently associated with any additional ions. Any field-based study should limit its parameters of study for this reason. Samples from waters with only the same ionic composition will yield the most representative and accurate results.</p> <p>The authors point out (pg. 2-11) that the relative concentration of bicarbonate is pH dependent, and that the dominant form of the ion in soil is bicarbonate at circumneutral pH. This gives further justification for limiting collection of samples to waters with <math>&gt; 6.0</math> pH.</p> <p>The authors have done an excellent job in discussing the many factors and general background information in Sections 1, 2, and 3. The discussion in these sections is valuable</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
	<p>for the reader, and is thorough and clear in its presentation. The background information is presented objectively and will be helpful and adequate for state water quality staff.</p> <p>If there is concern about the merits of keeping or excluding data, I recommend the question be directly discussed in the two introductory chapters. Even though the authors have discussed the reasons why they excluded chlorine and acidic conditions (actually multiple times throughout the document), a table that straightforwardly addresses the pros and cons could be included. List the pros and cons for excluding chloride and sites with <math>&lt; 6.0</math> pH is my recommendation.</p> <p>I agree with the inclusion of all other data, i.e., impaired and high quality streams, all stream sizes, and sampling from all seasons.</p>	
Reviewer 3	<p>I like the approach for using the ionic basis of: <math>[\text{SO}_4^{2-}] + [\text{HCO}_3^-] &gt; [\text{Cl}^-]</math> to develop the initial conductivity criteria. This step alone eliminates any problems due to the potential effect of road salts, especially throughout the Appalachians and the Eastern Seaboard in general. It would be an interesting exercise to run the same analyses with no sample exclusion, and then do a comparison of <math>\text{XC}_{95}</math> and <math>\text{HC}_{05}</math> for only a few selected sensitive genera and some important benthic assemblages (e.g., EPT). I would assume that these would not be too time consuming, but may be worthwhile if there are ecoregions with lower sample sizes than the very rich data set employed in this report.</p> <p>I would be a little concerned with any fall samples collected during an extreme drought period. If one assumes the normal two-component groundwater mixing model for eastern ecoregions, there is the possibility that a severe drought could result in over 95% of the stream</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

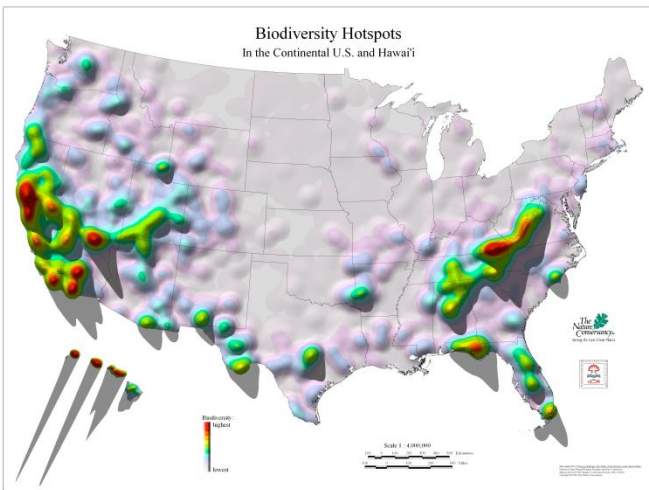
**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
	<p>flow coming from deep groundwater, and would represent an anomalous case for stream chemistry (successive years of drought may also be a very strong stressor on aquatic biota). It may be best to exclude any sample pairs (biota X conductivity) collected where gaged stream flows in a watershed, or a series of watersheds, dropped to below the 5<sup>th</sup> percentile of long-term flow records. Also, any exceptionally high-water events (greater than 99<sup>th</sup> percentile or perhaps 100-500 year storm events) may need to be considered if they occurred in the year before sampling. Benthic assemblage recovery (as cited in the report using the classic paper by Wallace, 1990) may take more than one year, depending on the species complex present in the stream and nearby refugia. Over my career, I learned quickly that there is no such thing as a normal year, and benthic and fish field collections need to be correlated with antecedent climatic conditions (e.g., temperature, flow, etc.).</p> <p>Not being very familiar with the water chemistry of western streams, I believe it may be important to think seriously about any exclusionary criteria for these lotic systems. However, I know that some of the mid-western and western states have good data bases with which to run the same analyses as done for ecoregion 69.</p>	
Reviewer 4	I have no problem with removing sites from the analyses where different ionic mixtures are likely to confound results (e.g., $[\text{Cl}^-] > [\text{SO}_4^{2-}] + [\text{HCO}_3^-]$ ) or where other stressors (e.g., $\text{pH} < 6$ ) may also contribute to confounded results. I do have some question on whether there could some other confounding caused by: 1) natural variation in conductivity at “reference” sites, and 2) variation in conductivity along a gradient of sites that may be considered “reference” in the sense of “least impacted” conditions vs. “minimally disturbed” in the sense described in the paper by Stoddard et al. (2006).	



**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
	<p>It is clear that there are clearly natural biodiversity “hotspots” in the ecoregions examined here (see example Nature Conservancy map). Some of these hotspots may partly be remnants of where biodiversity has been minimally disturbed by human activities, but some are where there is a combination of natural features (e.g., habitat, gradient, elevation, water chemistry) that combine to maximize biodiversity. My concern is that these natural “hotspots” may well be driving the <math>\text{XC}_{95}/\text{HC}_{05}</math> value particularly when aquatic life use potential is defined by a single aquatic life use, and therefore a single benchmark is derived. The effect of a single benchmark is that it may be under-protective of the most unique “hotspots” but overprotective of more typical habitats. I will address this comment more specifically below.</p> 	
Reviewer 5	<p>Data from sites with elevated TDS/SC but with ionic composition that differs from the dominant ion matrix (i.e. dominated by <math>\text{Ca}^{2+}</math>, <math>\text{Mg}^{2+}</math>, <math>\text{HCO}_3^-</math>, <math>\text{SO}_4^{2-}</math>), should be excluded from the datasets used for the analysis, as described by the document. Scientific literature is clear in demonstrating that the ionic composition of TDS influences the SC/TDS concentration at which toxic effects are observed (Mount et al. 1997). Scientific literature</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 1: Matrix Characterization:** The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?)

REVIEWER	COMMENT	RESPONSE
	is also clear in documenting that that $\text{Ca}^{2+}$ , $\text{Mg}^{2+}$ , $\text{HCO}_3^-$ , and $\text{SO}_4^{2-}$ are the predominant dissolved ions in most Appalachian coal-surface-mine influenced waters (Bryant et al. 2002; Pond et al. 2008; Fritz et al. 2010; Timpano et al. 2011; Agouridis et al. 2012; Bernhardt et al. 2012; Lindberg et al. 2012; Wood and Williams 2013; Pond et al. 2014; Sena et al. 2014).	



**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The analyses and discussions provided are adequate for the decision to include all stream samples regardless of catchment size. But, can the criteria developed be applied to great rivers, like Mississippi, Ohio, and Colorado rivers? These rivers support very different aquatic fauna, likely fewer sensitive genera, but some unique ones. If no large-river samples are included, could the criterion derived protect those unique taxa? Or, may the criterion be over-protective of large rivers?	
Reviewer 2	<p>It is excellent that all stream types and sizes were included in the data sets for the case studies, especially the smaller and intermittent streams. Smaller streams, both perennial and intermittent, are where valuable macroinvertebrate habitat is most often found. These are likely to have the appropriate streambed composition, rocks and logs for colonization, leaf litter, bank overhangs, and freedom of siltation – which are all crucial for macroinvertebrate life cycles, population abundance and diversity. So often only the larger, perennial streams and rivers are studied. The authors are “right on” when they point out that discharge from headwaters, intermittent and even ephemeral streams ultimately affect downstream stream reaches and rivers. This is often not understood or realized fully by program managers, who are not well versed in stream ecology, and policy makers. Additionally, the authors make an important point in that many macroinvertebrate taxa often use temporary streams for at least a portion of their life cycle. Much of my experience in stream ecology and water quality has been with the smaller streams and it is my belief that their value to the river system and its taxa cannot be over-emphasized. I thank the authors for their recognition of this.</p> <p>Exclusion of data from the larger catchment areas is, however, worthy of a little discussion here. The authors present four good reasons for not excluding them: 1) sensitive genera were</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	<p>found in the larger rivers; 2) inclusion of data from larger rivers did not significantly change the magnitude of the hazardous concentration; 3) Analysis of 3115 sites with drainage areas up to 17,986 sq km showed a very weak (a very weak, indeed!) correlation of conductivity and drainage area; and 4) background conductivity estimates for drainages <math>&gt; 155 \text{ sq km}</math> were within confidence bounds for establishment of background values.</p> <p>However, the EPA's Benchmark Report initial exclusion of larger streams – because sampling methods might differ for non-wadeable streams – has substantial merit. Sampling methods are indeed different for the larger rivers, and large river sampling requires greater resources (time, staff, boat/equipment) and, therefore, also happens less frequently. Collected macroinvertebrates in larger rivers can be low in numbers as well – due probably to a combination of factors: manmade channel morphology changes, river velocity too high, fewer colonization sites, poor habitat, deposition of sediment, anthropogenic contaminants, and difficulty in sampling at greater depths and velocities. Thus, more variability likely exists in data for macroinvertebrate databases for large rivers. However in this study, sensitive taxa were documented in the larger rivers, so perhaps collection methods and expertise in sampling has improved, but perhaps more importantly, these rivers are likely of higher quality than those here in the Midwest of which I am familiar and which are heavily impacted by agriculture.</p> <p>In conclusion, the authors have provided good discussion and support for the decision to include all samples regardless of catchment size. A bit more discussion as I have presented here might be helpful but probably is not necessary.</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	Lastly, I wish to reiterate the value of data from intermittent and ephemeral streams. These small streams provide irreplaceable habitat for macroinvertebrates, invertebrates, amphibians, aquatic/wet terrestrial species of all kinds. Their loss has been significant through ditching and tiling in agriculture, diverting and damming for irrigation, and in placing into underground pipes in urban development.	
Reviewer 3	<p>My personal opinion, and scientific bias, is to use only data from <u>wadeable streams</u> – this is the critical field design driver for stream assessment with EPA, and many of the eastern States and NGOs. EPA, along with many States, did a lot of work on developing such protocols to assure that there was robust physical, chemical, and biological data collected in order to make non-biased estimates of many important parameters. Indeed, many key biotic and habitat metrics were developed based solely on wadeable streams. Also, 1<sup>st</sup> through 3<sup>rd</sup> order streams may constitute 70-90% of stream km in an ecoregion, with larger streams (4<sup>th</sup> to 12<sup>th</sup> order) representing less than 10-30% of stream km.</p> <p>If one follows the River Continuum Theory, the 1<sup>st</sup> through 3<sup>rd</sup> (and perhaps some small 4<sup>th</sup>) order streams are where the real action is, and that the larger streams and rivers (large 4<sup>th</sup> to 5<sup>th</sup> and higher) start to reflect a major change in both ecological structure and function. OK, so one may collect some benthic organisms (genus may be the same, but probably different species) in the larger order streams that would also be found in lower order streams. However, stream processes in the larger order streams are so different I feel it would be difficult, and unjustifiable, to use this approach. Obviously, EPA would welcome this opportunity to be able to set conductivity criteria for large aquatic ecosystems (large stream and rivers), especially in light of the NPDES permits, etc.</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
Reviewer 4	My concern with this discussion is not so much with stream size as important variable, but other natural classification issues and some anthropogenic changes that might have occurred from human habitation and land disturbance that are not acute or readily controllable and are within a definition of “least impacted” streams. For example, in the mountainous regions of the WAP ecoregion in West Virginia for example, the relief has led to land uses (e.g., forestry, park, light agricultural, low density residential) that result in more highly forested ( $>90\%$ ) reference conditions. Along the edge of the WAP ecoregion in Ohio for example, the relief is more variable and farming and some other land use changes are somewhat more intense. “Least impacted” reference sites are much less likely to be “ $>90\%$ forested.” This broaches the important question of whether a single benchmark or multiple benchmarks to match tiered uses are more appropriate.	
Reviewer 5	My discussion below assumes that reference streams are low-order, small-drainage-area streams.  No, the document does not provide adequate basis for including all observations, regardless of stream size. There should be a stream size cutoff, and EPA should provide guidance on an appropriate cutoff. One factor in defining the stream-size range appropriate for the analysis concerns reference sites. As stated on page 2-1, line 23-24: “Genera that are not observed at reference sites ... are excluded from the data set.” Therefore, only streams of size classes where community compositions can be documented as being similar to those at reference sites should be included; or, only taxa found to be both occurring at reference sites and as characteristic of the higher-order streams should be included.	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	<p>There is a large volume of scientific literature supporting the understanding that aquatic communities and community compositions vary by stream size (e.g. Vannote et al. 1980). Scientific literature documents the taxonomic differences that occur between in the river continuum which extends from headwater (low-order) streams and the higher-order streams commonly known as rivers. For example, Grubauch et al. (1996) refer to the “rapid faunal replacement” that occurs in the mid-order reaches of an Appalachian river continuum; and they cite other studies with similar findings.</p> <p>The proposal to include both large-stream and small-stream (headwater stream, low-order stream) observations in the analysis dataset is not well supported by the logic in the paragraph starting on page 3-31, line 15.</p> <p>The first argument cited by the paragraph concerns Ephemeroptera taxa in large streams and cites Appendix B of US EPA (2011a) which states that Ephemeroptera occur at lesser richness in large streams with elevated SC than in large streams with low SC. This fact, in and of itself, is peripheral to the logic proposed by this document which concerns frequencies of occurrence by reference-site taxa. Appendix B (US EPA 2011) does not document that the relevant Ephemeroptera taxa -- those occurring in larger streams with low SC but not occurring in larger streams with high SC -- are taxa that also occur at reference sites. Even if it did, that additional fact would not provide full support because it does not document that the taxa composition high-SC high-order streams are altered in a manner that exceeds the 5%-of-reference-taxa loss threshold. If both high- and low-order streams are to be included in the analysis dataset, only taxa observed as characteristic of both high- and low-order streams</p>	



**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	<p>should be considered in the analysis. If conducting such analysis, the finding that a given taxon occurring at reference sites is also characteristic of high-order streams, should be based on more than a single occurrence by following the logic of the so-called extirpation concentration defined as the 95<sup>th</sup> percentile of capture probability and not as the maximum SC for observed occurrence. This precaution is justified by these organisms' mobility.</p> <p>The document (paragraph starting on page 3-31, line 15) also states that "conductivity and drainage area are very weakly correlated" within the areas studied. This fact is not of direct relevance to the argument that biological data from rivers and headwater streams should be mingled within datasets that are analyzed using the methods described. The use of "background" SC (i.e. 25<sup>th</sup> percentile) to approximate reference condition for large streams does not help the logic, in my view; "background" SC and "reference condition" are different concepts, with the "reference condition" concept as more restrictive.</p> <p>The document also states that "Inclusion of the data from large streams did not significantly change the magnitude of the HC<sub>05</sub>". That statement is supported by citing Suter et al. (2011), which is a conference presentation and not a peer-reviewed manuscript that is accessible to reviewers, potential regulatory commenters, etc.</p> <p>Most importantly: The method proposed by this document is novel, as admitted by the authors. However, when applied to headwater streams in coalfield areas, it is being applied in a context where numerous studies have found altered benthic macroinvertebrate communities in low-order streams influenced by major ions discharged by coal surface mines; and no peer-</p>	



**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 2: Catchment Size:** All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	reviewed studies I am aware of have found the opposite to be occurring. A comparable body of supporting science does not exist for the higher-order, high-drainage-area streams.	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 3: Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	Annual samples, particularly those collected in summer, likely miss some or many sensitive insect genera. However, as long as sampling time is NOT correlated with conductivity (e.g., sampling high-conductivity sites early, but low-conductivity sites later), this source of error is probably minor compared with other sources, such as selection of sites, sampling variability, and the temporal variation of conductivity. I would examine the correlation between site conductivity and sampling date (Julian Day).	
Reviewer 2	<p>The data of conductivity concentrations show that they do vary by season. This was addressed by comparing hazardous concentration values by season. “Due to the similarity at the low end of the sensitivity distribution (SD) between spring HC<sub>05</sub> and HC<sub>05</sub> of the full dataset” (pg.4-11), it was determined to use all data regardless of month. I question why this wasn’t also done for the fall (especially October) data? Granted, February – April exhibited the most noticeable change but October was significant as well. In Ecoregion 70 from the Watershed Assessment Branch database, September stood-out because it had significantly higher conductivity values (pg. 5-7), as did April with definitely lower values (pg. 5-8), although not as extensive as October’s. The box plot on pg. 5-9 for Ecoregion 70 shows the apparent seasonal variation of July – October.</p> <p>Reference sites, however, are stated to have conductivity levels “generally low and similar throughout the year although slightly higher in August, September and October,” (pg. 5-12). I think it is more than just “slightly” higher! On pg. 5-7, Ecoregion 70, the September values are so much higher that it is difficult for me to understand that the September data doesn’t adversely skew the results. Perhaps separate CCC and CMEC for the September timeframe is reasonable. Since my area of expertise is not statistics, I am not really able to investigate this myself and will rely on the authors’ determination that seasonal differences do not require weighting, and that the seasonal differences do not alter the results to any great degree.</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 3: Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	<p>On pg. 5-6, the conductivity background for Ecoregion 70 is &lt;200 uS/cm December – June, and &gt;200 uS/cm July – October. This seems to be enough of a distinction that perhaps all data need to be divided into two sets, one containing the December – June data and the second, the July – October data. It would seem that this would be sufficient rationale to have this separation but I am presenting this more as a question than a statement.</p> <p>As a side personal note: Here in the Midwest we have distinct seasons, and many parameters clearly show this in their values. I am accustomed to looking at the seasonal data and its use in planning for monitoring programs and watershed recovery plans. Having this specificity of data is more informative for these purposes than “lumping” or weighting of the data because it provides greater insight as to pollutant sources and causal relationships. For state staff, determination of sources of impairment is usually the overall objective and is frequently difficult to ascertain. Having a clear understanding of what is happening each month (when there is monthly data available) helps to provide insight. With that noted, I fully realize the objectives for those purposes and the objectives for this study are different. But it may be of value to the authors to understand how state WQ staff usually look at data and use it.</p> <p>Additionally, with these comments in mind, I must also add that I prefer limiting the amount of weighting when working with a dataset. On pg. 3-18 it is stated that if “the weighted HC<sub>05</sub> overlap the confidence bounds of un-weighted HC<sub>05</sub>, the un-weighted model is accepted.” This seems to be a logical and accurate decision. Further, it states that in general, “the use of unweighted SDs is easier and requires fewer data points.” I agree. Where weighting and manipulating the data can be reasonably minimized, I believe it should be. A balance must be made in the need for normalizing, scaling and weighting and the loss of variations that reflect the actual conditions.</p>	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 3: Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	Also pgs. 3-16 – 3.18, the three approaches to seasonality are given. It is well done.	
Reviewer 3	First, see part of the response to question 1. I would be careful including data for extreme stream flow conditions, which may occur in spring (high flows), summer (possible hurricanes), and fall (drought). Care should also be taken to examine any unusual antecedent conditions within watersheds to be studied. In our regional work, we needed to delete a few 1 <sup>st</sup> and 2 <sup>nd</sup> order sites due to extreme high flow conditions in the previous year that affected two subwatersheds in our study area.	
Reviewer 4	Since the analyses indicated the effects of seasonality are minor, I have no problem with how the paper dealt with this issue. Given the pattern in conductivity in some of the datasets where there are higher values in the late summer (e.g., August-September), a period that corresponds with typical lowest monthly flow periods, it might be of use to discuss how this might influence monitoring for compliance with any derived criteria. For example, the paper talked about a monthly weighting of conductivity values to determine the effect of seasonality on the criteria. If a State only collects data during a summer period (e.g., Aug/Sep) should the values be adjusted to the annual geometric mean to determine whether benchmarks are exceeded?	
Reviewer 5	I do not agree with the “not weight for season” decision. Research at Virginia Tech (Boehme 2013; Boehme et al. 2013) has demonstrated that composition of benthic macroinvertebrate samples from coalfield headwater streams varies seasonally, both in reference streams and in streams with elevated TDS originating from mining sources. Other research demonstrates seasonal differences in response by a multimetric index to contemporaneous SC in both reference streams and those affected by elevated SC/TDS (Timpano et al. 2011), meaning that community composition differs by season. Also, the document itself demonstrates clearly that SC in non-reference streams varies by season (Figures 4-2, 4-4, 5-2, and 5-4). Hence, I do not see scientific support for analysis using	

**TABLE 2. RESPONSE TO CHARGE QUESTIONS 1-3: DATA SET CONSIDERATIONS**

**CHARGE QUESTION 3: Seasonality:** The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.

REVIEWER	COMMENT	RESPONSE
	<p>methods described in the document of data sets that mingle samples from different seasons without a seasonality check, such as a check to determine if limit-defining taxa are seasonal.</p> <p>Section 3.14 describes a seasonality check procedure that compares spring, summer, and “all year” samples. That section defines spring as March – June. Elsewhere, the document describes “summer” as July – October, an unorthodox definition of that season. Does that definition of “summer” also apply in Section 3.14, and in the Case Study 1 and 2 seasonal analyses (Figures A-7 and B-7)? Seasonal definitions should be stated clearly.</p> <p>Seasonal HC<sub>05</sub> values were developed for Case Studies 1 and 2; spring and summer HC<sub>05</sub> values are similar for Case Study 1 (Figure A-7) but not for Case Study 2 (Figure B-7). On page 5-12 (Case Study 2), the document states: “In the final assessment, due to the similarity at the low end of the genus sensitivity distribution (SD) between the spring HC<sub>05</sub> and the HC<sub>05</sub> based on the full data set, the example ecoregional criteria were derived using all available data, regardless of the time of year they were collected.” Based on Figure B-7, I do not see the seasonal HC<sub>05</sub> values as similar.</p> <p>In conclusion, I see no justification for a procedure that would mingle data from all seasons with no seasonality check or adjustment for the data’s seasonal distribution. The Case Study 2 results justify the need for consideration of season. The fact that both community composition and water quality vary by season demonstrate that seasons should be considered separately in HC<sub>05</sub> development.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 4: Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).

REVIEWER	COMMENT	RESPONSE
Reviewer 1	<p>The method of deriving HC<sub>05</sub> is straight forward and clearly described. However, the method used to estimate the extirpation threshold (XC<sub>95</sub>) is confusing and problematic. XC<sub>95</sub> considers neither the direction of response of a genus to increased conductivity nor the relative frequency of the taxon (“probability of capture” in the text), two key factors for inferring extirpation. Therefore, it appears not possible to establish any consistent and meaningful relationship between XC<sub>95</sub> of a genus and its extirpation. The authors use a GAM model to refine XC<sub>95</sub>. That is helpful for those genera negatively affected by conductivity, but the threshold of extirpation for those genera positively or neutrally responding to increased conductivity over the range observed still remains indefinable. For example, XC<sub>95</sub> of <i>Cheumatopsyche</i> (A-29) is estimated to be &gt;3140µs/cm (A17), while the genus reached its highest “probability of capture” at this conductivity level. Even with a qualifying designation of “&gt;”, is this estimate really meaningful? The same designation (&gt;) is also given to those genera that have very different response curves, such as <i>Cheumatopsyche</i> and <i>Leuctra</i> in Fig. 3-1. When the values of XC<sub>95</sub> for genera that substantially differ in occurrence frequency and response to conductivity are treated equally, the SD curve is no longer interpretable and potentially misleading, at least in my opinion. Two options might be worth considering. First, presumably one can appropriately determine extirpation thresholds (i.e., XC<sub>95</sub> without &gt; designation) for more than &gt;10% of the genera. If so, he/she may put all other genera in a single category, “indefinitely high”. The authors may then use the first group of genera to define HC<sub>05</sub>. Second, the authors can look at how many genera declined down to ≤1% of the highest “probability of capture” in the max-conductivity bin in GAM models. If more than 10% or 20%, as in their case studies, they should be able to easily determine HC<sub>05</sub>, leaving out the idea of XC<sub>95</sub> entirely.</p>	
Reviewer 2	<p>As my knowledge base is centered on biological aspects of rivers and streams rather than statistics, several of my comments will be limited in this regard. I am listing the various thoughts which I had as I went through Section 3.1:</p> <ul style="list-style-type: none"> <li>• The inclusion of both high quality and impaired sites is correctly done. This provides a well-represented database, covering all levels of conditions and taxa, and</li> </ul>	



**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 4: Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).

REVIEWER	COMMENT	RESPONSE
	<p>at all times of the year. This reflects the variability that will exist because of seasons, habitats, and the effects of manmade influences in the river basin which affects the ionic composition.</p> <ul style="list-style-type: none"> <li>• The step-by-step explanation on pg.3-1 is helpful. More of this could be done to increase understanding of the calculation processes used in this document.</li> <li>• Improve clarity by explaining how the actual weighting and cumulative distributional function is done on pg. 3-1.</li> <li>• Good explanation on pg. 3-2.</li> <li>• Figure 3-2, pg. 3-4: Gives good general process flow. Is it possible that an actual mathematical example could follow along with each step?</li> <li>• The bullets on pg. 3-5 are thorough and give good support to adequacy of data. Sample size discussion is well done. Sensitivity analysis, which includes a representative proportion of sensitive genera, is well done. Having 90-120 genera and 500-800 sites are large numbers, and are seen throughout this document. This is excellent. It strengthens the development of the criterion, its applicability, and the justification of the concentrations determined. If only all studies could have such numbers!</li> <li>• Bootstrapping needs to be described more fully (for non-statistical readers). While the paragraph on pg. 3-7 is probably adequate for many, there are a considerable number of state agency or other watershed staff who have minimal statistical backgrounds. A few additional paragraphs detailing/giving examples of such exercises as bootstrapping would make the document more usable by the large range of agency staff.</li> <li>• In reference to pg. 3-9, lines 9-21, care must be taken to avoid too many repeated macroinvertebrate samplings in the same place over the course of a year. Repeated sampling is disruptive to the habitat and can diminish the taxa at the site. Unlike fish species, macroinvertebrates are less mobile, and, if young stages are removed, there may be fewer adults at the sites especially if there are no other small streams in the vicinity to repopulate.</li> </ul>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 4: Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).

REVIEWER	COMMENT	RESPONSE
	<ul style="list-style-type: none"> <li>I would like to see a little more specificity in describing sampling methods. Is there assurance that there was a standardized field sampling protocol observed for all biological sampling? It is important that all sampling crews used the same techniques. It is more of a problem between jurisdictions (states, cities, or private organizations which do monitoring) but can also occur within an agency. It is vital that, for example, an equal number of sweeps of the catch net are made at each site, or, the same number of individual samples comprise a composite.</li> <li>I see that my thoughts in the above bullet is addressed on the next page (pg.3-16), lines 8-15.</li> <li>The use of different protocols by different organizations and agencies is a very real concern to any large database that has merged several smaller data sets. It probably is one of the biggest and most pervasive problems. The importance of initial training, repeated review throughout the monitoring season, and dedicated adherence to the field sampling quality control document can't be overemphasized. The authors have (gratefully) recognized this problem and have provided how to address this: by comparing all-year HC values from one region to that of another comparable region. If the datasets have a large number of data points, I believe this would be an acceptable way to handle this.</li> <li>I'm not sure that I fully understand the third approach to seasonal variability in Section 3.1.4 (Assessing Seasonality, Life History, and Sampling Methods). However I do believe that the authors have done well in going step-by-step in their presenting of the third approach.</li> </ul>	
Reviewer 3	I really like this approach, since these are well developed exposure-response relationships at the genus level, assuming that any species within the genus would share a similar response (well-known for many fish genera exposures to numerous stressors). The entire sequence of CCC analysis and the derivation of the CCC for conductivity are very well-presented in Figure 3-2, as well as in the text. Also, the example in Figure 3-1 is good, giving the reader an example of how to derive the HC <sub>05</sub> of a genus sensitivity index – not a particularly easy concept to grasp unless one has some background in bioassay statistical techniques.	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 4: Criterion Continuous Concentration (CCC):** Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).

REVIEWER	COMMENT	RESPONSE
	One analytical - statistical comment: There have been a series of papers in recent years by King and Baker who use the TITAN model to examine stressor relationships with biota. It may be beneficial to explore this model to estimate conductivity-response as a check on the CCC.	
Reviewer 4	I generally found the approach to derive the XC <sub>95</sub> and HXC <sub>05</sub> relatively easy to understand, with perhaps a more step-by-step on how to calculate the weighted CFD values. Was this done using Excel, R, or some other application?	
Reviewer 5	<p>The methods for deriving the values are described clearly, especially when viewed in association with the examples presented.</p> <p>However, it is not quite clear what the CCC is intended to be within the context of a potential regulatory program. The CMEC is described as the maximum concentration likely to occur at a site where water quality satisfies the CCC 90% of the time, yet the CCC is also described elsewhere as a geometric mean. Which is it?</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 5: Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	Confidence levels (e.g., 90% or 95%) can be estimated only if the frequency distribution of data is known (e.g., normal). Did the authors check the data distribution before using Eq. 3-2? Is the critical value used here for normal distribution? If confirmed, the method is reasonable.	
Reviewer 2	<p><b>a) Does the analysis to derive the maximum exposure concentration (with temporal resolutions) characterize the maximum concentration that will result in meeting the CCC 90% of the time?</b> I can only provide comment in a limited manner. The annual geometric mean is appropriate for comparing different values and finding a central tendency or typical values for a set of numbers. It normalizes the ranges and removes the effect of large differences so that no one particular range of values dominates the weighting. This is appropriate for the intent of the calculations in this section/document. However, because I am not proficient in this, I am less sure of the maximum condition at any given station can be established by incorporating among-station and within-station variability. To achieve this, wouldn't the sampling sites and their particular data points need to be central in tendency and not exhibit values at the further reaches of the ranges? How was the 90% determined – review of that for the reader would be helpful.</p> <p><b>b) Is it reasonable to expect it to be a protective upper limit for sites in the data set?</b> Yes, I think it is appropriate for determining the upper limit. 90% is definitely a protective level. Indeed, there will likely be certain interests in watersheds who will contend that this is too stringent. However, based on the sensitive genera and maximum exposure concentrations found in this document, the data (and thus, the rationale) for establishing these levels is very strong and definitive. Using the paired analyses (daily measurements of conductivity paired with macroinvertebrate sampling) is a very strong statistical test and widely used in biological and environmental studies.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 5: Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

REVIEWER	COMMENT	RESPONSE
	<p><i>c) What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?</i> As I have mentioned the annual geometric mean and paired analyses are strong points. The subset of frequently sampled sites is a critical element. It would seem to me that it would be important that these are clearly representative of the majority of the sites, or does the annual geometric mean make this an unnecessary concern?</p> <p>The sampling of at least six times is also an important feature. I fully support the use of six times per year per site. I would increase the <i>n</i> to two in the spring (March – May) and two in the fall (Aug – Oct) and leave the remaining two for one in the summer and one in the winter. Greatest changes occur in the spring and fall months and therefore each warrant another sampling event to help capture this variability. Even with six samples, standard deviation will likely be high, especially if there are considerable differences between the sites, and even within the sites if weather, etc. are quite variable. Lastly, as much as one would like to have repeat sampling six times/site, it is often beyond the budget of many state 305(b), 303(d), and TMDL programs. Perhaps federal support can be made available for state criterion development.</p> <p>The flow chart in Figure 3-6 is helpful for overall process steps. But perhaps a working example of this could be placed in the Appendix and referenced here. I think having an example would be especially helpful to state water quality staff.</p> <p>In keeping with the above, I would suggest greater description of LOESS and a full example. Although such processes as LOESS and bootstrapping are familiar to tacticians and to those</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 5: Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

REVIEWER	COMMENT	RESPONSE
	who conduct these analyses regularly, many workers in the field of water quality programs haven't as much familiarity.	
Reviewer 3	Similar to my comments for Question 4, the sequence for determining the CMEC is very well described in Figure 3-6. I feel that the derivation of both the CEC and the CMEC are very robust, in part because of the availability of rich ecoregion data sets. I like the fact that there is careful trimming of the data set, followed by examining for unequal variances and for estimating Type I errors (there are often models published that do not perform these simple tests).	
Reviewer 4	This approach seems reasonable, however it seems that further empirical analyses of the consequences of this approach would be useful. For example, for sites that are achieving some biological benchmarks (e.g., IBI, ICI) what is the frequency that these are considered impaired based on the CCC and/or CMEC? Again my concern with single criteria rather than tiered criteria has some consequences with use of both of these benchmarks. An example of using tiered criteria and calculating CCCs and CMECs for both would be useful.	
Reviewer 5	<p>The logic for the CMEC derivation (Section 3-2) is not presented. Where did this equation come from, and where is the supporting logic? Has the validity of the proposed approach been checked using laboratory bioassays, or with any other method that uses measured data? If so or if not, that should be stated clearly.</p> <p>The assumption underlying the CMEC calculation appears to be that the CMEC is defined as a maximum concentration that is likely to occur at a site that satisfies the CCC (estimated as the HC<sub>05</sub>) 90% of the time. If one wishes to estimate a maximum concentration that is likely to occur at a site that satisfies the HC<sub>05</sub> 90% of the time, one must know the temporal distribution for the target variable – SC in this case. It appears that the CMEC equation has been derived assuming that SC will vary in time independently and as a normal distribution.</p>	



**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 5: Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

REVIEWER	COMMENT	RESPONSE
	<p>Has this been demonstrated with field data? Others have noted that water quality data rarely vary normally and are often autocorrelated (Helsel and Hirsch 2002, see Chapter 12 for temporal analysis).</p> <p>The proposed site selection procedure for the CMEC derivation is not adequate. The proposed procedure requires:</p> <p>At least 6 samples over a given year. A minimum of one sample in the spring (low conductivity, March–June), and one sample in the summer (high conductivity, July–October) are included to capture temporal variability.</p> <p>Desirable changes are:</p> <p>To remove the specific date designations from the second bullet, if the document goes forward with an intent for national application. Certainly, both high-concentration and low-concentration periods should be represented; but these periods may vary by time of year among regions, and among years (based on climate variability) for any given region.</p> <p>To add an additional criterion: that remaining samples should be evenly distributed throughout the year. If remaining samples are clustered within a given time of year, they will not be representative of the SC variability that occurs throughout the year; and, hence, would not be suitable for estimating a CMEC using statistical procedures.</p> <p>Also concerning CMEC: What are the units for the Y axes for Figure 4-9 and 5-9? Presumably, the Y axis (standard deviation) is expressed as log<sub>10</sub> SC, is that right? Whatever</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 5: Criteria Maximum Exposure Concentration (CMEC):** The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?

REVIEWER	COMMENT	RESPONSE
	<p>it is, it should be stated either in the axis label or in the figure caption. Also: If I understand the axis correctly, those numbers look quite low to me – I suggest they be checked.</p> <p>Also, the CMEC concept is not clearly defined by the document. For example, page xviii (Executive Summary) states “Below the CMEC, sites are expected to meet the CCC 90% of the time; i.e., a conductivity level that is protective of acutely toxic exposures for 95% of macroinvertebrate genera.” This sentence is not written correctly. Similarly, the Glossary defines the CMEC as “In this document, the CMEC is the conductivity level at which the CCC is met 90% of the time.” I think I understand what is meant by these sentences, but the language is not clearly stated.</p> <p>As an overall comment: I find that logic that underlies the CMEC as thinly supported, considering that its purpose is regulatory program development. The logic being applied here is statistical, not biological; and no biological data are presented as confirmation of results derived from statistical analyses.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 6: Duration:** Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The description and justification appear adequate. I am not aware of additional publications.	
Reviewer 2	<p>This approach relies “directly on paired in-situ measurements of conductivity and benthic macroinvertebrate assemblage composition,” pg. 3-20, a very reliable analysis test. Macroinvertebrates are indeed exposed to quite different conductivity levels throughout the year.</p> <p>The authors are quite correct that with only annual sampling, “it may be difficult to determine precisely how long conductivity levels can be above the CCC before extirpation...” (pg. 3-21, line 12-13) occurs. I would say it is <u>most</u> difficult and next to impossible to tell from one sample. Sampling only once is the reality, however, of many state bioassessment sampling programs. Nothing is better than having repeated (in the field) sampling for each site. Depending upon only one sample per year is what state programs would like to avoid but in many cases, it is all that they have. So from this standpoint, the approach seems to take this into consideration and makes sense. Lastly, lines 12-13 appear to support the argument of using only one sample/year as the basis to determine duration of CCC and CMEC. In general I believe that the authors have worked hard to provide adequate description and justification for the duration of the CCC and CMEC. The description and justification for the approach on pg. 3-22 to 3-23, line 1-16 and lines 1-16 is excellent. This is very well done.</p> <p>On a side note, is there a tag or footnote which could indicate that a data point(s) represents only one sampling per year? This would distinguish it from mean values from sites which have multiple sampling times during a year, thus allowing for all data to a dataset to be used, and yet allow the reader to know that some data are single data points and others are mean/geometric means. Seems this would be in the best interest for states wherein multiple databases are being used for criterion development or even just a single database which has some sites with only one sampling per year and some which have multiple samples. It is preferable to have more samples when possible, but it would be easy for state budget-cutters</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 6: Duration:** Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

REVIEWER	COMMENT	RESPONSE
	<p>to limit sampling to just one sample per year if that is all it takes to establish criterion development. “Why sample more if only one is needed?”</p> <p>Further, those interests who oppose water quality criterion in general (“infringement on private property rights”, “over-regulation for agriculture”, “costly programs for cities”, ...) would use the “only one sample per year” to justify their opposition to the criterion’s validity. The argument will be that there isn’t enough data and therefore the criterion is not based on “good science.”</p>	
Reviewer 3	Description and justification are more than adequate to support both CCC and CMEC. I am always a little leery about a CCC (or any water quality criteria that is based on a yearly value), since Figure 3-7 does illustrate very well the potential for large yearly variations in stream conductivity. In one of my forested study sites, conductivity may range from 75-100 $\mu\text{S}/\text{cm}$ in the spring to over 600 – 700 $\mu\text{S}/\text{cm}$ in late summer - early fall due to the dynamics of stream flow and forest transpiration.	
Reviewer 4	States like Ohio and localities such as the MSDGC (collected by MBI) commonly collect biological data paired with one or more weekly continuous regimes of conductivity data (e.g., Datasonde collectors). It seems that some of these sites can be used to examine duration questions in more detail. Such datasets have hourly values of conductivity collected over 7-10 days, once or twice a summer.	
Reviewer 5	<p>Section 3.3 discusses studies that are relevant to the duration question, but none of those studies address the question directly. I am not aware of relevant studies other than those discussed by the document.</p> <p>Answering this question would require continuous monitoring of water quality in association frequent benthic macroinvertebrate sampling, such as the data described by Boehme (2013), Boehme et al. (2013), and Timpano et al. (2013); but I am not aware of analyses by these or other authors that address this question directly.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 7: Frequency:** Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The recovery of macroinvertebrates strongly depends on nearest sources. If the exposure occurs at a local scale, three years may be enough for re-colonization. However, if the exposure happens at some broad scales, three years may be not enough.	
Reviewer 2	<p>This is one of the best sections in the document! Descriptions and reasoning are exceptionally well done throughout this entire portion. The details and thoroughness are reflective of excellent biological knowledge on the part of the authors.</p> <p>I only have a couple of comments: First, I am surprised at the high level of conductivity, &lt;960 uS/cm (pg. 3-27, line 10), before extirpation of sensitive crustaceans. This seems exceptionally high. As a general rule, crustaceans, and mollusks specifically, are front line indicators of contaminants and water quality pollutants. Because water passes through them, low pH, chemicals, and excessive suspended solids and siltation are known to affect them significantly and earlier than many other aquatic organisms.</p> <p>Secondly, I would have liked to have seen consideration given in the causal assessment methodology (Sec. 3.5, pg. 3-28, lines 15-20) of the relationship with “other known stressors such as metal toxicity, streambed erosion and siltation, and eutrophication.” These conditions do contribute as stressors, often co-exist during times of high conductivity, and seem to compound effects. I know from experience that during rain events and urban stormwater runoff (with increased suspended solids and accompanying high turbidity values), that conductivity also can substantially increase. A causal relationship seems to me to exist between increased turbidity and increased conductivity. This is not really addressed in the document. Do the increased conductivity values during rain events come exclusively from ions associated with concrete weathering, industrial runoff, fertilizer runoff, or, is the increase also coming from the suspended eroded soil particles (and their attachments)?</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 7: Frequency:** Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?

REVIEWER	COMMENT	RESPONSE
Reviewer 3	This section is well written and uses two classic papers (Niemi et al. and Wallace) to illustrate recovery rates in benthic organisms (insects primarily) from stressors. Recovery in stream fishes is not as clear since there may be multiple physical stressors that create long-term problems after water quality remediation (e.g., AMD), especially for lithophilic spawners. Generally, this section is highly supportive of CCC and CMEC.	
Reviewer 4	I think the discussion of the estimation of frequency is reasonable (not to be exceeded more than once in three years on average), but perhaps can be supplemented by some ambient analyses as another form of evidence. One suggestion might be to derive “biological stressor metrics” using the XC <sub>95</sub> values. For example, I have used the most sensitive 15th percent of conductivity weighted mean values by taxa to determine taxa “sensitive to conductivity.” For each site one can then generate the number of conductivity sensitive taxa present which can then be used to provide evidence that the count of sensitive taxa varies with conductivity as predicted under various duration, frequency and magnitude scenarios. This can also be used to compare potential tiered use responses to conductivity.	
Reviewer 5	Comment similar to the above response to Question 6.	



**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 8: Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	It is clear and adequately supported. This alternative endpoint explicitly identifies the stressor anions. However, it does not account for any other anions, which may be less abundant, but still significant, such as $\text{Cl}^-$ in many freshwaters. As a result, the criterion derived would be less applicable than conductivity-based criteria. In addition, this alternative is subject to the same criticisms I made early on estimating $\text{XC}_{95}$ and $\text{HC}_{05}$	
Reviewer 2	<p>In general and to the best of my understanding, yes, I believe the use of an alternate measurement endpoint is written reasonably clearly and is adequately supported. In instances where I felt more description or clarity is needed, I have listed it. As in Question #4, I am going to simply list individual comments which I noted as I went through Appendix F:</p> <ul style="list-style-type: none"> <li>• The correlation value for conductivity with the two ions is exceptionally tight (Figure 1) and provides excellent data justification for their use.</li> <li>• There are an exceptionally large number of paired samples (pg. F-4, line 12)! If only all studies and monitoring programs could have such a large dataset. Distribution of sampling sites was also excellent. The large background data set and the wide range of conductivity throughout the sampling area indeed allows for sound characterization of the extirpation concentration.</li> <li>• I believe that seasonal variation is less with these two ions. Except for September and December, there is greater similarity in their monthly values (pg. F-10, Figure F-5) than with the four ions. However, the table on the next page shows considerable variability. Nevertheless, the text says there was enough similarity on the low end of the genus sensitivity distribution to allow for criterion development. Is comparison of just the low end of the sensitivity distribution adequate? Certainly avoidance of extirpation for the most sensitive of the genera is the “goal” of the criterion, but do concentrations for moderately sensitive species, or, the extent of the ranges, have some role and should be discussed?</li> </ul>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 8: Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

REVIEWER	COMMENT	RESPONSE
	<ul style="list-style-type: none"> <li>Two areas of which more description and information might be helpful for the reader:               <ol style="list-style-type: none"> <li>pg. F-13, Figure F-6: Advantages of using log 10 to weight values</li> <li>LOWESS – pg. F-16, lines 3-6</li> </ol> </li> </ul> <p><b>The second part of Question #8:</b>            There is a very close correlation with the two ions. They are prevalent and widely distributed in the ecoregions. They have similarity on the low end of the genus sensitivity distribution – thus functioning in the statistical analyses similarly to the four ions.</p> <p>However, disadvantages of the two ions might include:</p> <p>Measurement of individual ions is more costly and time consuming and most sampling/monitoring programs measure for conductivity routinely, even the installed in-field monitoring instruments can give continuous readout on conductivity value. Conductivity measurement is easy, quick, and inexpensive. The four ions comprising conductivity measurements are equally as widespread in distribution or perhaps more so than the two ions. Most monitoring programs only do conductivity because of limits on budgets. Are the four ions less affected by low pH values?</p>	
Reviewer 3	<p>Just a very general comment to start with in this response to question 8. In our water quality laboratory, we generally do a complete cation and anion scan since these are easy on an ion chromatograph. Although primarily interested in Ca and Mg, we also analyze for K and Na, and have found these to be important cations in some streams. The anion scan is important in that it also gives a few other ions that appear sometimes in our study streams, although we do nutrient scans on other instruments due to sensitivity and detection limits.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 8: Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

REVIEWER	COMMENT	RESPONSE
	So, based on the discussion in Appendix F, I would be very comfortable with using the alternative measurement endpoint, but only as a last resort if the water quality data is not adequate for a data set (meaning no measurement of Ca, Mg or Cl). There is not much difference in the slopes of Figure F-1 (c) and F-1 (d), although there is less scatter in F-1 (d) with the addition of Cl. (Note: why wasn't a test for equality of slopes performed or did I miss it somewhere in this section?).	
Reviewer 4	The correlation between conductivity and the alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) is so strong that I think most users (e.g., States) will focus on conductivity given its cost and ability to cheaply monitor it continuously. Because of this I did not analyze this as closely as some other parts of the report, but it seems to result in a similar type of benchmark.	
Reviewer 5	<p>The alternate endpoint is clear and is supported by the scientific information that is available at this time. However, additional investigations are warranted as a means of thoroughly documenting the appropriateness of <math>[\text{HCO}_3^- + \text{SO}_4^{2-}]</math> as a biotic condition indicator that would provide information other than that which is provided by SC.</p> <p>It is clear that <math>\text{HCO}_3^-</math> and <math>\text{SO}_4^{2-}</math> are the two dominant anions in most Appalachian coal-surface-mine influenced waters (Pond et al. 2008; Timpano et al. 2011; Agouridas et al. 2012; Pond et al. 2014; Sena et al. 2014). Since numerous studies have found elevated SC to be closely associated with benthic macroinvertebrate community alterations and taxa losses, it seems quite reasonable to use <math>[\text{HCO}_3^- + \text{SO}_4^{2-}]</math> as a measurement endpoint – although no more reasonable than use of SC itself. However, I see this relationship as a reflection of the geochemical processes that drive ion release from the mine spoils and not necessarily as a causative indicator. For that matter, the sum of <math>\text{Ca}^{++}</math> and <math>\text{Mg}^{++}</math>, which are typically the two dominant cations, could also be used as a measurement endpoint to the same effect.</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 8: Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

REVIEWER	COMMENT	RESPONSE
	<p>I do not see support for an argument that <math>[\text{HCO}_3^- + \text{SO}_4^{2-}]</math> would be a “better” endpoint than SC; I do not presume it to be more or less representative as an indicator of the “actual toxicant” because the actual toxicant or toxicants is/are unknown.</p> <p>I find the role of <math>\text{HCO}_3^-</math> in the observed phenomena to be quite puzzling. <math>\text{HCO}_3^-</math> is often elevated (relative to ecoregion 69 reference levels) in an adjacent ecoregion (Ridge and Valley, 67) due to natural conditions. Therefore, why would similar <math>\text{HCO}_3^-</math> levels contribute to benthic macroinvertebrate impairments in the coalfields? Are the taxa that different? Is it possible that the ratio of <math>\text{HCO}_3^-</math> to other ions present acts as an ecotoxicological influence? The fact that Mount et al. (1997) found <math>\text{HCO}_3^-</math> to be more directly associated with lab-test organism toxicity than most of the other ions studied does support a potential ecotoxicological role for <math>\text{HCO}_3^-</math>. However, Mount et al. (1997) also found <math>\text{Mg}^{2+}</math> to be associated with those toxicities; and scientific literature (e.g. Pond et al. 2008 and 2014, and other studies) demonstrates that <math>\text{Mg}^{2+}</math> is also quite elevated in mining-influenced high-SC streams; and <math>\text{Mg}^{2+}/\text{Ca}^{2+}</math> ratios are often altered in mining.</p> <p>I suggest that EPA investigate benthic macroinvertebrate status in mine-influenced <math>\text{SC} &gt; 300</math> <math>\mu\text{S}/\text{cm}</math> waters where <math>\text{SO}_4^{2-}</math> concentrations are quite low and <math>\text{HCO}_3^-</math> is the predominant anion to determine if biotic condition in such waters is consistent with expectations based on studies to date. The current proposed document would apply <math>\text{HC}_{05}</math> levels as criteria in such waters but it is not clear that such inclusion is warranted. Such waters have not been represented in any of the existing studies that associate SC levels with biological effects. Mine-spoil leaching studies (Agouridis et al. 2012; Daniels et al. 2013; Sena et al. 2014) indicate that <math>\text{SO}_4^{2-}</math> concentrations decline with progressive leachings more rapidly than SC/TDS, suggesting that <math>\text{HCO}_3^-</math> remains as an important solution component and becomes the dominant anion. Hence, one would expect effluents from aging mine-spoil fills constructed</p>	

**TABLE 3. RESPONSE TO CHARGE QUESTIONS 4-8: EXAMPLE CRITERIA CALCULATIONS**

**CHARGE QUESTION 8: Alternate measurement endpoint:** Is the example alternate measurement endpoint ( $[\text{HCO}_3^- + \text{SO}_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?

REVIEWER	COMMENT	RESPONSE
	with non-pyritic spoils to approach a condition: SC/TDS remains elevated and $\text{HCO}_3^-$ concentrations are also elevated (relative to reference) but $\text{SO}_4^{2-}$ concentrations have declined substantially from the elevated levels that characterize leachates from fresh mine spoils to a concentration much lower than $[\text{HCO}_3^-]$ . To my knowledge, biological effects of such waters have not been studied. If my understanding of mine spoil geochemistry is correct: Frequencies of occurrence by such waters will increase with time as the existing stock of mine-spoil fills that have been constructed throughout central Appalachia age and their leachate chemistries change.	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	<p>The process is clearly described, but it does not seem to fully address the concerns of SAB. To apply the conductivity criteria (HC<sub>05</sub>) developed for one region to another, we need to make sure both water chemistry and macroinvertebrate fauna to be comparable. The authors did a good job for assessing water-chemistry comparability. However, their responses to SAB's other comments do not seem adequate. The authors are correct that SD does not require the same set of genera. However, it does require that the distribution of XC<sub>95</sub> (at least for sensitive genera, tolerant ones are not used anyway) in the new area is similar to in the original region. Say, two regions share all genera, but if the new region happens to have more highly-sensitive genera that meet the minimum sample size (25 samples) than in the original region, its HC<sub>05</sub>, if derived, may be lower, and thus the original HC<sub>05</sub> would be less protective. The authors need to address the importance of biological comparability.</p> <p>Yes, water quality criteria of EPA established based lab tests is applicable across the nation or multiple regions. This is because a standard lab procedure is used (test species and experimental setting). Here, we do not have a standard set of genera with the same occurrence frequencies and same environmental conditions to derive a universal HC<sub>05</sub>. Extrapolating conductivity criteria to beyond the original region may be risky even if water chemistry is comparable, as I argued above.</p>	



**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
Reviewer 2	<p><b>First part of Question #9:</b> Geographic applicability is approached by the background-matching method. The background conductivity of the original area and the new area should be similar. Also the ionic mixture for the background should be the same. The authors have described the elements of this very well and I believe this will be the most exacting and appropriate way to apply criterion to a new region – a vitally important facet to ensure use by all states and for a conductivity criterion to be widely implemented.</p> <p><b>Second part of Question #9:</b> Yes. The only area I would question would be the SAB’s recommendation that “consideration be given to the species composition of stream communities, which might be different in different states . . .” (pg. D-2, line 15-17). I interpret this to mean a direct, species by species comparison. However, the authors of this document have used a taxonomic sensitivity distribution model which doesn’t do this, rather, it looks at a set of species/genera and how the communities in general respond to a stressor. I believe they have provided satisfactory support for their choice. My tendency towards the SAB’s recommendation is because my experience lies with species’ inventories and direct counts for abundance and diversity as compared to reference streams. This preference also goes back to whether one prefers to “lump” data or “split” data – a long known philosophical debate among biologists!</p> <p><b>Third part of Question #9:</b> Yes, I believe it does. This question is similar to the first part of #9 and I really don’t have anything additional to add.</p>	
Reviewer 3	<p><b>First Comment:</b> I wonder if the general statement could be made that the process would be applicable for any Ecoregion III level imbedded within any Ecoregion II level. This seems logical to me, since the original development of ecoregions was designed to address</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
	<p>similarities in geology (and other parameters) that would translate to similar, but not identical, stream ionic concentrations. In the examples in Section 3.6, the ecoregions are adjoining so there may be a very high probability that stream chemistries may be similar.</p> <p><b>Second Comment:</b> A good test may be to examine two Ecoregion III level watersheds that are not contiguous, or at a large distance from each other, or perhaps two watersheds close to each other and two distant from each other.</p> <p><b>Sidebar:</b> In regard to section 4.1.1, I calculated background conductivity using the Y-intercept method developed by Dodds (for estimating background nutrients in mid-western streams) for some 152 probability-based stream sites that we sampled in Ecoregion 69 over the years. My estimate of background conductivity was 82 <math>\mu\text{S}/\text{cm}</math> and the estimate in the report was 80 <math>\mu\text{S}/\text{cm}</math>. I was pleased that these two estimates were in close agreement, especially since the analytical approaches were quite different. It is not that the Dodds' technique is so great (unless there is an adequate sample size), but similar background estimates indicate that the EPA approach for estimating background conductivity is consistent with other potentially useful statistical techniques.</p>	
Reviewer 4	<p>To derive criteria as proposed, the process for assessing general geographic applicability of conductivity is fine. As I discussed above, derivation of benchmarks under a tiered series of aquatic life use may need to accommodate modifications to the derivation approach. The geographic applicability approach generally compares whether the range/variability in conductivity in background conductivity is similar between regions. I am not sure it addresses conditions where a subset of streams may have uniquely (and predictably) lower conductivity that need to be considered separately. Unfortunately, this is somewhat confounded with accurately identifying "background" conditions, particularly in the Ohio region of ecoregion 70. I think this paper would be well served to be placed within the conceptual framework of the Biological Condition Gradient framework (Davies and Jackson,</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
	<p>2006) and the reference site framework of Stoddard et al. (2006) For example, Appendix D defines “Background conductivity as the range of ionic concentrations naturally occurring in the environment that has not been influenced by human activity.” Several paragraphs later, in weighing lines of evidence, it asks: “Are conductivity values at natural background (least-disturbed) sites similar in the new area compared to the original area?” As defined by Stoddard et al. (2006), least impacted is the best available physical, chemical and biological habitat conditions given today’s state of the landscape. With a naturally occurring measure such as conductivity, this definition can be important. Is the single benchmark or recommended criteria a reflection of “minimally disturbed” (site condition in the absence of significant human disturbance) or of least impacted conditions? Tiered uses allow a State to recognize that a subset of sites may approach minimally impacted and the associated criteria can form a baseline to protect that level of condition. Conversely, if a State has another class of sites with an appreciably higher level of acceptable development across the landscape and these sites are still considered least impacted (and these cannot be managed in a way to reduce the conductivity footprint), then different criteria for certain stressors may be applicable. For nonpoint sources of pollutant the CWA talks about controlling stressors that can be feasibly addressed with best management practices.</p> <p>One way to more closely examine the influence of tiered aquatic life uses and tiered water quality criteria would be construct a number of human disturbance indices or gradients (e.g., Bryce et al. 1999; Wang et al. 2008) and then relate them back to well-founded biological condition gradient exercises that classify sites into six ranges of biological condition based on definitions for ten components of biological condition (Davies and Jackson, 2006). This has been done for many States. It may be that the species that comprise the upper tiers of the BCG (e.g., 1-2) could well be the ones that drive the selection of the XC95 value, and absence</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
	<p>of more tolerant taxa from reference sites would drive the conductivity benchmark lower. Conversely, these species may occur in too few reference sites at lower tiers that represent “least impacted” conditions (e.g., BCG tier 3-4).</p> <p>Appendix C, in describing how to use weight-of-evidence to examine the geographic applicability, did not examine the different aquatic life use streams across this ecoregion (particularly EWH vs WWH). Table C-7 indicates that there were no reference sites; however, biocriteria for WWH streams in this ecoregion is based on the 25th percentile of “least impacted” reference sites (EWH based on 75th percentile of sites statewide). Analysis of conductivity values at reference sites and by aquatic life use would be important evidence for this analyses.</p> <p>Base flow seemingly is an important variable not considered other than in a general way in this appendix. Elevated conductivity values at sites in Appendix C in August and September correspond with the lowest estimate monthly average flows by month in Ohio (USGS ungagged model output). Some more explicit consideration of local flow influences on ionic strength would be useful. It is my experience that local base flows in headwater streams can vary considerably within across this region.</p>	
Reviewer 5	<p>For the most part, the process is clearly described generally, especially when the Chapter 3 description is viewed in association with the Case Study 3 example. However, the process is not fully supported as a reasonable process for regulatory development as described. Certainly, the described process might be used by resource managers in the “new” ecoregion to inform management decisions, but regulatory development would be a different and more serious application.</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
	<p>The process described depends upon a “matching” of background SCs for two regions. The document’s glossary defines the term “background conductivity” as representing the “conductivity for a region that occurs naturally and not as the result of human activity” so that is clear.</p> <p>A deficiency in the Geographic Applicability description concerns the process for determining a background SC when using a distribution of water quality data from a given region. As noted by the document: “the 25th centile is conventionally used ... However, when land cover modification (or other anthropogenic disturbance) is pervasive, selection of a centile lower than 25% may be justifiable.” The fact that the 25<sup>th</sup> percentile is based on assumptions and not on rigorous analysis should be noted. The document references US EPA (2000a), but US EPA (2000a) only asserts that a percentile within the range of 5<sup>th</sup> to 25<sup>th</sup> percentile can be used to represent a reference value without citing supporting studies or analyses. As stated by US EPA (2000a), page 4-8: “Both the 75th percentile for reference streams and the 5th to 25th percentile from a representative sample distribution are only recommendations. The actual distribution of the observations should be the major determinant of the threshold point chosen.” As far as I can tell, the decision concerning which centile to select as a “background SC” indicator is being left to judgment, yet the outcome of this process could be significant as a determinant of the “new” ecoregion’s HC<sub>05</sub> according to the process described here. A decision which centile to select can make a big difference in the resulting background estimate. In regional databases I have available for analysis, the 25<sup>th</sup> percentiles of SC distributions differ from the 5<sup>th</sup> percentiles by multiples ranging from 3 to 10.</p>	



**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 9: General:** Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?

REVIEWER	COMMENT	RESPONSE
	<p>One of the authors of this document has published a peer-reviewed study (Griffith, 2014) that derives 25<sup>th</sup> percentiles SC distributions for ecoregions from throughout the US; that study describes a method summarized in the abstract as “followed EPA methods to estimate reference values” but does not describe the resulting values as “background”. In fact, the author states that “Much discussion exists in the literature as to whether estimates like mine are true estimates of background or at least current reference conditions ...”.</p> <p>The process for assessing geographic applicability of conductivity criteria to a new area is not fully supported. What is missing is a biological confirmation. Are benthic macroinvertebrate communities for the two ecoregions in streams of similar sizes comprised of similar taxa? Do the limit-defining taxa also occur in the “new” ecoregion? A method for evaluating biological data as a means of answering such questions should be described. If background conductivity, background ionic signature, and benthic macroinvertebrate taxa for the two ecoregions are all similar and the limit-defining taxa are present in both ecoregions, it would be reasonable to consider applying an HC<sub>05</sub> value developed for one ecoregion to another.</p> <p>The underlying assumption of the Geographic Applicability process, as described, is as follows: If background SC estimates for two regions are similar, then sensitivity of regional taxa to elevated SC will be similar as well. The document cites no studies to support the validity of that assumption.</p>	



**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 10: Geographic applicability to a new area within an ecoregion:** Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The process described is clear (Section 3.6), and the case study analyses support the limited extrapolation. However, I am not clear why one would not include samples from the whole ecoregion in the criterion development at the first place. Even if one ecoregion includes streams in more than one state, it appears much easier to combine raw data from all states involved, standardize them (e.g., sub-sample size), and then develop the region-wide criteria, than to rely on extrapolation.	
Reviewer 2	<p>The authors have been meticulous in setting the parameters for the study and then clearly describing in this document the process for assessing geographic applicability to locations within the same ecoregion but outside of the parent data sets. First, as they stated on pg. 3-32, most streams in an ecoregion tend to have a similar conductivity regime and ionic composition of dissolved salts. This is generally true, but exceptions do occur, and they wisely caution to have care when applying the example ecoregion criteria to any one particular stream reach. Specific changes in rock composition or feeder streams with springs can alter a particular reach. Good job in recognizing this.</p> <p>Regional background conductivity is defined well on pg. 3-33, lines 7-13. Continuing, they clearly point out that for a data set from one geographic area to be applicable to another similar area, there needs to be: a) similar background conductivity levels and ion composition, and b) a comparison of the confidence intervals of the background data set of the new area to those of the original area; confidence bounds for background estimated from the example criterion data set overlapped with the confidence bounds for the background of the rest of Ecoregion 69. The weight-of-evidence assessment for applicability of criteria to the new area adds much to the soundness of the approach. This validation of background specifically for the Ohio portion of Ecoregion 70 was done with a weight-of-evidence. A weight-of-evidence process is something which state staff are accustomed to doing and thus can identify easily with this and conduct it.</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 10: Geographic applicability to a new area within an ecoregion:** Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
	<p>Excellent discussion on pg. 3-38, lines 14-25, of causes of and considerations when the background conductivity is greater in the new area than in the original area. I appreciated seeing such a complete listing; well thought-out and applicable. The summary of 3.6.3.4 is also well done. The discussion on pg. 5-18, Section 5.3 showed further rationale for the reliability of this process: using first through fourth-order streams (thus maintaining some uniformity of catchment size), extensive data sets, probability-based designs, methods comparable across the assessments and QA/QC. The approach has been well done.</p> <p>In Section 4.3, the utilization of the background-matching approach for geographical applicability was effective in Ecoregion 69 as well as in Section 5.3 and Ecoregion 70. The new portion was estimated at the 25<sup>th</sup> percentile, comparing with the background conductivity estimates of the original set. All chloride-dominated samples were removed before estimating background conductivity, thus keeping the same ionic mixture for the new area the same as the example criteria. The importance of keeping data inputs all of the same “category” for a quality comparison assessment is more valuable and fundamental to good statistics than satisfying an approach that believes all data should be included. Thus, this answers previous questions of whether there should be exclusion of particular ions.</p>	
Reviewer 3	Based on the analyses presented, I do not have any problem with the process. Indeed, the case study analyses in Sections 4.3 and 5.3 do support the criteria application. Basically, I feel that the parent data set is a training set, and the conductivity estimates outside of this set should be well within statistical bounds.	
Reviewer 4	The same comments I provided above apply to this charge question.	
Reviewer 5	Answer is similar to that provided to Question 9 above: The assumption here is that taxa comprising benthic macroinvertebrate communities within the smaller area are similar to those that occur within the larger area. This validity of this assumption should be verified	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 10: Geographic applicability to a new area within an ecoregion:** Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
	before extending criteria developed in one area to another, regardless of ecoregion boundaries.	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 11: Geographic applicability to a new area in another ecoregion:** Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The applicability analysis is clearly described, but again I am not fully convinced that the approach is sufficient for the reason described in my comments on Question 9. The case study appears to support the approach, but the new ecoregion (68) is just next to the original one (69). The result may differ substantially if the new ecoregion is further away and associated with very different benthic fauna. It is difficult to generalize the effectiveness of this background-matching approach based on this special case study.	
Reviewer 2	<p>As in Question #10 and within-ecoregion, the analysis for the background-matching approach for geographic applicability to a new area in another ecoregion was well presented. The monitoring and sampling procedures (Alabama, Kentucky, Tennessee) are very clearly described, well-defined, specific, and a pleasure to read. The Results in Section 6.2 are presented point by point. It is helpful to have these points in paragraphs 1-4 on pg. 6-10 – 6-11.</p> <p>Confidence intervals are greater in Ecoregion 70 than the other two regions (pg.6-12, Table 6.4). Perhaps reasons should be given for this. The difference is quite notable. Also in this table it is mentioned that the WABbase data set for Ecoregion 69 included samples without genus identification, meaning that identification was carried just to the Family level. Although it is always better to be able to key down to genus, this is not unusual. This is a problem for stream monitoring/sampling programs and will probably only get worse as fewer individuals are training in entomology.</p> <p>Verification of applicability of example criterion from an original ecoregion to a new ecoregion using the background-matching method was done by independently estimating the HC<sub>05</sub>. Verification is an important step – a necessary “hurdle” - and when it is successful the method can be confirmed to be reliable and its use can proceed. I compliment the authors for presenting the first demonstration of successfully applying a criterion to a new region.</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 11: Geographic applicability to a new area in another ecoregion:** Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
	<p>This is a major step in expanding agencies' ability to establish a criterion for a parameter which has a significant role in the health of aquatic organisms.</p> <p>I do have concerns about the need for about 500 samples in order to achieve consistent results with the HC<sub>05</sub> derivation (pg. 6-13). The need for a large data set is well understood, but it may be difficult for some entities to have that many samples. The mention that there were different sampling methods, and that some methods tend to collect different types of genera, is appreciated. While it would be better to have uniform sampling methods, in reality that doesn't always happen. The authors tried to restore confidence that a large variety of taxa were nevertheless represented. It might be wise to recheck the methods/protocols used, verifying that each method was used about equally throughout the data set.</p> <p>The authors were very specific and thorough in their description on pg. 6-13. The applicability was well presented.</p> <p>As I have mentioned previously, my expertise lies in the biological aspects rather than the statistical analyses. With this in mind, it would be helpful to have more information about bootstrapping and an example by which one could follow. This would be an Appendix supplement I realize, but I think it would be useful to staff in watershed programs.</p>	
Reviewer 3	I think this section was well written, and that the data and analyses do support the application of conductivity criteria to other areas.	
Reviewer 4	I think the background matching analysis is a sound approach for comparing applicability to another ecoregion. Again, I have the same caveat about potentially doing this in a tiered use framework.	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 11: Geographic applicability to a new area in another ecoregion:** Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
Reviewer 5	<p>The background matching concept is clearly described, generally, but certain details are murky.</p> <p>As mentioned above, the process for defining a “background SC” should be more explicit – if it is to be used as an essential component of a criterion definition process as described by the document.</p> <p>One detail in the Case Study 3 example is not clear. Multiple data sets were used for Case Study 1 and for Case Study 2. A background SC estimate (<math>\pm</math> CI) was derived for each of these datasets. Then, Case Study 3 uses a single background SC estimate from each of the two case studies for the “matching” analysis but it does not state a clear rationale for selection. More specifically: the value described as the “the example Criterion data set” (as described on page 4-21) was used to represent Case Study 1 (94 <math>\mu</math>S/cm; 95% CI 86–101 <math>\mu</math>S/cm); however, the background SC estimated for the corresponding data set (“example Criterion derivation data set” as described on page 5-22) was not used to represent Case Study 2; rather, the “WABbase data set, probability sample subset” (147 <math>\mu</math>S/cm; 95% CI 136–159 <math>\mu</math>S/cm) was used to represent Case Study 2. What is the rationale for deciding which background SC estimate to use in procedures such as the Case Study 3 example?</p> <p>Also, it is not clear to me how the 95% CIs are derived for the background SC estimates. Maybe that procedure is described somewhere in the document but I am not finding it.</p> <p>In my view, the underlying rationale for use of the background matching approach alone to establish HC<sub>05</sub> values is not clearly supported (as stated above in response to question 9). Justification of the process would require biological confirmation; no process for biological confirmation is described.</p>	



**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 12: Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

REVIEWER	COMMENT	RESPONSE
Reviewer 1	<p>Ephemeral streams typically support highly-mobile taxa (e.g., beetles) and taxa of short life cycles (e.g., some chironomids). They may collectively share most genera with perennial streams, as shown by Grubbs (2010), but the occurrence frequencies and abundance of most shared taxa are most likely to be much lower. My experience is that not many sensitive genera live in temporal streams. As a result, these streams may be “over-protected” by a HC<sub>05</sub> established for perennial streams. However, I agree that they are important components of stream networks, but should be protected by separate conductivity criteria. Temporal streams have recently attracted much research interest.</p> <p>Following references are relevant:</p> <p>Lake, P.S. 2011. <i>Drought and aquatic ecosystems: effects and responses</i>. Chichester, UK. Wiley-Blackwell.</p> <p>Steward, A.L., D. von Schiller, K. Tockner, J.C. Marshall, and S.E. Bunn 2012. When the river runs dry: human and ecological values of dry riverbeds. <i>Frontiers in Ecology and the Environment</i> <b>10</b>: 202–209</p> <p>Williams, D.D. 2006. <i>The biology of temporary waters</i>. New York, NY: Oxford University Press.</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 12: Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

REVIEWER	COMMENT	RESPONSE
Reviewer 2	<p>Yes, the discussion in Section 3.6.2 well supports the field-based method for applicability to ephemeral streams. The support is well defined by the publications cited. It clearly discusses that macroinvertebrates are found in intermittent and ephemeral streams. Grubbs' (2010) research provides excellent quantifying results. In my experience, I believe the abundance and diversity of macroinvertebrates in ephemeral and intermittent streams is far greater than larger rivers which (in this part of the country) have nearly all been channelized, straightened, trees/brush removed, island and meanders removed, streambeds laden with silt, and hydrologically modified.</p> <p>I found the short discussion of the various adaptations to survive temporary dry periods (pg. 3-20, lines 27-30) to be exceptionally helpful. Very seldom is this addressed or even widely known, however, it is a significant fact among many of the taxa of ephemeral streams. It was a pleasure to see this included. And the continuing discussion of the use of upstream temporary streams for part of their life cycle (pg. 3-30 - 3-31) is also accurate and equally important to include. This fact, and the documented presence of the "vast majority (91 out of 108) of macroinvertebrate taxa were observed in both the perennial and temporary channels" (Grubbs (2010), provides strong rationale for the applicability of field-based method/criterion for conductivity to ephemeral streams. Upstream water quality conditions affect lower reaches' aquatic life and the exposure to harmful levels of conductivity (and all other contaminants as well).</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 12: Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

REVIEWER	COMMENT	RESPONSE
	<p>As I've mentioned in my response in Question #2, intermittent and ephemeral streams (even "often-wet" depressions in fields, wet meadows and pasture drainages, wet areas in riparian corridors or nearby river valleys, etc.) provide habitat for macroinvertebrates – at least for a portion of their life cycle. The value of these small streams and temporary wet areas as habitat for many taxa has not been appreciated nor understood by many property owners, developers, and policy decision-makers. The decision to include ephemeral streams in this criterion development is especially important and gratefully appreciated by biologists such as myself. The information provided here in this section also provides strong rationale for the current debate on "navigable waters" regulations.</p> <p>Please refer back to my response for Question #2 for my other previous comments.</p>	
Reviewer 3	<p>My basic comment is that we are not doing enough to protect either intermittent or ephemeral streams, and I would recommend that EPA take a stronger stance on these important characteristics of the watershed. It is hard enough to protect 1<sup>st</sup> order streams in the United States, but trying to gain protection for zero order streams is almost impossible. Consequently, and where applicable, any paired analyses with conductivity and benthic organisms would be beneficial to support the importance of ephemeral streams.</p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 12: Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

REVIEWER	COMMENT	RESPONSE
Reviewer 4	I would examine Ohio EPA's Primary Headwater Assessment data. It focused on streams generally less than about a square mile and includes ephemeral as well as perennial and interstitial streams. They have collected conductivity data and have sites in the WAP ecoregion as part of their studies. Ohio University also collected primary headwater stream data as part of a study in the vicinity of the Portsmouth nuclear facility that might be of use. In a neighboring ecoregion (Interior Plateau), MBI has data from around 100 or so streams around Hamilton Co., although many have urban impacts.	
Reviewer 5	<p>I answer this question with two caveats: (1) I am not trained in aquatic ecology and have no professional experience dealing with the subject of this question, and (2) I am well aware of legal issues concerning "Waters of the United States," and that the role of ephemeral streams within that framework is at issue; in answering this question, I take no position on that issue.</p> <p>Because I lack experience with the precise issue, I reviewed several articles on the topic including some cited by the document and some not. In my reading of the literature, it became clear that there is significant overlap among taxa residing in "temporary" streams (as ephemeral streams are often called in these studies) and those residing in permanent streams. Therefore, it is reasonable to expect that HC<sub>05</sub> levels derived from analysis of biological data collected from low-order permanent streams will be protective of most taxa occurring in ephemeral streams. However, because some taxa occurring in low-order permanent streams do not occur in ephemeral streams, and vice versa, it is not reasonable to expect that the HC<sub>05</sub></p>	

**TABLE 4. RESPONSE TO CHARGE QUESTIONS 9-12: GEOGRAPHIC APPLICABILITY**

**CHARGE QUESTION 12: Applicability to ephemeral streams:** In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?

REVIEWER	COMMENT	RESPONSE
	<p>levels derived from low-order permanent streams to achieve the goals of the document's HC<sub>05</sub> derivation method exactly. Hence, it is my view that a process for applying HC<sub>05</sub> levels derived from permanent streams to ephemeral streams would require biological confirmation. One method of biological confirmation could be to verify the presence of the permanent streams' limit-defining taxa within the ephemeral streams; other methods are also possible.</p> <p>Articles reviewed to reach this opinion include DeJong et al. (2013), del Rosario and Resh (2000), Delucchi (1988), Feminella (1996), Grubb (2010), Price et al. (2003), Stout and Wallace (2003), and Williams (1996):</p>	



**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	The process is clearly described. I have a number of concerns. First, the fish sampling methods (sampling gears and distance) used by different agencies/programs are not detailed, but likely inconsistent. For example, if one set of samples were collected over a reach of 40-time channel width, but another over 20-time channel width, the former likely capture more species. As a result, the occurrence frequency of a species may vary with sampling method or data sources, introducing noise into the analysis. The sample comparability of the various sources needs to be evaluated. Second, the sample sub-setting based on major basins is well defensible, but much less so when based on stream size. Although many fish species prefer streams of certain sizes, they also occur in streams of different sizes. Any numerical thresholds seem arbitrary. The modification to the bootstrapping process is reasonable; however, my earlier criticisms to XC <sub>95</sub> estimation, its relevancy to species extirpation and SD curves are applicable here.	
Reviewer 2	<p>I believe that the work done to derive a fish HC<sub>05</sub> was exceptionally well done. Data sets were very large and a number of considerations which are fish-specific were incorporated. These provide validity to the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values. The suitability of the method – to be applicable to fish as well as macroinvertebrates – is especially reflective of the quality of this work and its usability for widespread application to aquatic organisms.</p> <p>The following are thoughts and notes which I made as I progressed through the section.</p> <ul style="list-style-type: none"> <li>On page G-3, the section accurately makes the connection between loss of macroinvertebrates (impacted by high conductivity levels) and the stress this puts on fish by decreased food availability. This is clearly an additional justification for a</li> </ul>	



**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC<sub>05</sub> FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
	<p>conductivity criterion – fish losses reduce the natural resource quality of the stream, reduces recreational potential, increases the number of threatened and endangered species and possible extirpation.</p> <p>Should other stressors be included in the above discussion? For example: poor habitat, lack of water depth diversity, high ammonia levels, high suspended solids, low DO ... all are stressful to fish either directly or indirectly. This weakens fish so that the effects of other stresses, such as the ionic imbalance from high conductivity levels, are likely enhanced. In other words, should other types of stress be taken into the analyses with conductivity because of the increased level of impact there might be?</p> <ul style="list-style-type: none"> <li>• Of the six fish species listed as relatively tolerant of elevated conductivity, pg. G-4, lines 4-7, I believe that <i>Micropterus dolomieu</i> (smallmouth bass) perhaps shouldn't be included. It is generally intolerant of pollutants and poor water quality. Even <i>Lepomis cyanellus</i> (green sunfish) prefers somewhat good conditions, even though it can be found in eutrophic waters.</li> <li>• Identification of fish to the species level is indeed the routine for fish sampling, unlike macroinvertebrates, and this does lend itself to species-level XC<sub>95</sub> values.</li> <li>• The fish analyses used a combined data set for fish from portions of four contiguous ecoregions and seven states! Seven data sets collected between 1991 and 2009, 1657 sampling events across 1364 distinct sites, gives great spatial and temporal data.</li> </ul>	

**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
	<p>What an amazing quantity and variety of data! This extensive data base is difficult to find in the environmental arena. Kudos for bringing together such an excellent base from which to assess for a fish conductivity criterion.</p> <ul style="list-style-type: none"> <li>A concern: Reference sites were not identified in the dataset, however the document seems to imply that 134 sites, which were &gt;90% forested, were likely such. What is the reason for not identifying, and assuring, that reference sites were included? Could the data not have been identified, perhaps as a separate grouping within the dataset? How can you be sure that there were adequate reference sites in the initial sample collections? Heavily forested does not necessarily assure that water quality will be of high quality. This exact situation happened with a monitoring program which I designed and implemented for the stream system running through Omaha, Nebraska. After careful searching, I found what appeared to be a minimally impacted small stream of which for most its length it flowed through rolling wooded hills. Although some small acreages, occasional houses, and further back, a new development were in among the hills and woods, the stream appeared to be minimally impacted and exhibited great fish and macroinvertebrate habitat. It appeared to be the perfect reference stream (there was no existing data available to use as a guide). Eventually, it became exceedingly apparent that I chose poorly as fecal coliform levels repeatedly were some of the highest of the 24 sampling sites in the system and some of the other parameters were also not appreciably better than any of the other sites.</li> </ul>	

**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
	<ul style="list-style-type: none"> <li>Description on pg. G-4, Lines 15 -28 is very thorough; provides the reader with a detailed understanding of the area, conditions, etc. In G-7, there is good use of excluding larger catchment areas, sites with low pH and high chloride levels, and sites which were too small to support fish. This strengthens the data analyses. The discussion of fish on pg. G-10 regarding the exclusion of sites where there is question of the presence or absence of a species is also good. I am surprised to read that brook trout (<i>Salvelinus fontinalis</i>) are stocked. While they are native, they are few in number and are especially sensitive to poor water quality conditions. I believe that I disagree that they should be included. I have concerns about counting any of the stocked fish species because of the possibility of affecting the XC<sub>95</sub> estimates. Stocking is a manmade “condition”, largely to improve recreational fishing; the expected life span is pretty much irrelevant and independent of the stream’s condition.</li> <li>The paragraph on pg. G-11, lines 9-15, is not fully clear to me. It would be helpful to have it explained a bit more fully. The following lines on that page, lines 16-31, are well done. I would say, however, if a sensitive taxon is found in a waterway in which it is unexpected (outside the distribution of that species), there is the possibility that perhaps the species range had not been accurately established originally. Or that it had expanded its range – either way, it would probably be best to check with local fish biologists before exclusion.</li> </ul>	
Reviewer 3	Unless I am missing something, there is no Section 3.4 in Appendix G. So, there are many obvious differences between benthic and fish data in assessing conductivity effects. The benthic data is at the genus level – good enough, but fishes are easily identifiable (with a few	

**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC<sub>05</sub> FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
	<p>exceptions) to the species level. However, fish sampling is more time consuming so not as many samples are collected in comparison to the benthic collections. There are a number of other considerations as well discussed on pages G-1 and G-2.</p> <p>I liked the approach where fish data was lumped from four ecoregions, which resulted in a data base of over 1,437 observations. The clarity and validity of the modified method was adequate to derive the XC<sub>95</sub> and HC<sub>05</sub> values. I also liked the data filters that were employed for the analysis – these eliminated a lot of potential problems with the data analysis.</p> <p><u>OK – here is where I am unhappy with the fish data.</u> First, both rainbow trout and brown trout must be excluded from the data set. These are exotic, introduced species and even though there are established populations of these two species, that is not a good reason to include them. Many folks are trying hard to protect native species throughout the Appalachians, and including them as well as carp just does not make sense. I would follow the listings for introduced and exotic species, as found in Wiley and Hocutt, as the cut for potentially introduced species into an ecoregion. Also, just because the two trout species are recreationally important, that is not a good enough reason to include them in the analysis. If the work done by Tim King is valid, then there are only a few places where one needs to worry about brook trout introductions. After all, this is the native salmonid of the Appalachians.</p> <p>I like Figure G-9 very much. First, it showed the species sensitivity distributions for many fish species. More importantly, it illustrated that even within a genus, there was wide variation in the response to conductivity, e.g., <i>Etheostoma</i> and <i>Cottus</i> sp. The derivation of the fish HC<sub>05</sub> is excellent, but the hazardous concentration of 392 µS/cm seems a little high</p>	

**TABLE 5. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: *FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)***

**CHARGE QUESTION 13: General:** The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.

REVIEWER	COMMENT	RESPONSE
	to me, but that may be a reflection of the species that are common in my research sites.	
Reviewer 4	Again, I generally found no real problem with the statistical approach and found the modified bootstrap methodology reasonable. My main concerns are related to how the regions are combined, given biogeographical differences in fish distributions across ecoregions. The argument is made that, in a manner similar to SSDs generated for toxicity testing, it is not important that the species that make up those below the HC <sub>05</sub> do not occur in Ohio. I am not sure this is reasonable for a natural “stressor” such as conductivity. The benchmark is driven by coldwater and rare species that do not occur in Ohio. If this analysis was conducted with Ohio ecoregion 70 data alone, perhaps with a lower threshold of sample size, perhaps it is possible that other sensitive fish species would replace the most sensitive taxa in Appendix G. It is likely, however, that those sensitive taxa are inhabitants of the EWH tiered use in Ohio rather than the WWH use. Again, I think that some discussion of tiered uses is essential to how a State might apply this approach.	
Reviewer 5	I have no professional expertise or activities that concern fish. I can see that the data analysis procedure is similar. However, my scientific knowledge of fish and of environmental characteristics that influence their condition and behavior is insufficient to allow my informed opinion about this part of the document.	



**TABLE 6. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)**

**CHARGE QUESTION 14: Protection:** Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
Reviewer 1	Just as the case studies for macroinvertebrates, the conductivity criterion derived here appears to be protective for fish in the study region in practice. However, EPA needs to address the lack of biological relevancy of XC <sub>95</sub> to species extirpation and vague interpretation of SD curves, as I described earlier.	
Reviewer 2	Yes, the analyses for the fish example criterion is protective of fish. The HC <sub>05</sub> of 392 uS/cm (95% CI 256-424 uS/cm) is appropriate. Good discussion in the Results, and as I observed in Question #13, many strong attributes accompany the fish analyses.	
Reviewer 3	<p>It appears the case study criteria for the benthic organisms would also be protective of fish populations. I don't think any additional data and analyses are needed. I look at this approach as a two-factor method where the benthic conductivity criteria would drive the fish protection, and perhaps <i>vice versa</i> in special situations.</p> <p>FINAL COMMENT: I would like to have seen a little more done with assemblages, e.g., EPT, intolerants, tolerants, etc. However, the genus level for bugs and the species level for fish approaches are great, especially with the highly robust data sets.</p>	
Reviewer 4	My experience with conductivity and other ionic strength parameters is that macroinvertebrates are generally more sensitive as a group than fish, but I think because of sample size considerations the most sensitive fish are often being excluded from the analyses here. The limited distribution of certain fish species (and macroinvertebrate genera) that are often excluded may in themselves be evidence that the aquatic life potential may vary with some natural as well as anthropogenic impacts. As monitoring programs mature, samples sizes in these areas are continually growing, so I think that State water quality standards programs need to be continually exploring their databases to refine aquatic life uses and the criteria designed to protect these uses. Modification to the approach for deriving a single ecoregion criterion for all streams requires an adequate monitoring program with robust critical program elements (Yoder and Barbour 2009) that provide the data and the capability to conduct the analyses described in this document. For	



**TABLE 6. RESPONSE TO CHARGE QUESTIONS 13-14: SUPPORTING INFORMATION: *FIELD-BASED HC05 FOR FISH IN APPALACHIAN STREAMS (APPENDIX G)***

**CHARGE QUESTION 14: Protection:** Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.

REVIEWER	COMMENT	RESPONSE
	example, a program needs to be capable of accurately classifying and controlling for natural features that influence biological assemblages (e.g., stream gradient, size, elevation, base flow, etc.). Tiered uses require robust data with the ability to recognize the influence of anthropogenic influences on the landscape and the ability to address what stressors may or may not be feasibly controllable.	
Reviewer 5	Same answer as for number 13.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 1	xviii	--	On definition of XC <sub>95</sub> . It is not clear what do the authors mean by “effectively absent” here.	
	xix	--	First, this definition of extirpation is very different from what commonly used in the literature of conservation biology (absent from a region or regions, but still present somewhere else), and then may cause confusion. Second, it is also hard to determine when a genus is no longer a viable resource or unlikely to fulfill its functions, particularly when only 200 individuals are counted from a site.	
	1-1	8	“ . . waters dominated by . . ” Do the authors mean “waters naturally dominated by”? If not, one would not be able to apply the method to streams contaminated, say, by NaCl from road deicing. Clarify.	
	1-2	11-12	If any studies/data show Ecoregion III effectively capture the natural variation of conductivity across space, cite them. If not, the authors need to justify this decision.	
	2-1	9-11	Here the authors set the threshed of conductivity for extirpation as the level below which 95% observations occur. Above this conductivity level, the taxon is assumed to be no longer a viable resource or unlikely to fulfill its function. However, how they actually did this is much more complex (P3-13). They can leave the details to later, but they need to give readers some idea about how it is actually done. Otherwise, one may reject their method right away because above this threshold, a genus may be still common and viable!	
	2-10	Figure 2-1	Does increase of ion concentration always lead to decline of macroinvertebrate and fish species? I thought that the relative or true abundance of tolerant species may increase, just as shown in the case studies (B19-30). Modify.	
	2-15	21-22	See my earlier comments regarding species extirpation.	
	2-17	22	The authors need a newer citation, if available.	
	2-19	1	Replace “many states” with “most states”(?)	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 1 (continued)	2-19	17	“Freshwater insects are among the most sensitive. .” This statement is too general. Freshwater insects differ greatly in their sensitivity to human disturbances. Most EPT species are sensitive to organic pollution, but most chironomid genera are tolerant and so are most other dipterans (true flies). Modify.	
	2-20	17-18	“In other words, . . .value.” This threshold may make sense if a taxon rapidly decreases with conductivity; however, it makes no sense if a taxon increases or does not change with conductivity over the range observed. The authors addressed this issue on page 3-13, however, they need to bring their arguments and solutions up here. The authors also state that “In other words, the probability of 0.05 that an observation of a genus occurs above its XC <sub>95</sub> conductivity value.” Does this statement really hold when observations in a large bin are down-weighted in calculating XC <sub>95</sub> ? Even if this statement is valid, it is still confusing. Readers might interpret it as that one should expect to capture a genus at 5 of 100 sites where conductivity is greater than its XC <sub>95</sub> . The authors need to give a clearer and biological meaningful interpretation of XC <sub>95</sub> .	
	3-3	Figure 3-1	This figure is confusing. First, when 85% of the samples contained <i>Cheumatopsyche</i> in the bin with conductivity >1000 µs/cm, how is the genus assumed to extirpate to close to do so? Second, even when the “probability of capture” of a genus declined with increasing conductivity, the taxon may be still be common at its XC <sub>95</sub> , approximately 20-40% of the samples for <i>Stenonem</i> and <i>Leuctra</i> , respectively in the case study. Considering the impact detectability of a genus associated with small sample size (200 counts) and limited sampling period (once a year), the probability of occurrence could be much higher. One can argue that both genera are strong, at least far away from extirpation. It appears difficult, if possible at all, to consistently relate the extirpation of a genus to its XC <sub>95</sub> . The authors addressed this issue later in pages 3-12 and 3-13, but they need to give a full treatment when interpreting the figure or when introducing XC <sub>95</sub> in P2-20.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 1 (continued)	3-3	Figure 3-1	I am also concerned about their use of the term, “probability of capture”. In the literature, two terms, occurrence probability and detectability, are typically used to describe observations of a species, occurrence probability and detectability. The former describes the probability of a species to occur at a site (any spatial unit). The later refers to the probability of a species to be detected when it is present at a site. “The proportion of samples with a genus present at a conductivity level” could be taken as an estimate of occurrence probability only if detectability is assumed to be 1, something that rarely occurs in practice. I suggest the authors to use “relative frequency,” rather than probability of capture, which has been commonly used to refer to the % of individuals captured by a sample. In addition, the authors estimated the proportion of genus observation for a conductivity bin, rather than a conductivity level. Modify and clarify the terminology.	
	3-5	20	“background . . . region;” It would be helpful to clearly state how similar conductivity among reference sites is similar enough.	
	3-8	8-12	One major source of salinity in freshwater waters snow zone is road de-icing. The conductivity criteria described here will not be applicable to assess the impact of NaCl used for de-icing? (also see my earlier comment on this issue)	
	3-10	26-31	See my comments on the relationship between XC <sub>95</sub> and extirpation earlier.	
	3-11	1	Did the authors assess how bin delineation affected CDF? The description here is a bit vague regarding how they balanced the number of bins and the size of bins. Clarify.	
	3-11	eq 3-1	“x” needs to be defined.	
	3-12	Figure 3-3	Replace “cumulative probability” with “cumulative proportion” to be consistent with the text (L2), and to avoid the issue of imperfect detectability. I also suggest adding a third panel figure to show a concave increasing curve like the one for <i>Corydalis</i> (B-42). This type of curve really means a positive response of a genus to conductivity (B-29 for the same genus). For the 2 <sup>nd</sup> ( <i>Nigronia</i> ) and 3 <sup>rd</sup> types of curves ( <i>Corydalis</i> ), it is not possible to relate XC <sub>95</sub> to extirpation, as I argued earlier. Glad to see the authors starting to address this critical issue here. However, the issue needs to be fully treated much earlier. I also do not see it being a data-	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			distribution issue, but a fundamental limitation of CDF. CDF curves of Types 2-3 are also not anomalies, but they are normal and frequent, as shown in the case studies. Revise.	
Reviewer 1 (continued)	3-13	1-7	Yes, the qualifying designation helps for understanding HC <sub>05</sub> , but relating XC <sub>95</sub> to extirpation remains conceptually flawed. See my comments to Charge Question 4 for possible options.	
	3-13	21-22	Replace “mean curve” with “fitted curve”. Also, what is the confidence limit? 95% or 90%? Clarify.	
	3-17	1-3	A further concern is whether sampling dates/period is related to conductivity. If most high-conductivity sites were sampled in spring (March-June), but low-conductivity sites in summer, one likely underestimates the occurrences of sensitive taxa in the latter and then overestimate HC <sub>05</sub> conductivity. Correlation between conductivity and sampling time can be used to identify the bias.	
	3-19	eq 3-2	This equation needs to be written in a standard math format as follows: $CMEC = 10^{(\bar{X} + z_{\alpha} \times \sigma_r)}$	
	3-21	10	“ . . and often more than 4 days”. Above CCC? Clarify.	
	3-19	11-14	Is “the one-tail critical value” half of the number of the standard deviation required for 90% confidence limits? The authors also define $\bar{X}$ twice here and differently. Clarify or correct.	
	3-26	27	“More than 90% of . . insects.” This statement is too broad. In many streams, insects took less than 90% of all individuals. Add “often” or “frequently”.	
	F-16	eq F-1	Re-write the equation in a standard math format	
	G-3	7-8	This statement is too broad. Many adult insects, such as winter stoneflies, actually only move over a short distance.	
	G-10	3	Add “hybrids” after “immature specimen”	
	G-18	14	Do the authors mean a selected <b>minimum</b> size ranging from 0-60 occurrence? If so, how it can be zero? Clarify.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2	vii	5-4	Space needed between “of” and “survey”	
	xi	5-10	Delete the “and” before Kentucky	
	xvii	Continuing paragraph	Paired analyses is a strong statistical test. Exceptional number of field samples, sites and years of sampling!	
	xvii and 2-3	10-12	Conductivity is described for eastern and western montane ecoregions but nothing is said about the Midwest – it’s a major portion of the mid-section of the country and probably should be included.	
	xviii	Executive Summary	Very well done.	
	2-4	25	Figure 2-1 is located five pages away; could it be moved closer to pg.2-4?	
	2-3 to 2-7	Section 2.2.1	Thorough; excellent overview and foundational information	
	2-8	Table 2-1	Parentheses should encase 2012 in: Samarina (2007); Ruhl et al 2012.	
	2-11	Sections 2.2.3, 2.3, 2.5	Also excellent information; valuable for water quality staff to better understand the causes and mechanisms.	
	3-1	29-30	Could there be a bit more information with the “weighted CDF model”?	
	3-2	1-3	An accompanying short explanation of the statistical package R would be helpful.	
	3-10	Section 3.1.1.3.	Specific description of the sampling methods as well as assurances that adherence to standardized sampling techniques were observed, would be nice. Perhaps sampling details are in the Appendices.	
	3-16	8-15	Good recognition of the variance in sampling protocols among different agencies or monitoring groups. My concern is whether this variability can be “handled” by the process? The authors believe that it does.	
	3-23	Section 3.4	Well done.	



**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	3-27	10	Would not extirpation for the most sensitive crustacean occur before 960 uS/cm (of <960 uS/cm)? This is a high level of conductivity and mollusks are “canary” indicators of contaminants and water soluble stressors.	
	3-30	1-8	This was new information for me and find it very interesting. I did not know that Ephemeroptera can tolerate such low pH conditions if the conductivity is high. Good information; well done throughout all of Section 3.6.	
	3-32	Figure 3-8	An amazingly weak correlation! I might not have expected this, but it is clear.	
	3-34	7	“illiustrated” is misspelled.	
	4-1 to 4-27	Section 4	Figures and tables are very helpful.	
	4-8 & 4-9	Figures 4-3, 4-4	These figures show large increases in conductivity in October. Seems that these higher values would affect the calculations for HC <sub>05</sub> and the HC <sub>05</sub> when simply looking at the figures. I understand the explanations given but am not 100% sure that they are complete enough for the less-trained in statistics.	
	4-27	4-11	Clarity of the use of the one day sampling/grab sample serving for CMEC and CCC.	
	5-4	Figure 5-1	Number of and distribution of sampling sites is exceptional.	
	5-6	12-15	Wouldn't the <200uS/cm Dec through June and >200 July through Oct provide support for the argument that there is a seasonal difference, thus calling for separation of data by seasons or seasonal weighting?	
	5-7	Figure 5-2	Seasonal variation is clearly shown for September. Difficult to understand how this wouldn't skew the results.	
	5-12	1 <sup>st</sup> paragraph	Personal Comment: Here in the Midwest we have distinct seasons, and water quality parameters often reflect this. Having unweighted, monthly/seasonal data is helpful to state agency staff who are trying to determine sources of pollutants and causal relationships. Determining sources of impairment are challenging and a clear understanding of what is happening each month (when there is monthly data available!), provides insight.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	5-14	Figure 5-7	I concur with the acceptability of the hazardous concentration of 338 uS/cm but some interest groups may believe it is too stringent.	
	5-18	Line 7	Delete the second “for.”	
	5-20	Figure 5-10	Delete “and” in the figure’s title: “...southeastern Ohio into <u>and</u> Kentucky”	
	A-1	Figure A-1	A sentence or two describing LOWESS would be very helpful.	
	A-4	Figure A-4	Good to address other water quality parameters; informative.	
	A-5	1 <sup>st</sup> paragraph	My initial thoughts when reading this were that the confounders listed would have an effect on conductivity, and as I’ve stated in my response to one of the charge questions, I do believe that additional stressors can indirectly increase the damage done by high conductivity levels. Sorting it out, however, is an immensely difficult undertaking, requiring considerable data much uncertainty. However, in this document it was determined that confounders were not an issue.	
	A-7	7-8	“Removal of poor habitat samples from the data set had almost no effect on the SD model or HC05.” Based on the work of this study, this appears to be true. Unfortunately, if the removal of poor habitat doesn’t affect conductivity, then those who oppose habitat restoration projects can use this as an argument in support of their position.	
	A-11	1-7	Weighting, by its very purpose, brings a comparable ‘status’ to a data set with variable values and is a method by which calculations can be made. But care must be taken to correctly do the weighting to ensure correct representation of the data is seen in the results. I believe the authors have endeavored to do it well.	
	A-13	5	Not sure if this is recalculating to make data “fit” expectations. What was the RBP score used for the first calculations?	
	A-14	Section A.3.	Confidence intervals – there are some immensely large ranges.	
	B-9	1-5	Perhaps the RBP 130 score is not the correct level to use – could it have removed too many of poor and moderately poor sites?	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	B-9	17	Paired conductivity and biological data is a time-tested statistical test in environmental research; reliable and strong.	
	Appendix B	All figures	Well done; very helpful in conveying the relationships.	
	B-13	2-4	Of the 13 factors that were listed as being considered for having a casual relationship between conductivity and macroinvertebrates, some of them have had only minimal or no discussion in this document. Those are: nutrients, deposited sediments, selenium, settling ponds and dissolved oxygen. Selenium and metals were addressed in Appendix G.4.5.	
	Appendix C		Excellent data design, rationales, descriptions; Table C-1 very helpful.	
	C-4	Table “C-2”	The numbering of the tables is incorrect. It should be Table C-1 because it is a continuing of the table on the previous page.	
	C-5	Table “C-2”	Same as above	
	C-5	1-6	Helpful for understanding the information in the table.	
	C-7	4	“(see Table C-3)” should be: (see Table C-4). The numbering of the tables for the rest of the section is now ‘off’.	
	C-8	Table “C-3”	Should be “Table C-2”	
	C-14	Top of page	Figure “C-3” should be Figure C-2. Very interesting geological information.	
	C-14	Bottom of page	Figure “C-4” should be Figure C-3.	
	C-15	Figure “C-5”	Should be Figure C-4. Strong relationship in the cumulative distribution between the Criterion data set and the Ohio data set; gives significant strength to the analyses.	
	C-16	Figure “C-6”	Should be Figure C-5. Very good illustration of distributions’ overlap and the ranges overlap; Strong.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	C-18	Figure “C-7”	Should be Figure C-6.	
	C-19	4-6 and 22-23	Good points on looking at reasons for absence of sensitive species for evidence that the regions are different. However, in stating that “current conditions may not allow re-colonization,” means that habitat is poor, and this conflicts with the previous determination in the document that habitat quality doesn’t affect the analyses. Here it appears to factor-in.	
	C-21	6-7	The likelihood that “watersheds with >90% native vegetation are more likely to have low conductivity” are also likely to have better quality stream habitat. Indirectly this also supports the role of habitat.	
	C-22	Table C-10	Would have liked to see responses for 23-27, but this is an example of how there might be descriptions or verifications not clear or missing. Thanks to the authors for presenting it as it is.	
	C-24	Table “C-4”	Should be Table C-11. Very good table; informative and well presented.	
	C-26	2-3	“(see Figure C-7)” should be Figure C-6. “(see Figure C-8)” should be Figure C-7.	
	C-26	Figure “C-8”	Should be Figure C-7.	
	C-27	Figure “C-9”	Should be Figure C-8.	
	C-28	C.4.1	Appreciated the descriptions of the regions.	
	C-30	References	Brady, K: ... - overly bold underlining. Kahneman, D. - are there pages for the book?	
	F-3	Figure F-1	The tight correlations in the scatter plots are very good.	
	F-6	Table F-1	Excellent table; SO <sub>4</sub> + HCO <sub>3</sub> clearly significant.	
	F-21	Section F.5	Summary and Tables F-5 and F-6 are helpful. I wonder how the CCC of 160 mg/L and the CMEC at 300 mg/L compares with other regions around the country?	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	F-22	8	“(see Figure F-10)” should be Figure F-11.	
	F-23	Figure “F-10”	Should be Figure F-11.	
	F-24	Figure “F-11”	Should be Figure F-12. I would like to see more explanation of the bootstrapping method.	
	F-25	Figure “F-12”	Should be Figure F-13.	
	F-54	F.10	3 references should have underlining of the authors if the format is to be kept the same throughout: Barbour, MT..., Newman, MC..., and R Development Core Team	
	G-1	28-29	While I understand the need for minimum sample sizes of 500-800 macroinvertebrate sample and 800-1000 fish samples, I wonder if state agencies will be able to have that many in their databases for each ecoregion? Has there been any checking with other states to see if most can meet this?	
	G-2; G-3	1-30; 1-13	Clear, accurate, and helpful discussion.	
	G-20	Figure G-8	Second to last line: “and 75-80 species evaluated.” The text on page G-18, line 29, said that it is 89, not 80. And in another, the number was 87.	
	G-20		Section 3.4 seems to be missing. Pages go from G.3.2. on pg. G-17 to G.4 on page G-20. If 3.4 is there, I didn’t see it.	
	G-25	Figure “G-7”	Should be Figure G-11.	
	G-27	G.4.6	Multivariant analysis for fish was interesting, especially the finding that catchment area and habitat significantly contributed to the model.	
	G-28	4 <sup>th</sup> line down	“Catchement” is misspelled: should be catchment.	
	G-34	G.6	The format of entries in the Appendix G’s references is not exactly the same as in the main reference section; To maintain the format here, Gerritsen, J.,... needs to have: a)semicolons b) initials follow the last name, c) uniformity in use of periods.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 2 (continued)	G-36	3, 7,12	“Availble” is misspelled. Should be available.	
	G-42	Table G-7, title	“Ecoregions observed are the ecoregions where the species was collected in the combined data set” - needs to be bold	
	7-1	References	<p>Reference Section</p> <p>The entire section does not maintain one particular format. The following are problems:</p> <ol style="list-style-type: none"> <li>1) The initials on authors who are not the first and last author are not uniformly handled. In picking a uniform format, I suggest placing the initials in front of the surname. And periods following the initials.</li> <li>2) Parentheses around the year of publication or just a period? Some entries have parentheses, others do not. Some have periods, others not.</li> <li>3) A period after the journal name – or not?</li> <li>4) Titles in small or all capital letters?</li> <li>5) Listing of pages referenced in books – often missing.</li> <li>6) The agency’s name followed by its abbreviation in parentheses, or, the reverse?</li> </ol> <p>The following is the first author’s last name on every entry that I suggest be changed to meet a standard format and have one or more of the above problems. For me to re-write each faulty entry would be too time consuming.</p> <p>APHA Barbour Berra Bradley Boelter Brinck Clark Cormier (2010) Dahm Duncan</p>	



**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			Dunlop (delete the second “Water Quality”) Echols (2009 a) Echols (2008b) Efron Entrekin Evans (2008a) (2008b) Evans (3001) Frag Fox Godwin Gregory Griffith Haluszczak Harper Hem Higgins Hill Hille Hitt Hopkins Hynes Jackson (2007) Jackson (2005) Kaushal (2005) Kaushal (2013) – check to see if it is now published. Kelly Kennedy (2003), (2004), (2005) Kimmel Komnick Lasier	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			<p>Lefebvre, O. <u>and</u> R. Moletta  Likens (1970)  Merricks  Meyer  Mount  Mullins  Newman (2000), (2001)  Nelson  NYSDEC  Omernik (1987), (1995)  Paul, M.J. <u>and</u> J.L. Meyer  Pond (2004), (2010), (2008), (2014)  Posthuma  Remane  Sams – spell out USGS: U.S. Geological Survey  Scanlon – add “and” just prior to the last author.  Smithson  Soucek  Stauffer  Stubblefield  Suter (2007), (2001)  U.S. EPA (1985), (1987), (2000a, 2000b), (2003), (2006),(2009), (2010) (2011a, 2011b, 2011c)  Van Dam – add “and” just prior to last author  Veil  Wallace – entomol. Needs capitalizing.  Werner – remove comma and add “and” between authors; Delete “Wright et al. 1993”  Wood (2008)  Woods (2002), (1996)  Ziegler (2007), (2010)</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			Zielinski	
Reviewer 3	Entire report	Entire report	Capitalize States where appropriate – eliminates the confusion between noun (States) and verb (state or states) forms. Perhaps also capitalize Tribe. Check foreword.	
	xiii		FORWORD should be FOREWORD	
	xiii	8, 14, 17	Capitalize States	
	2-4	15	Split ¶ after effluents (before Ionic)	
	2-10	Figure 2-1	Change black font to white on right side of all three blue blocks. One never uses black on blue.	
	2-17	18	Period after al (in et al. check entire document)	
	4-6	5	Figure 4.1 rather than 4.2	
	4-10	Figure 4-5	Cannot see data points!!!! Faint!!!!	
	5-11	Figure 5-5	Cannot see data points!!!! Faint!!!!	
Reviewer 4	--	--	<b>References Cited</b>	
			Bryce, S. A., D.P. Larsen, R.M. Hughes and P.R. Kaufmann, 1999. Assessing relative risks to aquatic ecosystems: A Mid-Appalachian case study. Journal of the American Water Resources Association. 35: 1752-1688.	
			Davies, S.P. and S.K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251-66.	
			Stoddard, J., P. Larsen, C. P. Hawkins, R. Johnson, and R. Norris. 2006. Setting expectations for the ecological condition of running waters: the concept of reference conditions. Ecological Applications 16:1267-1276.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	--	--	Wang, L., T. Brenden, P. Seelbach, A. Cooper, D. Allan, R. Clark Jr. and M. Wiley. 2008 Landscape Based Identification of Human Disturbance Gradients and Reference Conditions for Michigan Streams. <i>Environ Monit Assess</i> (2008) 141:1–17.  Yoder, C.O. and M.T. Barbour. 2009. Critical technical elements of state bioassessment programs: a process to evaluate program rigor and comparability. <i>Environmental Monitoring and Assessment</i> 150(1-4): pp 31-42.	
	Entire Report	--	As I mentioned in my general comments I think conductivity criteria should be considered in a tiered aquatic life use framework. There is natural variation in “background” conductivity due to variation in precipitation, base flow, etc., and within the range of “least impacted” to “minimally impacted” reference sites (as defined by Stoddard et al. 2006) there are variations due to human occupation of the landscape (e.g., agriculture, residential). My fear is that the criteria may not be stringent enough for the minimally impacted regions, and too stringent for land uses that State’s would not considered to be impaired, but rather to be consistent with the swimmable/fishable goals of the act. This is not a suggestion that least impaired would encompass mine-related acute impacts or other impacts that are feasibly controllable.	
	Glossary	--	“Background”– The definition of background I think is a bit “murky” given that later in text in includes both minimally impacted and least impacted reference sites. I would also add definitions of “least impacted” and “minimally impacted” reference sites and “tiered aquatic life uses.”	
	1-3	--	The document talks about how the protection of 95% of genera with this method is comparable with the protection of 95% of “species” in the lab toxicity approach. Some genera are more speciose than others. Is it truly similar? Not a major comment.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	2-3	25-28	This supports the contention that “background” levels can vary depending on how reference sites are defined. Within an ecoregion there could be subwatersheds with lower or higher conductivity than neighboring watersheds. Because of this, the natural distribution and abundance of those taxa that are most sensitive to conductivity can vary. These differences can be due to natural or some level of anthropogenic impacts (not acute or controllable sources). A discussion of tiered uses, I think is warranted for a stressor such as conductivity or perhaps other “natural” stressors such as habitat.	
	2-3; 2-12	29-30; 3-6	Because precipitation can influence conductivity (the variation in seasonal concentrations, that is higher in late summer, tends to be months when precipitation is lowest) it seems that measures of base flow may be important in resolving within regional background variation in conductivity. My experience from Ohio is that the Exceptional Warmwater Habitat streams often have high base flows. In any case, I think discussion of potentially tiering conductivity benchmarks should be discussed.	
	2-16	22-29	Along a gradient of stress, such as conductivity, the probability of capturing an individual of a genus decreases with increasing stress. How does sampling methodology potentially influence derivation of benchmarks. If a taxa is not collected when the sample size is small, there is some likelihood it is present when not represented in the catch. Is there a way to use methods that count many more organisms (e.g., Ohio EPA method can have abundance estimates greater than a 1000-2000) to determine the bias when using methods that only count 200 or 300 individuals. Thus what is considered “mortality” may be partly under sampling. Although the trend in taxa response with conductivity may be similar between methods, the actual benchmarks could change if “sensitive taxa” show up at somewhat higher conductivities.	
	3-5	20	What is considered “similar background conditions?” If there are two groups of sites, one centered on a conductivity of 150 and the other on 250 and both are considered background, is that similar enough? When do you decide that you have two groups of sites that might be within the same ecoregion but that differ enough in conductivity and taxa that two tiers are needed?	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	3-8	22-29	This is where the concept of tiered uses could be tested. If minimally impacted reference sites can be distinguished from “least impacted” then the general not observed at reference sites could differ. The BCG process that has been used in a number of states results in output that classifies sites into different BCG tiers with reference sites usually varying between tiers 1-4 depending on definition. Sites are rarely classified as BCG1, but data sets where sites identified as BCG tiers 1-2 could be compared to those defined as BCG tiers 3-4 to see how this might affect the derivation of criteria. It would certainly influence which sites occur or do not occur at reference sites.	
	3-17	18	Because of the assumption of using samples from both seasons to generate the criteria, this implies that looking for exceedences of conductivity should be based on a geometric mean value from monthly samples. For determining water quality violations of the criteria are we expected to take monthly samples including both spring and summer periods or is there a methodology to adjust the “expected” criteria if only summer samples are taken as is common for many monitoring programs. Some more specific guidance on sampling for what would be considered violations or exceedences of criteria would be useful.	
	3-35	4-14	This is where I think some explicit definition of reference site conditions (minimally impacted distinguished from least impacted) is advisable (sensu Stoddard et al. 2006). In addition, an independent human disturbance gradient could be used to estimate the anthropogenic footprint.	
	3-35	23	Would an independent human disturbance gradient measure be useful here?	
	3-39	--	Given the “fuzziness” of deriving background conditions I think this can be difficult. Again the concept of tiered uses is important in this context and I think there is a need for a more explicit approach to distinguish between minimally impacted, least impacted and best attainable. For example a 90% forest threshold for a watershed may not be feasible in many areas and it is an important question whether conductivity levels are actually feasibly controllable in all cases.	
	4-17	5-19	It is easier to read if the X-axis of the plot is in actual conductivity units rather than as the log of conductivity.	



**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	4-24	--	Is it possible that this graph argues for the existence of two tiers of expectations? For example could there be unique, less common areas where conductivity is naturally very low and the most sensitive taxa more narrowly distributed and thus rare and less likely to meet a threshold of 25 sites for a genus to occur, and then more typical sites where more sensitive taxa are not found as frequently?	
	A-13	--	General Question: The section analyzed differences of HC <sub>05</sub> values when low habitat scores were eliminated as potentially confounding variables. To explore the different benchmarks that might occur under tiered uses, perhaps the RPH habitat could be used to establish “reference” cutoffs under a crude tiered use scenario (reference sites with habitat scores > 160 vs. > 180). Ideally this could be done with a State like Ohio where tiered uses (EWH vs. WWH) are clearly defined.	
	C-1	19	Here is an example where I think the concept of background needs to be better quantified. At a minimum it should be related to Stoddard et al. (2006) definition of minimally vs. least impacted conditions. Ideally some form of a human disturbance score can be calculated. There is scattered mention of > 90% forested as a reference benchmark, but that may be hard to find even in the WAP ecoregion of Ohio. For States to apply these benchmarks I think it argues for detail discussion of tiered uses and reference or background conditions.	
	C-2	15-16	Clarify whether background is minimally disturbed or least impacted.	
	C-2	17-18	The concept of subregions and local variations in base flow, which may not be easily predictable needs to be discussed more. Base flow seems to be a very important influence on conductivity given the variation observed in conductivity by month. Conductivity is usually higher overall in late summer early fall when flow is a minimum, but base flow makes up the greatest % of stream flow. For very small headwater streams, base flows can vary substantially within a region depending on the complexity of groundwater systems and points where streams become gaining flows. In larger streams I think these likely average out within a region, but in small streams may be important and may result in difference in rare and sensitive macro taxa.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	C-3	Table C-1	Regional properties – need to add consideration of base flow to this table. The sandstone aquifers in SE Ohio tend to have conductivities of 450-600 which indicates as the percent of flow that consists of groundwater increases the more likely that higher conductivity will occur. May also want to see Ohio primary headwater assessment data from Ohio (streams < 1 sq mi) that has conductivity data <a href="http://www.epa.ohio.gov/portals/35/wqs/headwaters/PHWH_Compendium.pdf">http://www.epa.ohio.gov/portals/35/wqs/headwaters/PHWH_Compendium.pdf</a>	
	C-17	Table C-7	This table indicates that reference sites do not occur in Ohio, but Ohio does have reference sites. Also, would it be difficult to get forest cover for the Ohio sites? My guess is that few approach the 90% forest cover mentioned (page C-21, line 8) for further south. This has implications for attainability and setting feasible and controllable benchmarks.	
	C-17	Table C-10	Again reference sites are available for Ohio.	
	G-2	23-24	The comparison of species vs. genus level XC <sub>95</sub> values identified that genus values represented the more tolerant species in the genus. Why wouldn't that apply to the macroinvertebrate analyses and does this suggest that the lower conductivity sites in a region are not adequately protected? Would this be resolved with tiered uses and perhaps a lower threshold that would let more rare and sensitive species into a higher tier use?	
	G-7	4-5	Ohio fish sites are not all in sites I would call "perennial." Although very few of the sites dry completely, many small headwater sites can occur in what I would term interstitial streams that have periods where flows in riffles are subsurface, although permanent pools remain.	
	G-7	9	Again we have the 90% forested "benchmark" for reference with little discussion of what this means.	
	G-20	4-6	It would be useful to see data on sites that missed the N=25 cutoff in this table to see if they characterize tiered uses or local high quality sites that may be important or some unique restricted distribution that might be important.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 4 (continued)	G-22	9	In addition to tiered uses on some states, most states characterize coldwater uses as unique from warmwater uses. Ohio for example identifies species considered as characteristic of its coldwater uses. How would removal of coldwater taxa change the HC <sub>05</sub> values?	
	G-23	26-31	If coldwater benchmarks were delineated separately from warmwater benchmarks, how would this affect this conclusion?	
Reviewer 5	Entire report	--	<p><b>Comments Concerning Potential Regulatory Programs that Apply SC = ~300 <math>\mu</math>S/cm as Criterion:</b></p> <p>I offer this comment in the event that EPA does, eventually, issue an SC criteria document; and if such document would be accompanied by regulatory implementation guidelines or guidance to the states.</p> <p>Significant research concerning major ion release by mine spoil fills has been conducted at Virginia Tech, University of Kentucky, West Virginia University, and elsewhere; much has been learned about Appalachian mine spoils and their release of major ions when exposed to environmental processes. Results of these studies are described in a variety of publications including Orndorff et al. (2010), Agouridis et al. (2012), Daniels et al. (2013), Odenheimer (2013), Odenheimer et al. (2013), Evans et al. (2014), Sena (2014), and Sena et al. (2014).</p> <p>These results lead me to conclude that a regulatory program that restricts water discharges to &lt;300 <math>\mu</math>S/cm throughout the mining period is likely to have the effect of serving as an effective prohibition on Appalachian surface mining, as it is unlikely that this level can be achieved on most or all mine sites during the active mining process. It is possible that advanced spoil management and handling methods can be developed that would enable &lt;300 <math>\mu</math>S/cm level to be achieved after mining and reclamation are complete on mine sites that have adequate low SC/TDS spoil materials available for use in constructing soils and hydrologic media. Such practice will require that high SC/TDS materials be placed in non-hydrologic locations, such as within or beneath highly compacted spoil zones; and constriction of functional hydrologic media above those compacted zones.</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			Because materials giving rise to low SC/TDS drainage waters are often highly weathered, such practice would be consistent with associated practices required to restore forest plant communities.	
Reviewer 5 (continued)	Entire report	--	<p><b>Terminology: Conductivity / Specific Conductance:</b> Throughout, including page xix, Glossary; text under “Conductivity”, and page 2-12, line 22: “specific conductivity” is not a well-recognized term. “Specific conductance” is commonly used in scientific literature to describe the electrical conductivity of water at 25°C:</p> <ul style="list-style-type: none"> <li>• see US EPA, Analytical Methods and Laboratories; <a href="http://water.epa.gov/scitech/methods/">http://water.epa.gov/scitech/methods/</a>, Method 120.1, Conductance (Specific Conductance, umhos at 25°C).</li> <li>• see Hem J.D. 1989. Page 3-18, “Specific Electrical Conductance”.</li> <li>• Consider US EPA Storet Code 00095 (“SPECIFIC CONDUCTANCE (UMHOS/CM @ 25C)”.</li> <li>• Consider also US EPA SW-846, Method 9050A, Specific Conductance.</li> </ul> <p>The text states (page 2-13, lines 11-13) “The term “specific conductivity” indicates the measurement has been standardized to 25°C, a reference temperature (Wetzel 2001).” This may be true, but the Wetzel (2001) reference is far less well known than the references I have cited above. I am not aware of other instances where the term “specific conductivity” is used and accepted. In my experience, the term “specific conductance” is used more commonly to express this concept.</p> <p>Also: Operating instructions for hand-held meters, such as those commonly used by field personnel to monitor water quality often use the term “conductivity” as a measure of the raw electrical conductivity value; and “specific conductance” as the temperature-standardized value. Operators of these meters often have the choice of setting the readout for “conductivity” or “specific conductance.” In my discussions with personnel representing agencies that supervise water quality databases and in my review of such databases themselves, I have observed</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			<p>considerable difficulties when the word “conductivity” is used to represent a water quality variable. Such difficulties include lack of knowledge by certain agency personnel concerning whether the variable is intended to represent actual electrical conductivity of temperature-standardized conductivity (specific conductance); and apparent mingling of records representing both types of measurements under the “conductivity” heading. I strongly encourage EPA to follow what is a routine and well-recognized scientific practice: To reserve the word “conductivity” for raw electrical conductivity measures; and to use the “specific conductance” for temperature-standardized measures. EPA’s current practice (as represented by this document) assigns two different meanings to the word “conductivity.” Perhaps document authors have adopted this practice in response to usage of “conductivity” in Pond et al. (2008)? I call attention, however, to Pond et al. (2014), which uses the term “specific conductance” to designate the temperature-standardized value, as per current scientific practice and most other EPA documents that I am familiar with.</p> <p>In my comments, I have used the term “specific conductance” (SC) to represent a 25°C-standardized value.</p>	
Reviewer 5 (continued)	Entire report	--	<p><b>Terminology: Extirpation Concentrations:</b> In my view, this term is being used inappropriately.</p> <ul style="list-style-type: none"> <li>• Webster defines the term “extirpate” to mean “to destroy completely.” Other dictionaries have similar meanings. That is not an appropriate term here because of the way in which the document quantifies the term: 5% of observations occur at concentrations higher than the so-called extirpation concentration (expressed as XC<sub>95</sub>). The capture probability figures within the document (Figs. 3-1 and 3-4) indicate that individuals of certain taxa are being observed at concentrations &gt;2x XC<sub>95</sub>.</li> <li>• Any water quality measure is highly variable in time. The method described here recommends “measurements of the agent(s) should be paired in space and time with biological sampling”. If measurements were timed to acquire</li> </ul>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
			samples at the point in time within any given stream where SC is at its highest point during a given genera's life cycle, the term "extirpation concentration" would be more justifiable – but such targeting is not described by the method presented.	
Reviewer 5 (continued)	Entire report	--	<b>Terminology: Field-based Method.</b> In my view, a term such as "field data analysis method" would be more appropriate because the method described includes no actual field activities data collection; it relies solely on secondary data that have been obtained for other purposes.	
	Entire report	--	<b>General Comments: Box Plots:</b> <ul style="list-style-type: none"> <li>There are a number of box plots that show distributions of SC and related ions among months throughout the document (e.g. Figs 4-2, 4-3, 4-4, etc.). Given that most that most of the annual data are distributed quite unevenly among months, I would suggest displaying numbers of observations used to generate each monthly box plot.</li> <li>There are a number of box plots throughout the document. Suggest stating at some point quantities represented by the box plots (I presume median, 25<sup>th</sup>, and 75<sup>th</sup> percentiles for the boxes; 90<sup>th</sup> and 10<sup>th</sup> percentiles for the tails? Are all observations lying outside of the 10<sup>th</sup> and 90<sup>th</sup> percentiles represented as data points, or only some?) Suggest stating the nature of box plot representations explicitly at some point in the manuscript. The caption for the first box plot would be a logical place to do this.</li> </ul>	
	xxi	--	Glossary: Suggest that the term "Reference Site" be added to the glossary, given the importance of Reference Sites to the proposed method (e.g. Section 3.1.1.2.5. Exclusion of disturbance or pollution-dependent genera: "Genera that are not observed at reference sites or are estuarine or marine organisms are excluded from the data set.")	
	2-8	Table 2-1	Table should be annotated to communicate that these are examples only (i.e. this is not a comprehensive or exhaustive list.).	
	2-15	1-2	"charged particles" are not equivalent to "dissolved ions".	



**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	2-15	28	“Physiological Mechanisms”: My scientific background and training does not enable me to evaluate this section.	
	2-19	17-20	“Freshwater insects are among the most sensitive organisms relative to other taxa, including zooplankton, fish and amphibians (see Appendix G of this report; Kennedy et al. 2004; Echols et al. 2009b; Lazorchak et al. 2011; Consbrock et al. 2011; Williams et al. 2011).” The statement is poorly supported by the references provided. Lazorchak et al. (2011), Consbrock et al. (2011), and Williams et al. (2011) are citations of conference presentations; hence, supporting documentation is not available to the public or to reviewers and, hence, are inappropriate, in my view, as a means of providing scientific support for a statement with this level of significance in a (potential) regulatory document. Kennedy et al. (2004) compared sensitivity of <i>Isonychia bicolor</i> , a mayfly, to only one other taxon, <i>Ceriodaphnia dubia</i> , which does not typically inhabit the flowing waters where <i>Isonychia</i> are generally found. Echols et al. (2009b) also worked with <i>Isonychia bicolor</i> ; they compared laboratory-derived toxicity values for <i>Isonychia</i> with comparable values obtained from the literature for other species (Table 1), some of which were aquatic insects; and the aquatic insect species were not the most sensitive for most measures. Appendix G derives a species sensitivity distribution (SSD) for fish (Figure G-12); visual comparison of this distribution to the benthic macroinvertebrate SSDs (Figures 4-7 and 5-7) indicates fish as more sensitive throughout most of the SC range. I am not saying the statement in question is in error as I have not looked into that topic with depth. I am saying that statement is of great significance relative to the regulatory program proposed by this document; and the statement is poorly supported as currently presented.	
	3-5	Section 3.1.1.2	Selection and Adequacy of Data Sets: An additional selection criterion should be that observed SC levels should be well distributed over the population of streams used for the analysis, when those streams are stratified using measured characteristics.	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	3-6	13-14	“As a general rule of thumb, the minimum sample size to estimate an XC <sub>95</sub> using this field-based method is 25 observations of the genus in the region.” Suggest that this rule of thumb be investigated further to determine if minimum number of observations should be expressed alternatively as a fraction of dataset size. If an SSD dataset were to include 500 samples, 25 observations would constitute 5% of the dataset; but if the SSD dataset were to include 2500 samples, the 25 observations would constitute only 1% of the dataset. Does 25 observations of a taxon remain as an adequate number as dataset size increases?	
	4-11	7-9	“Samples collected from the WVDEP-identified reference sites indicate that conductivity levels are generally low and similar throughout the year, although slightly higher in summer/fall months of August, September, and October ...” My interpretation of “slightly higher” is not consistent with its use in this sentence. Figures 4-2 and 4-4 indicates that mean SC during the 3 months listed is >2x the mean SC during most other months.	
	5-12	1-3	Same comment as for page 4-11, lines 7-9.	
	7-1 and forward	--	Reference formatting is inconsistent.	
	Entire report	--	Seasonal definitions are not clear. For example, page 4-11, lines 7-9 refer to August, September, and October as “summer/fall months” while page 3-19, line 5 refers to the July-October period as “summer.”	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	--	--	<p><b>References Cited</b></p> <p>Agouridis, C., P. Angel, T. Taylor, C. Barton, R. Warner, X. Yu, C. Wood. 2012. Water quality characteristics of discharge from reforested loose dumped mine spoil in eastern Kentucky. <i>Journal of Environmental Quality</i> 41:454–468.</p> <p>Bernhardt E.S., B.D. Lutz, R.S. King, J.P. Fay, C.E. Carter, A.M. Helton, D. Campagna, J. Amos. 2012. How many mountains can we mine? Assessing the regional degradation of central Appalachian rivers by surface coal mining. <i>Environmental Science and Technology</i> 46:8115--8122.</p> <p>Boehme E.A. 2013. Temporal dynamics of benthic macroinvertebrate communities and their response to elevated specific conductance in headwater streams of the Appalachian coalfields. M.S. Thesis, Virginia Tech.</p> <p>Boehme E.A., S.H. Schoenholtz, C.E. Zipper, D.J. Soucek, A.J. Timpano. 2013. Benthic macroinvertebrate community temporal dynamics and their response to elevated specific conductance in Appalachian coalfield headwater streams. P. 7-22 in: 2013 Powell River Project Research and Education Program Reports. Virginia Tech. <a href="http://www.prp.cses.vt.edu/Reports_13/Reports_13.html">http://www.prp.cses.vt.edu/Reports_13/Reports_13.html</a></p> <p>Bryant, G., S. McPhilliamy, H. Childers. 2002. A Survey of the Water Quality of Streams in the Primary Region of the Mountaintop/Valley Fill Coal Mining, October 1999 to January 2001, USEPA Region III, Wheeling, WV.</p> <p>Daniels W., Z. Orndorff, M. Eick, C.E. Zipper. Predicting TDS release from Appalachian mine spoils. p. 275-285. In: J.R.Craynon (ed.), <i>Environmental Considerations in Energy Production</i>. Society for Mining, Metallurgy, and Exploration. Englewood, CO.</p> <p>De Jong, G.D. and S.P. Canton. 2013. Presence of long-lived taxa and hydrologic permanence. <i>Journal of Freshwater Ecology</i> 28(2): 277-282.</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	--	--	<p>del Rosario RB, Resh VH (2000) Invertebrates in intermittent and perennial streams: is the hyporheic zone a refuge from drying? J N Am Benthol Soc 19:680–696</p> <p>Delucchi, C. M. 1988. Comparison of community structure among streams with different temporal flow regimes. Canadian Journal of Zoology 66: 579-586.</p> <p>Evans D.M., C.E. Zipper, P.F. Donovan, W.L. Daniels. 2014. Long-term trends of specific conductance in waters discharged by coal-mine valley fills in central Appalachia, USA. Journal of the American Water Resources Association 50: DOI: 10.1111/jawr.12198</p> <p>Feminella, J.W. 1996. Comparison of benthic macroinvertebrate assemblages in small streams along a gradient of flow permanence. Journal of the North American Benthological Society 15: 651-669.</p> <p>Fritz K.M., S. Fulton, B.R. Johnson, C.D. Barton, J.D. Jack, D.A. Word, &amp; R.A. Burke. 2010. Structural and functional characteristics of natural and constructed channels draining a reclaimed mountaintop removal and valley fill coal mine. Journal of the North American Benthological Society. 29:673-689.</p> <p>Griffith M.B. 2014. Natural variation and current reference for specific conductivity and major ions in wadeable streams of the coterminous U.S. Freshwater Sciences 33: 1-17.</p> <p>Grubaugh J.W, J.B. Wallace, E.S. Houston. 1996. Longitudinal changes of macroinvertebrate communities along an Appalachian stream continuum. Can J Fish Aquat Sci 53:896–909</p> <p>Grubbs S.A. 2010. Influence of flow permanence on headwater macroinvertebrate communities in a Cumberland Plateau watershed, USA. Aquatic Ecology. 45: 185-195.</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	--	--	<p>Helsel D.R., R.M. Hirsch. 2002. Statistical Methods in Water Resources. U.S. Geological Survey. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation Chapter A3.</p> <p>Hem J.D. 1989. Study and Interpretation of the Chemical Characteristics of Natural Water. U.S. Geological Survey, Water Supply Paper 2254.</p> <p>Lindberg, T. T., E. S. Bernhardt, R. Bier, A. M. Helton, R. B. Merola, A. Vengosh, and R. T. Di Giulio. 2011. Cumulative impacts of mountaintop mining on an Appalachian watershed. Proceedings of the National Academy of Sciences 108:20929-20934, with online supporting information.</p> <p>Mount, D. R., J. M. Gulley, J. R. Hockett, T. D. Garrison, &amp; J. M. Evans. 1997. Statistical models to predict the toxicity of major ions to <i>Ceriodaphnia dubia</i>, <i>Daphnia magna</i>, and fathead minnows (<i>Pimephales promelas</i>). Environmental Toxicology and Chemistry 16:2009-2019.</p> <p>Odenhimer J.L. 2013. Determining a Total Dissolved Solids Release Index from Overburden in Appalachian Coal Fields. M.S. Thesis, West Virginia University.</p> <p>Odenheimer J., J. Skousen, L.M. McDonald. 2013. Predicting total dissolved solids release from overburden in Appalachian coal fields. In: J.R. Craynon (ed.). Environmental Considerations in Energy Production. Society for Mining, Metallurgy, and Exploration. Englewood, CO.</p> <p>Orndorff Z.W., W.L. Daniels WL, M. Beck, M.J. Eick. 2010. Leaching potentials of coal spoil and refuse: Acid-base interactions and electrical conductivity pp 736–766. In: Barnhisel RI (ed.), Proc Am Soc Min Reclam Ann Meetings, Pittsburgh, PA. 5–11 Jun. 2010. Amer Soc Mining &amp; Rec</p>	

**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	--	--	<p>Pond, G.J., M.E. Passmore, F.A. Borsuk, L. Reynolds, &amp; C.J. Rose. 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus level macroinvertebrate bioassessment tools. <i>Journal of the North American Benthological Society</i> 27:717-737.</p> <p>Pond, G.J., M.E. Passmore, N.D. Pointon, J.K. Felbinger, C.A.Walker, K.J.G. Krock, J.B. Fulton, W.L. Nash. 2014. Long-term impacts on macroinvertebrates downstream of reclaimed mountaintop mining valley fills in central Appalachia. <i>Environmental Management</i>. DOI 10.1007/s00267-014-0319-6</p> <p>Price K., A. Suski, J. McGarvie, B. Beasley, J.S. Richardson. 2003. Communities of aquatic insects of old-growth and clearcut coastal headwater streams of varying flow persistence. <i>Can J For Res</i> 33:1416–1432</p> <p>Sena K.L. 2014. Influence of Spoil Type on Afforestation Success and Hydrochemical Function on a Surface Coal Mine in Eastern Kentucky. M.S. thesis, University of Kentucky.</p> <p>Sena K., C. Barton, P. Angel, C. Agouridis, R. Warner. 2014. Influence of spoil type on chemistry and hydrology of interflow on a surface coal mine in the eastern US coalfield. <i>Water, Air, &amp; Soil Pollution</i> 225: 1-14.</p> <p>Stout B., J.B. Wallace. 2003. A Survey of Eight Major Aquatic Insect Orders Associated with Small Headwater Streams Subject to Valley Fills from Mountaintop Mining. Appendix in <i>Mountaintop Mining/Valley Fills in Appalachia. Final Programmatic Environmental Impact Statement</i>. U.S. Environmental Protection Agency, Philadelphia, PA</p> <p>Timpano A.J., S.H. Schoenholtz, D.J. Soucek, C.E. Zipper. 2014. Salinity as a limiting factor for biological condition in mining influenced central Appalachian headwater streams. <i>Journal of the American Water Resources Association</i>. DOI: 10.1111/jawr.12247</p>	



**TABLE 7. SPECIFIC OBSERVATIONS ON THE DOCUMENT**

Reviewer	Page	Line	COMMENT	RESPONSE
Reviewer 5 (continued)	--	--	<p>Timpano A.J., D. Soucek, S. Schoenholtz, C. Zipper. May 2013. Continuous conductivity monitoring for predicting macroinvertebrate community structure in coal mining-influenced streams. Society for Freshwater Science 2013 Annual Meeting, 19-23 May, Jacksonville, Florida. Abstract ID 7546.</p> <p>Timpano A.J., S. Schoenholtz, C. Zipper, D. Soucek. 2011. Levels of dissolved solids associated with aquatic life effects in headwater streams of Virginia's Central Appalachian coalfield region. Final report prepared for Virginia Department of Environmental Quality; Virginia Department of Mines, Minerals, and Energy; and Powell River Project. April 2011.</p> <p>U.S. EPA (Environmental Protection Agency). (2000a) Nutrient criteria technical guidance manual: Rivers and streams. Office of Water, Washington, DC. EPA/822/B-00/002.</p> <p>Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, C.E. Cushing. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37: 130-137.</p> <p>Williams D.D. 1996. Environmental constraints in temporary fresh waters and their consequences for the insect fauna. J N Am Benthol Soc 15:634-6</p>	

## APPENDIX A - INDIVIDUAL REVIEWER COMMENTS

### Comments from Reviewer 1

## Peer Review Comments on EPA's Draft Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity

### Reviewer 1

#### I. GENERAL IMPRESSIONS

It is great to see U.S. EPA developing a field-based method for establishing aquatic-life criteria for conductivity, an increasingly important stressor for freshwater ecosystems. The effects of abiotic (habitat, flow, water chemistry) and biotic factors (e.g., competition and predation) on the responses of a taxon to increased conductivity are complex and poorly understood. It is therefore sensible to use a field-based approach, rather than lab tests, to derive the criterion. I also believe that genera are the best choice of taxonomic units because the sensitivities of species from the same genus are often similar, and the identifications at this level are normally more accurate and less costly than at the species level. Clumping taxonomic data to the genus level also increases the number of taxon occurrences and makes more taxa available in a region for establishing a conductivity criterion. Overall, the document is well written. However, I have several major concerns, particularly on the concept and measure ( $XC_{95}$ ) of taxon extirpation and associated statistical analysis. The vague and inconsistent relationship between  $XC_{95}$  and extirpation appears to have compromised the rigor of the process of criterion development. I am also worried about how specific case studies in the document are used to justify extrapolation of a conductivity criterion developed for one region to another region. In addition, some terms (e.g., probability of capture) need to be more clearly defined, and equations need to be constructed in a standard format.

#### II. RESPONSE TO CHARGE QUESTIONS

##### Questions 1-3: Data Set Considerations

***1. Ion Matrix Characterization:*** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $Ca^{2+}$ ) plus magnesium ( $Mg^{2+}$ ) and the anions bicarbonate ( $HCO_3^-$ ) plus sulfate ( $SO_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[SO_4^{2-}] + [HCO_3^-] > [Cl^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[SO_4^{2-}] + [HCO_3^-] > [Cl^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).*

Comments: If NaCl in stream water is from natural sources, it would be appropriate to exclude those samples dominated by Cl<sup>-</sup>. However, if it is clearly from human activities, such as road de-icing, exclusion of the samples will make the conductivity criterion derived not applicable to NaCl

contamination, a major stressor in streams of the snow zone. It seems to make sense to include all sampling sites where  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-]$  is **naturally** greater than  $[\text{Cl}^-]$ .

**2. Catchment Size:** *All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: The analyses and discussions provided are adequate for the decision to include all stream samples regardless of catchment size. But, can the criteria developed be applied to great rivers, like Mississippi, Ohio, and Colorado rivers? These rivers support very different aquatic fauna, likely fewer sensitive genera, but some unique ones. If no large-river samples are included, could the criterion derived protect those unique taxa? Or, may the criterion be over-protective of large rivers?

**3. Seasonality:** *The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: Annual samples, particularly those collected in summer, likely miss some or many sensitive insect genera. However, as long as sampling time is NOT correlated with conductivity (e.g., sampling high-conductivity sites early, but low-conductivity sites later), this source of error is probably minor compared with other sources, such as selection of sites, sampling variability, and the temporal variation of conductivity. I would examine the correlation between site conductivity and sampling date (Julian Day).

#### **Questions 4-8: Case Studies: Example Criteria Calculations**

**4. Criterion Continuous Concentration (CCC):** *Please comment on the clarity of the method to derive the  $\text{XC}_{95}$  and  $\text{HC}_{05}$  (Section 3.1, Deriving a CCC).*

Comments: The method of deriving  $\text{HC}_{05}$  is straight forward and clearly described. However, the method used to estimate the extirpation threshold ( $\text{XC}_{95}$ ) is confusing and problematic.  $\text{XC}_{95}$  considers neither the direction of response of a genus to increased conductivity nor the relative frequency of the taxon (“probability of capture” in the text), two key factors for inferring

extirpation. Therefore, it appears not possible to establish any consistent and meaningful relationship between XC<sub>95</sub> of a genus and its extirpation. The authors use a GAM model to refine XC<sub>95</sub>. That is helpful for those genera negatively affected by conductivity, but the threshold of extirpation for those genera positively or neutrally responding to increased conductivity over the range observed still remains indefinable. For example, XC<sub>95</sub> of *Cheumatopsyche* (A-29) is estimated to be >3140µs/cm (A17), while the genus reached its highest “probability of capture” at this conductivity level. Even with a qualifying designation of “>”, is this estimate really meaningful? The same designation (>) is also given to those genera that have very different response curves, such as *Cheumatopsyche* and *Leuctra* in Fig. 3-1. When the values of XC<sub>95</sub> for genera that substantially differ in occurrence frequency and response to conductivity are treated equally, the SD curve is no longer interpretable and potentially misleading, at least in my opinion. Two options might be worth considering. First, presumably one can appropriately determine extirpation thresholds (i.e., XC<sub>95</sub> without > designation) for more than >10% of the genera. If so, he/she may put all other genera in a single category, “indefinitely high”. The authors may then use the first group of genera to define HC<sub>05</sub>. Second, the authors can look at how many genera declined down to ≤1% of the highest “probability of capture” in the max-conductivity bin in GAM models. If more than 10% or 20%, as in their case studies, they should be able to easily determine HC<sub>05</sub>, leaving out the idea of XC<sub>95</sub> entirely.

**5. Criteria Maximum Exposure Concentration (CMEC):** *The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?*

Comments: Confidence levels (e.g., 90% or 95%) can be estimated only if the frequency distribution of data is known (e.g., normal). Did the authors check the data distribution before using Eq. 3-2? Is the critical value used here for normal distribution? If confirmed, the method is reasonable.

**6. Duration:** *Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Comments: The description and justification appear adequate. I am not aware of additional publications.

**7. Frequency:** *Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*



Comments: The recovery of macroinvertebrates strongly depends on nearest sources. If the exposure occurs at a local scale, three years may be enough for re-colonization. However, if the exposure happens at some broad scales, three years may be not enough.

I am not aware of additional publications.

**8. Alternate measurement endpoint: Is the example alternate measurement endpoint ( $\text{HCO}_3^- + \text{SO}_4^{2-}$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?**

Comments: It is clear and adequately supported. This alternative endpoint explicitly identifies the stressor anions. However, it does not account for any other anions, which may be less abundant, but still significant, such as  $\text{Cl}^-$  in many freshwaters. As a result, the criterion derived would be less applicable than conductivity-based criteria. In addition, this alternative is subject to the same criticisms I made early on estimating  $\text{XC}_{95}$  and  $\text{HC}_{05}$ .

#### **Questions 9-12: Geographic Applicability**

**9. General: Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?**

Comments: The process is clearly described, but it does not seem to fully address the concerns of SAB. To apply the conductivity criteria ( $\text{HC}_{05}$ ) developed for one region to another, we need to make sure both water chemistry and macroinvertebrate fauna to be comparable. The authors did a good job for assessing water-chemistry comparability. However, their responses to SAB's other comments do not seem adequate. The authors are correct that SD does not require the same set of genera. However, it does require that the distribution of  $\text{XC}_{95}$  (at least for sensitive genera, tolerant ones are not used anyway) in the new area is similar to in the original region. Say, two regions share all genera, but if the new region happens to have more highly-sensitive genera that meet the minimum sample size (25 samples) than in the original region, its  $\text{HC}_{05}$ , if derived, may be lower, and thus the original  $\text{HC}_{05}$  would be less protective. The authors need to address the importance of biological comparability.

Yes, water quality criteria of EPA established based lab tests is applicable across the nation or multiple regions. This is because a standard lab procedure is used (test species and experimental setting). Here, we do not have a standard set of genera with the same occurrence frequencies and same environmental conditions to derive a universal  $\text{HC}_{05}$ . Extrapolating conductivity criteria to beyond the original region may be risky even if water chemistry is comparable, as I argued above.

***10. Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

Comments: The process described is clear (Section 3.6), and the case study analyses support the limited extrapolation. However, I am not clear why one would not include samples from the whole ecoregion in the criterion development at the first place. Even if one ecoregion includes streams in more than one state, it appears much easier to combine raw data from all states involved, standardize them (e.g., sub-sample size), and then develop the region-wide criteria, than to rely on extrapolation.

***11. Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

Comments: The applicability analysis is clearly described, but again I am not fully convinced that the approach is sufficient for the reason described in my comments on Question 9. The case study appears to support the approach, but the new ecoregion (68) is just next to the original one (69). The result may differ substantially if the new ecoregion is further away and associated with very different benthic fauna. It is difficult to generalize the effectiveness of this background-matching approach based on this special case study.

***12. Applicability to ephemeral streams: In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?***

Comments: Ephemeral streams typically support highly-mobile taxa (e.g., beetles) and taxa of short life cycles (e.g., some chironomids). They may collectively share most genera with perennial streams, as shown by Grubbs (2010), but the occurrence frequencies and abundance of most shared taxa are most likely to be much lower. My experience is that not many sensitive genera live in

temporal streams. As a result, these streams may be “over-protected” by a HC<sub>05</sub> established for perennial streams. However, I agree that they are important components of stream networks, but should be protected by separate conductivity criteria. Temporal streams have recently attracted much research interest.

Following references are relevant:

- Lake, P.S. 2011. *Drought and aquatic ecosystems: effects and responses*. Chichester, UK. Wiley-Blackwell.
- Steward, A.L., D. von Schiller, K. Tockner, J.C. Marshall, and S.E. Bunn 2012. When the river runs dry: human and ecological values of dry riverbeds. *Frontiers in Ecology and the Environment* **10**: 202–209
- Williams, D.D. 2006. *The biology of temporary waters*. New York, NY: Oxford University Press.

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

***13. General: The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.***

Comments: The process is clearly described. I have a number of concerns. First, the fish sampling methods (sampling gears and distance) used by different agencies/programs are not detailed, but likely inconsistent. For example, if one set of samples were collected over a reach of 40-time channel width, but another over 20-time channel width, the former likely capture more species. As a result, the occurrence frequency of a species may vary with sampling method or data sources, introducing noise into the analysis. The sample comparability of the various sources needs to be evaluated. Second, the sample sub-setting based on major basins is well defensible, but much less so when based on stream size. Although many fish species prefer streams of certain sizes, they also occur in streams of different sizes. Any numerical thresholds seem arbitrary. The modification to the bootstrapping process is reasonable; however, my earlier criticisms to XC<sub>95</sub> estimation, its relevancy to species extirpation and SD curves are applicable here.

***14. Protection: Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.***

Comments: Just as the case studies for macroinvertebrates, the conductivity criterion derived here appears to be protective for fish in the study region in practice. However, EPA needs to address the lack of biological relevancy of XC<sub>95</sub> to species extirpation and vague interpretation of SD curves, as I described earlier.

### III. SPECIFIC OBSERVATIONS

Page	Line	Comment or Question
xviii	--	On definition of XC <sub>95</sub> . It is not clear what do the authors mean by “effectively absent” here.
xix	--	First, this definition of extirpation is very different from what commonly used in the literature of conservation biology (absent from a region or regions, but still present somewhere else), and then may cause confusion. Second, it is also hard to determine when a genus is no longer a viable resource or unlikely to fulfill its functions, particularly when only 200 individuals are counted from a site.
1-1	8	“ . . waters dominated by . . . ” Do the authors mean “waters naturally dominated by”? If not, one would not be able to apply the method to streams contaminated, say, by NaCl from road deicing. Clarify.
1-2	11-12	If any studies/data show Ecoregion III effectively capture the natural variation of conductivity across space, cite them. If not, the authors need to justify this decision.
2-1	9-11	Here the authors set the threshed of conductivity for extirpation as the level below which 95% observations occur. Above this conductivity level, the taxon is assumed to be no longer a viable resource or unlikely to fulfill its function. However, how they actually did this is much more complex (P3-13). They can leave the details to later, but they need to give readers some idea about how it is actually done. Otherwise, one may reject their method right away because above this threshold, a genus may be still common and viable!
2-10	Figure 2-1	Does increase of ion concentration always lead to decline of macroinvertebrate and fish species? I thought that the relative or true abundance of tolerant species may increase, just as shown in the case studies (B19-30). Modify.
2-15	21-22	See my earlier comments regarding species extirpation.
2-17	22	The authors need a newer citation, if available.
2-19	1	Replace “many states” with “most states”(?)
2-19	17	“Freshwater insects are among the most sensitive. . .”. This statement is too general. Freshwater insects differ greatly in their sensitivity to human disturbances. Most EPT species are sensitive to organic pollution, but most chironomid genera are tolerant and so are most other dipterans (true flies). Modify.
2-20	17-18	“In other words, . . . value.” This threshold may make sense if a taxon rapidly decreases with conductivity; however, it makes no sense if a taxon increases or does not change with conductivity over the range observed. The authors addressed this issue on page 3-13, however, they need to bring their arguments and solutions up here. The authors also state that “In other words, the probability of 0.05 that an observation of a genus occurs above its XC <sub>95</sub> conductivity value.” Does this statement really hold when observations in a large bin are down-weighted in calculating XC <sub>95</sub> ? Even if this statement is valid, it is still confusing.



Page	Line	Comment or Question
		Readers might interpret it as that one should expect to capture a genus at 5 of 100 sites where conductivity is greater than its XC <sub>95</sub> . The authors need to give a clearer and biological meaningful interpretation of XC <sub>95</sub> .
3-3	Figure 3-1	<p>This figure is confusing. First, when 85% of the samples contained <i>Cheumatopsyche</i> in the bin with conductivity &gt;1000 µs/cm, how is the genus assumed to extirpate to close to do so? Second, even when the “probability of capture” of a genus declined with increasing conductivity, the taxon may be still be common at its XC<sub>95</sub>, approximately 20-40% of the samples for <i>Stenonem</i> and <i>Leuctra</i>, respectively in the case study. Considering the impact detectability of a genus associated with small sample size (200 counts) and limited sampling period (once a year), the probability of occurrence could be much higher. One can argue that both genera are strong, at least far away from extirpation. It appears difficult, if possible at all, to consistently relate the extirpation of a genus to its XC<sub>95</sub>. The authors addressed this issue later in pages 3-12 and 3-13, but they need to give a full treatment when interpreting the figure or when introducing XC<sub>95</sub> in P2-20.</p> <p>I am also concerned about their use of the term, “probability of capture”. In the literature, two terms, occurrence probability and detectability, are typically used to describe observations of a species, occurrence probability and detectability. The former describes the probability of a species to occur at a site (any spatial unit). The later refers to the probability of a species to be detected when it is present at a site. “The proportion of samples with a genus present at a conductivity level” could be taken as an estimate of occurrence probability only if detectability is assumed to be 1, something that rarely occurs in practice. I suggest the authors to use “relative frequency,” rather than probability of capture, which has been commonly used to refer to the % of individuals captured by a sample. In addition, the authors estimated the proportion of genus observation for a conductivity bin, rather than a conductivity level. Modify and clarify the terminology.</p>
3-5	20	“background . . . region,” It would be helpful to clearly state how similar conductivity among reference sites is similar enough.
3-8	8-12	One major source of salinity in freshwater waters snow zone is road de-icing. The conductivity criteria described here will not be applicable to assess the impact of NaCl used for de-icing? (also see my earlier comment on this issue)
3-10	26-31	See my comments on the relationship between XC <sub>95</sub> and extirpation earlier.
3-11	1	Did the authors assess how bin delineation affected CDF? The description here is a bit vague regarding how they balanced the number of bins and the size of bins. Clarify.
3-11	eq 3-1	“x” needs to be defined.

Page	Line	Comment or Question
3-12	Figure 3-3	Replace “cumulative probability” with “cumulative proportion” to be consistent with the text (L2), and to avoid the issue of imperfect detectability. I also suggest adding a third panel figure to show a concave increasing curve like the one for <i>Corydalis</i> (B-42). This type of curve really means a positive response of a genus to conductivity (B-29 for the same genus). For the 2 <sup>nd</sup> ( <i>Nigronia</i> ) and 3 <sup>rd</sup> types of curves ( <i>Corydalis</i> ), it is not possible to relate XC <sub>95</sub> to extirpation, as I argued earlier. Glad to see the authors starting to address this critical issue here. However, the issue needs to be fully treated much earlier. I also do not see it being a data-distribution issue, but a fundamental limitation of CDF. CDF curves of Types 2-3 are also not anomalies, but they are normal and frequent, as shown in the case studies. Revise.
3-13	1-7	Yes, the qualifying designation helps for understanding HC <sub>05</sub> , but relating XC <sub>95</sub> to extirpation remains conceptually flawed. See my comments to Charge Question 4 for possible options.
3-13	21-22	Replace “mean curve” with “fitted curve”. Also, what is the confidence limit? 95% or 90%? Clarify.
3-17	1-3	A further concern is whether sampling dates/period is related to conductivity. If most high-conductivity sites were sampled in spring (March-June), but low-conductivity sites in summer, one likely underestimates the occurrences of sensitive taxa in the latter and then overestimate HC <sub>05</sub> conductivity. Correlation between conductivity and sampling time can be used to identify the bias.
3-19	eq 3-2	This equation needs to be written in a standard math format as follows: $CMEC = 10^{(\bar{X} + z_{\alpha} \times \sigma_r)}$
3-21	10	“ . . and often more than 4 days”. Above CCC? Clarify.
3-19	11-14	Is “the one-tail critical value” half of the number of the standard deviation required for 90% confidence limits? The authors also define $\bar{X}$ twice here and differently. Clarify or correct.
3-26	27	“More than 90% of . . insects.” This statement is too broad. In many streams, insects took less than 90% of all individuals. Add “often” or “frequently”.
F-16	eq F-1	Re-write the equation in a standard math format
G-3	7-8	This statement is too broad. Many adult insects, such as winter stoneflies, actually only move over a short distance.
G-10	3	Add “hybrids” after “immature specimen”
G-18	14	Do the authors mean a selected <b>minimum</b> size ranging from 0-60 occurrence? If so, how it can be zero? Clarify.



### Comments from Reviewer 2

## Peer Review Comments on EPA's Draft Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity

Reviewer 2

### I. GENERAL IMPRESSIONS

It was a pleasure to review the US EPA draft document, *Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity*. As a biologist who has worked both in the field and in managerial capacity for water quality monitoring and bioassessment, the area of specific conductivity has long been more or less overlooked. This is largely due to two reasons: 1) most city, watershed and state monitoring programs don't quite know what to do with conductivity data, and 2) the role of conductivity in its effect on aquatic and benthic organisms is not well understood. The need for a criterion is basic to the improvement in these areas. This document's clear and strong guidance in providing a method for development of conductivity criterion and for a method to make it applicable to adjoining regions is an immensely valuable new and long over-due tool for monitoring programs. The document provides considerable information on the effects of high conductivity levels on macroinvertebrates and fish, and provides strong, data-supported rationale for its approaches and methods. Very large data sets, paired analyses, and strong/reliable/widely used statistical models were used in all of the analytical processes. The biological information in all sections was especially accurate, thorough, and clearly written. It was a pleasure to read those sections and to learn new information. The statistical material was less clear for me, but that is more of a deficiency on my part than that of the document. In that regard, perhaps more explanation of several of the calculation processes could be provided, and a full, working example of each (probably placed in the Appendices), would be helpful for water quality staff who have limited statistical training. With such examples, staff could follow the step-by-step process. I realize this might be viewed by the authors as somewhat of an unnecessary effort, but I believe it would help the document's usability by a greater number of staff with varying backgrounds and knowledge base.

Thank you for the opportunity to review this excellent document. The document is well done and its conclusions correct. I believe it will be a valuable guidance for development of much needed criterion for conductivity. It will provide an immensely important function in the improvement of water quality and ecological health for the nation's streams and rivers.

### II. RESPONSE TO CHARGE QUESTIONS

#### Questions 1-3: Data Set Considerations

**1. Ion Matrix Characterization:** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set*

*(e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[SO_4^{2-}] + [HCO_3^-] > [Cl^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).*

Comments: I believe it is appropriate to remove samples (data) from the data set which might move the results from reflecting the true condition. The more “types” of data included in a database, outliers, for example, the more general/less specific will be the results – and therefore, less accurate.

The question for this study was “how to derive example criteria for conductivity for flowing waters dominated by calcium, magnesium, sulfate and bicarbonate ions” (pg. xvi), and not for flowing waters dominated by chloride ions. All of these ions are predominant throughout the study’s ecoregions because of the geology, physiography, vegetation, animal life, climate, soils, water quality, and hydrology found here. However, calcium, magnesium, sulfate and bicarbonate come from weathering of limestone and dolomite (the geological composition of this region) and are the ions which have the greatest impact on specific conductivity which is the intent of this study. Although chloride ions are also prevalent, the decision to exclude chloride anions is logical and appropriate.

Additionally, the decision to exclude sample sites with  $< 6$  pH is also probably wise, although this is perhaps less definitive. Acidity directly affects conductivity by causing calcium and magnesium to become more mobile with decreasing pH, thus having a clear role in conductivity levels. But the level of its effect and its associated variables – such as temperature – would then also need to be considered, increasing the study’s data needs and broadening the question. On the converse, acidic conditions do exist in waters of this geographical region because of anthropogenic influences such as urban stormwater runoff, surface mining runoff, gas/oil extraction waste water, and aerial deposition. And from this standpoint only, there might be adequate justification for its inclusion. However, since  $< 6$  pH waters in this study were not large in number, the decision to include or exclude could go either way. Would their inclusion have had much influence? A basic rule of thumb for most scientific studies is “the more specific the testing or measuring, the more specific and accurate will be the results.”

Toxicities of ions differ, and keeping the data collection and the subsequent analyses limited to the four ions ensures data results free of the additional variables inherently associated with any additional ions. Any field-based study should limit its parameters of study for this reason. Samples from waters with only the same ionic composition will yield the most representative and accurate results.

The authors point out (pg. 2-11) that the relative concentration of bicarbonate is pH dependent, and that the dominant form of the ion in soil is bicarbonate at circumneutral pH. This gives further justification for limiting collection of samples to waters with  $> 6.0$  pH.

The authors have done an excellent job in discussing the many factors and general background information in Sections 1, 2, and 3. The discussion in these sections is valuable for the reader, and is thorough and clear in its presentation. The background information is presented objectively and will be helpful and adequate for state water quality staff.

If there is concern about the merits of keeping or excluding data, I recommend the question be directly discussed in the two introductory chapters. Even though the authors have discussed the reasons why they excluded chlorine and acidic conditions (actually multiple times throughout the document), a table that straightforwardly addresses the pros and cons could be included. List the pros and cons for excluding chloride and sites with  $< 6.0$  pH is my recommendation.

I agree with the inclusion of all other data, i.e., impaired and high quality streams, all stream sizes, and sampling from all seasons.

***2. Catchment Size: All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.***

Comments: It is excellent that all stream types and sizes were included in the data sets for the case studies, especially the smaller and intermittent streams. Smaller streams, both perennial and intermittent, are where valuable macroinvertebrate habitat is most often found. These are likely to have the appropriate streambed composition, rocks and logs for colonization, leaf litter, bank overhangs, and freedom of siltation – which are all crucial for macroinvertebrate life cycles, population abundance and diversity. So often only the larger, perennial streams and rivers are studied. The authors are “right on” when they point out that discharge from headwaters, intermittent and even ephemeral streams ultimately affect downstream stream reaches and rivers. This is often not understood or realized fully by program managers, who are not well versed in stream ecology, and policy makers. Additionally, the authors make an important point in that many macroinvertebrate taxa often use temporary streams for at least a portion of their life cycle. Much of my experience in stream ecology and water quality has been with the smaller streams and it is my belief that their value to the river system and its taxa cannot be over-emphasized. I thank the authors for their recognition of this.

Exclusion of data from the larger catchment areas is, however, worthy of a little discussion here. The authors present four good reasons for not excluding them: 1) sensitive genera were found in the larger rivers; 2) inclusion of data from larger rivers did not significantly change the magnitude of the hazardous concentration; 3) Analysis of 3115 sites with drainage areas up to 17,986 sq km showed a very weak (a very weak, indeed!) correlation of conductivity and drainage area; and 4)

background conductivity estimates for drainages > 155 sq km were within confidence bounds for establishment of background values.

However, the EPA's Benchmark Report initial exclusion of larger streams – because sampling methods might differ for non-wadeable streams – has substantial merit. Sampling methods are indeed different for the larger rivers, and large river sampling requires greater resources (time, staff, boat/equipment) and, therefore, also happens less frequently. Collected macroinvertebrates in larger rivers can be low in numbers as well – due probably to a combination of factors: manmade channel morphology changes, river velocity too high, fewer colonization sites, poor habitat, deposition of sediment, anthropogenic contaminants, and difficulty in sampling at greater depths and velocities. Thus, more variability likely exists in data for macroinvertebrate databases for large rivers. However in this study, sensitive taxa were documented in the larger rivers, so perhaps collection methods and expertise in sampling has improved, but perhaps more importantly, these rivers are likely of higher quality than those here in the Midwest of which I am familiar and which are heavily impacted by agriculture.

In conclusion, the authors have provided good discussion and support for the decision to include all samples regardless of catchment size. A bit more discussion as I have presented here might be helpful but probably is not necessary.

Lastly, I wish to reiterate the value of data from intermittent and ephemeral streams. These small streams provide irreplaceable habitat for macroinvertebrates, invertebrates, amphibians, aquatic/wet terrestrial species of all kinds. Their loss has been significant through ditching and tiling in agriculture, diverting and damming for irrigation, and in placing into underground pipes in urban development.

***3. Seasonality: The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.***

Comments: The data of conductivity concentrations show that they do vary by season. This was addressed by comparing hazardous concentration values by season. “Due to the similarity at the low end of the sensitivity distribution (SD) between spring HC<sub>05</sub> and HC<sub>05</sub> of the full dataset” (pg.4-11), it was determined to use all data regardless of month. I question why this wasn't also done for the fall (especially October) data? Granted, February – April exhibited the most noticeable change but October was significant as well. In Ecoregion 70 from the Watershed Assessment Branch database, September stood-out because it had significantly higher conductivity values (pg. 5-7), as did April with definitely lower values (pg. 5-8), although not as extensive as October's. The box plot on pg. 5-9 for Ecoregion 70 shows the apparent seasonal variation of July – October.



Reference sites, however, are stated to have conductivity levels “generally low and similar throughout the year although slightly higher in August, September and October,” (pg. 5-12). I think it is more than just “slightly” higher! On pg. 5-7, Ecoregion 70, the September values are so much higher that it is difficult for me to understand that the September data doesn’t adversely skew the results. Perhaps separate CCC and CMEC for the September timeframe is reasonable. Since my area of expertise is not statistics, I am not really able to investigate this myself and will rely on the authors’ determination that seasonal differences do not require weighting, and that the seasonal differences do not alter the results to any great degree.

On pg. 5-6, the conductivity background for Ecoregion 70 is <200 uS/cm December – June, and >200 uS/cm July – October. This seems to be enough of a distinction that perhaps all data need to be divided into two sets, one containing the December – June data and the second, the July – October data. It would seem that this would be sufficient rationale to have this separation but I am presenting this more as a question than a statement.

As a side personal note: Here in the Midwest we have distinct seasons, and many parameters clearly show this in their values. I am accustomed to looking at the seasonal data and its use in planning for monitoring programs and watershed recovery plans. Having this specificity of data is more informative for these purposes than “lumping” or weighting of the data because it provides greater insight as to pollutant sources and causal relationships. For state staff, determination of sources of impairment is usually the overall objective and is frequently difficult to ascertain. Having a clear understanding of what is happening each month (when there is monthly data available) helps to provide insight. With that noted, I fully realize the objectives for those purposes and the objectives for this study are different. But it may be of value to the authors to understand how state WQ staff usually look at data and use it.

Additionally, with these comments in mind, I must also add that I prefer limiting the amount of weighting when working with a dataset. On pg. 3-18 it is stated that if “the weighted HC<sub>05</sub> overlap the confidence bounds of un-weighted HC<sub>05</sub>, the un-weighted model is accepted.” This seems to be a logical and accurate decision. Further, it states that in general, “the use of unweighted SDs is easier and requires fewer data points.” I agree. Where weighting and manipulating the data can be reasonably minimized, I believe it should be. A balance must be made in the need for normalizing, scaling and weighting and the loss of variations that reflect the actual conditions.

Also pgs. 3-16 – 3.18, the three approaches to seasonality are given. It is well done.

#### **Questions 4-8: Case Studies: Example Criteria Calculations**

***4. Criterion Continuous Concentration (CCC): Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).***

Comments: As my knowledge base is centered on biological aspects of rivers and streams rather than statistics, several of my comments will be limited in this regard. I am listing the various thoughts which I had as I went through Section 3.1:



- The inclusion of both high quality and impaired sites is correctly done. This provides a well-represented database, covering all levels of conditions and taxa, and at all times of the year. This reflects the variability that will exist because of seasons, habitats, and the effects of manmade influences in the river basin which affects the ionic composition.
- The step-by-step explanation on pg.3-1 is helpful. More of this could be done to increase understanding of the calculation processes used in this document.
- Improve clarity by explaining how the actual weighting and cumulative distributional function is done on pg. 3-1.
- Good explanation on pg. 3-2.
- Figure 3-2, pg. 3-4: Gives good general process flow. Is it possible that an actual mathematical example could follow along with each step?
- The bullets on pg. 3-5 are thorough and give good support to adequacy of data. Sample size discussion is well done. Sensitivity analysis, which includes a representative proportion of sensitive genera, is well done. Having 90-120 genera and 500-800 sites are large numbers, and are seen throughout this document. This is excellent. It strengthens the development of the criterion, its applicability, and the justification of the concentrations determined. If only all studies could have such numbers!
- Bootstrapping needs to be described more fully (for non-statistical readers). While the paragraph on pg. 3-7 is probably adequate for many, there are a considerable number of state agency or other watershed staff who have minimal statistical backgrounds. A few additional paragraphs detailing/giving examples of such exercises as bootstrapping would make the document more usable by the large range of agency staff.
- In reference to pg. 3-9, lines 9-21, care must be taken to avoid too many repeated macroinvertebrate samplings in the same place over the course of a year. Repeated sampling is disruptive to the habitat and can diminish the taxa at the site. Unlike fish species, macroinvertebrates are less mobile, and, if young stages are removed, there may be fewer adults at the sites especially if there are no other small streams in the vicinity to repopulate.
- I would like to see a little more specificity in describing sampling methods. Is there assurance that there was a standardized field sampling protocol observed for all biological sampling? It is important that all sampling crews used the same techniques. It is more of a problem between jurisdictions (states, cities, or private organizations which do monitoring) but can also occur within an agency. It is vital that, for example, an equal number of sweeps of the catch net are made at each site, or, the same number of individual samples comprise a composite.
- I see that my thoughts in the above bullet is addressed on the next page (pg.3-16), lines 8-15.
- The use of different protocols by different organizations and agencies is a very real concern to any large database that has merged several smaller data sets. It probably is one of the biggest and most pervasive problems. The importance of initial training, repeated review throughout the monitoring season, and dedicated adherence to the field sampling quality control document can't be overemphasized.  
The authors have (gratefully) recognized this problem and have provided how to address this: by comparing all-year HC values from one region to that of another comparable region. If the datasets have a large number of data points, I believe this would be an acceptable way to handle this.

- I'm not sure that I fully understand the third approach to seasonal variability in Section 3.1.4 (Assessing Seasonality, Life History, and Sampling Methods). However I do believe that the authors have done well in going step-by-step in their presenting of the third approach.

**5. Criteria Maximum Exposure Concentration (CMEC): The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?**

Comments:

**a) Does the analysis to derive the maximum exposure concentration (with temporal resolutions) characterize the maximum concentration that will result in meeting the CCC 90% of the time?**

I can only provide comment in a limited manner. The annual geometric mean is appropriate for comparing different values and finding a central tendency or typical values for a set of numbers. It normalizes the ranges and removes the effect of large differences so that no one particular range of values dominates the weighting. This is appropriate for the intent of the calculations in this section/document. However, because I am not proficient in this, I am less sure of the maximum condition at any given station can be established by incorporating among-station and within-station variability. To achieve this, wouldn't the sampling sites and their particular data points need to be central in tendency and not exhibit values at the further reaches of the ranges? How was the 90% determined – review of that for the reader would be helpful.

**b) Is it reasonable to expect it to be a protective upper limit for sites in the data set?** Yes, I think it is appropriate for determining the upper limit. 90% is definitely a protective level. Indeed, there will likely be certain interests in watersheds who will contend that this is too stringent. However, based on the sensitive genera and maximum exposure concentrations found in this document, the data (and thus, the rationale) for establishing these levels is very strong and definitive. Using the paired analyses (daily measurements of conductivity paired with macroinvertebrate sampling) is a very strong statistical test and widely used in biological and environmental studies.

**c) What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?** As I have mentioned the annual geometric mean and paired analyses are strong points. The subset of frequently sampled sites is a critical element. It would seem to me that it would be important that these are clearly representative of the majority of the sites, or does the annual geometric mean make this an unnecessary concern?

The sampling of at least six times is also an important feature. I fully support the use of six times per year per site. I would increase the *n* to two in the spring (March – May) and two in the fall (Aug – Oct) and leave the remaining two for one in the summer and one in the winter. Greatest changes occur in the spring and fall months and therefore each warrant another sampling event to help capture this variability. Even with six samples, standard deviation will likely be high, especially if there are considerable differences between the sites, and even within the sites if

weather, etc. are quite variable. Lastly, as much as one would like to have repeat sampling six times/site, it is often beyond the budget of many state 305(b), 303(d), and TMDL programs. Perhaps federal support can be made available for state criterion development.

The flow chart in Figure 3-6 is helpful for overall process steps. But perhaps a working example of this could be placed in the Appendix and referenced here. I think having an example would be especially helpful to state water quality staff.

In keeping with the above, I would suggest greater description of LOESS and a full example. Although such processes as LOESS and bootstrapping are familiar to tacticians and to those who conduct these analyses regularly, many workers in the field of water quality programs haven't as much familiarity.

**6. Duration: Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?**

Comments: This approach relies “directly on paired in-situ measurements of conductivity and benthic macroinvertebrate assemblage composition,” pg. 3-20, a very reliable analysis test. Macroinvertebrates are indeed exposed to quite different conductivity levels throughout the year.

The authors are quite correct that with only annual sampling, “it may be difficult to determine precisely how long conductivity levels can be above the CCC before extirpation...” (pg. 3-21, line 12-13) occurs. I would say it is most difficult and next to impossible to tell from one sample. Sampling only once is the reality, however, of many state bioassessment sampling programs. Nothing is better than having repeated (in the field) sampling for each site. Depending upon only one sample per year is what state programs would like to avoid but in many cases, it is all that they have. So from this standpoint, the approach seems to take this into consideration and makes sense. Lastly, lines 12-13 appear to support the argument of using only one sample/year as the basis to determine duration of CCC and CMEC. In general I believe that the authors have worked hard to provide adequate description and justification for the duration of the CCC and CMEC. The description and justification for the approach on pg. 3-22 to 3-23, line 1-16 and lines 1-16 is excellent. This is very well done.

On a side note, is there a tag or footnote which could indicate that a data point(s) represents only one sampling per year? This would distinguish it from mean values from sites which have multiple sampling times during a year, thus allowing for all data to a dataset to be used, and yet allow the reader to know that some data are single data points and others are mean/geometric means. Seems this would be in the best interest for states wherein multiple databases are being used for criterion development or even just a single database which has some sites with only one sampling per year and some which have multiple samples. It is preferable to have more samples when possible, but it would be easy for state budget-cutters to limit sampling to just one sample per year if that is all it takes to establish criterion development. “Why sample more if only one is needed?”

Further, those interests who oppose water quality criterion in general (“infringement on private property rights”, “over-regulation for agriculture”, “costly programs for cities”, ...) would use the “only one sample per year” to justify their opposition to the criterion’s validity. The argument will be that there isn’t enough data and therefore the criterion is not based on “good science.”

**7. Frequency: Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?**

Comments: This is one of the best sections in the document! Descriptions and reasoning are exceptionally well done throughout this entire portion. The details and thoroughness are reflective of excellent biological knowledge on the part of the authors.

I only have a couple of comments: First, I am surprised at the high level of conductivity, <960 uS/cm (pg. 3-27, line 10), before extirpation of sensitive crustaceans. This seems exceptionally high. As a general rule, crustaceans, and mollusks specifically, are front line indicators of contaminants and water quality pollutants. Because water passes through them, low pH, chemicals, and excessive suspended solids and siltation are known to affect them significantly and earlier than many other aquatic organisms.

Secondly, I would have liked to have seen consideration given in the causal assessment methodology (Sec. 3.5, pg. 3-28, lines 15-20) of the relationship with “other known stressors such as metal toxicity, streambed erosion and siltation, and eutrophication.” These conditions do contribute as stressors, often co-exist during times of high conductivity, and seem to compound effects. I know from experience that during rain events and urban stormwater runoff (with increased suspended solids and accompanying high turbidity values), that conductivity also can substantially increase. A causal relationship seems to me to exist between increased turbidity and increased conductivity. This is not really addressed in the document. Do the increased conductivity values during rain events come exclusively from ions associated with concrete weathering, industrial runoff, fertilizer runoff, or, is the increase also coming from the suspended eroded soil particles (and their attachments)?

**8. Alternate measurement endpoint: Is the example alternate measurement endpoint ( $\text{HCO}_3^- + \text{SO}_4^{2-}$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?**

Comments: In general and to the best of my understanding, yes, I believe the use of an alternate measurement endpoint is written reasonably clearly and is adequately supported. In instances where I felt more description or clarity is needed, I have listed it. As in Question #4, I am going to simply list individual comments which I noted as I went through Appendix F:

- The correlation value for conductivity with the two ions is exceptionally tight (Figure 1) and provides excellent data justification for their use.

- There are an exceptionally large number of paired samples (pg. F-4, line 12)! If only all studies and monitoring programs could have such a large dataset. Distribution of sampling sites was also excellent. The large background data set and the wide range of conductivity throughout the sampling area indeed allows for sound characterization of the extirpation concentration.
- I believe that seasonal variation is less with these two ions. Except for September and December, there is greater similarity in their monthly values (pg. F-10, Figure F-5) than with the four ions. However, the table on the next page shows considerable variability. Nevertheless, the text says there was enough similarity on the low end of the genus sensitivity distribution to allow for criterion development. Is comparison of just the low end of the sensitivity distribution adequate? Certainly avoidance of extirpation for the most sensitive of the genera is the “goal” of the criterion, but do concentrations for moderately sensitive species, or, the extent of the ranges, have some role and should be discussed?
- Two areas of which more description and information might be helpful for the reader:
  - 3) pg. F-13, Figure F-6: Advantages of using log 10 to weight values
  - 4) LOWESS – pg. F-16, lines 3-6

#### **The second part of Question #8:**

There is a very close correlation with the two ions. They are prevalent and widely distributed in the ecoregions. They have similarity on the low end of the genus sensitivity distribution – thus functioning in the statistical analyses similarly to the four ions.

However, disadvantages of the two ions might include:

Measurement of individual ions is more costly and time consuming and most sampling/monitoring programs measure for conductivity routinely, even the installed in-field monitoring instruments can give continuous readout on conductivity value. Conductivity measurement is easy, quick, and inexpensive. The four ions comprising conductivity measurements are equally as widespread in distribution or perhaps more so than the two ions. Most monitoring programs only do conductivity because of limits on budgets. Are the four ions less affected by low pH values?

#### **Questions 9-12: Geographic Applicability**

**9. General:** *Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?*



Comments:

**First part of Question #9:**

Geographic applicability is approached by the background-matching method. The background conductivity of the original area and the new area should be similar. Also the ionic mixture for the background should be the same. The authors have described the elements of this very well and I believe this will be the most exacting and appropriate way to apply criterion to a new region – a vitally important facet to ensure use by all states and for a conductivity criterion to be widely implemented.

**Second part of Question #9:**

Yes. The only area I would question would be the SAB's recommendation that "consideration be given to the species composition of stream communities, which might be different in different states . . ." (pg. D-2, line 15-17). I interpret this to mean a direct, species by species comparison. However, the authors of this document have used a taxonomic sensitivity distribution model which doesn't do this, rather, it looks at a set of species/genera and how the communities in general respond to a stressor. I believe they have provided satisfactory support for their choice. My tendency towards the SAB's recommendation is because my experience lies with species' inventories and direct counts for abundance and diversity as compared to reference streams. This preference also goes back to whether one prefers to "lump" data or "split" data – a long known philosophical debate among biologists!

**Third part of Question #9:**

Yes, I believe it does. This question is similar to the first part of #9 and I really don't have anything additional to add.

***10. Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

Comments: The authors have been meticulous in setting the parameters for the study and then clearly describing in this document the process for assessing geographic applicability to locations within the same ecoregion but outside of the parent data sets. First, as they stated on pg. 3-32, most streams in an ecoregion tend to have a similar conductivity regime and ionic composition of dissolved salts. This is generally true, but exceptions do occur, and they wisely caution to have care when applying the example ecoregion criteria to any one particular stream reach. Specific changes in rock composition or feeder streams with springs can alter a particular reach. Good job in recognizing this.

Regional background conductivity is defined well on pg. 3-33, lines 7-13. Continuing, they clearly point out that for a data set from one geographic area to be applicable to another similar area, there needs to be: a) similar background conductivity levels and ion composition, and b) a comparison of the confidence intervals of the background data set of the new area to those of the original area;



confidence bounds for background estimated from the example criterion data set overlapped with the confidence bounds for the background of the rest of Ecoregion 69. The weight-of-evidence assessment for applicability of criteria to the new area adds much to the soundness of the approach. This validation of background specifically for the Ohio portion of Ecoregion 70 was done with a weight-of-evidence. A weight-of-evidence process is something which state staff are accustomed to doing and thus can identify easily with this and conduct it.

Excellent discussion on pg. 3-38, lines 14-25, of causes of and considerations when the background conductivity is greater in the new area than in the original area. I appreciated seeing such a complete listing; well thought-out and applicable. The summary of 3.6.3.4 is also well done. The discussion on pg. 5-18, Section 5.3 showed further rationale for the reliability of this process: using first through fourth-order streams (thus maintaining some uniformity of catchment size), extensive data sets, probability-based designs, methods comparable across the assessments and QA/QC. The approach has been well done.

In Section 4.3, the utilization of the background-matching approach for geographical applicability was effective in Ecoregion 69 as well as in Section 5.3 and Ecoregion 70. The new portion was estimated at the 25<sup>th</sup> percentile, comparing with the background conductivity estimates of the original set. All chloride-dominated samples were removed before estimating background conductivity, thus keeping the same ionic mixture for the new area the same as the example criteria. The importance of keeping data inputs all of the same “category” for a quality comparison assessment is more valuable and fundamental to good statistics than satisfying an approach that believes all data should be included. Thus, this answers previous questions of whether there should be exclusion of particular ions.

***11. Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

Comments: As in Question #10 and within-ecoregion, the analysis for the background-matching approach for geographic applicability to a new area in another ecoregion was well presented. The monitoring and sampling procedures (Alabama, Kentucky, Tennessee) are very clearly described, well-defined, specific, and a pleasure to read. The Results in Section 6.2 are presented point by point. It is helpful to have these points in paragraphs 1-4 on pg. 6-10 – 6-11.

Confidence intervals are greater in Ecoregion 70 than the other two regions (pg.6-12, Table 6.4). Perhaps reasons should be given for this. The difference is quite notable. Also in this table it is mentioned that the WABbase data set for Ecoregion 69 included samples without genus identification, meaning that identification was carried just to the Family level. Although it is always better to be able to key down to genus, this is not unusual. This is a problem for stream monitoring/sampling programs and will probably only get worse as fewer individuals are training in entomology.

Verification of applicability of example criterion from an original ecoregion to a new ecoregion using the background-matching method was done by independently estimating the HC<sub>05</sub>. Verification is an important step – a necessary “hurdle” – and when it is successful the method can be confirmed to be reliable and its use can proceed. I compliment the authors for presenting the first demonstration of successfully applying a criterion to a new region. This is a major step in expanding agencies’ ability to establish a criterion for a parameter which has a significant role in the health of aquatic organisms.

I do have concerns about the need for about 500 samples in order to achieve consistent results with the HC<sub>05</sub> derivation (pg. 6-13). The need for a large data set is well understood, but it may be difficult for some entities to have that many samples. The mention that there were different sampling methods, and that some methods tend to collect different types of genera, is appreciated. While it would be better to have uniform sampling methods, in reality that doesn’t always happen. The authors tried to restore confidence that a large variety of taxa were nevertheless represented. It might be wise to recheck the methods/protocols used, verifying that each method was used about equally throughout the data set.

The authors were very specific and thorough in their description on pg. 6-13. The applicability was well presented.

As I have mentioned previously, my expertise lies in the biological aspects rather than the statistical analyses. With this in mind, it would be helpful to have more information about bootstrapping and an example by which one could follow. This would be an Appendix supplement I realize, but I think it would be useful to staff in watershed programs.

***12. Applicability to ephemeral streams: In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?***

Comments: Yes, the discussion in Section 3.6.2 well supports the field-based method for applicability to ephemeral streams. The support is well defined by the publications cited. It clearly discusses that macroinvertebrates are found in intermittent and ephemeral streams. Grubbs’ (2010) research provides excellent quantifying results. In my experience, I believe the abundance and diversity of macroinvertebrates in ephemeral and intermittent streams is far greater than larger

rivers which (in this part of the country) have nearly all been channelized, straightened, trees/brush removed, island and meanders removed, streambeds laden with silt, and hydrologically modified.

I found the short discussion of the various adaptations to survive temporary dry periods (pg. 3-20, lines 27-30) to be exceptionally helpful. Very seldom is this addressed or even widely known, however, it is a significant fact among many of the taxa of ephemeral streams. It was a pleasure to see this included. And the continuing discussion of the use of upstream temporary streams for part of their life cycle (pg. 3-30 - 3-31) is also accurate and equally important to include. This fact, and the documented presence of the “vast majority (91 out of 108) of macroinvertebrate taxa were observed in both the perennial and temporary channels” (Grubbs (2010), provides strong rationale for the applicability of field-based method/criterion for conductivity to ephemeral streams. Upstream water quality conditions affect lower reaches’ aquatic life and the exposure to harmful levels of conductivity (and all other contaminants as well).

As I’ve mentioned in my response in Question #2, intermittent and ephemeral streams (even “often-wet” depressions in fields, wet meadows and pasture drainages, wet areas in riparian corridors or nearby river valleys, etc.) provide habitat for macroinvertebrates – at least for a portion of their life cycle. The value of these small streams and temporary wet areas as habitat for many taxa has not been appreciated nor understood by many property owners, developers, and policy decision-makers. The decision to include ephemeral streams in this criterion development is especially important and gratefully appreciated by biologists such as myself. The information provided here in this section also provides strong rationale for the current debate on “navigable waters” regulations.

Please refer back to my response for Question #2 for my other previous comments.

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

**13. General:** *The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.*

Comments: I believe that the work done to derive a fish HC<sub>05</sub> was exceptionally well done. Data sets were very large and a number of considerations which are fish-specific were incorporated. These provide validity to the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values. The suitability of the method – to be applicable to fish as well as macroinvertebrates – is especially reflective of the quality of this work and its usability for widespread application to aquatic organisms.

The following are thoughts and notes which I made as I progressed through the section.

- On page G-3, the section accurately makes the connection between loss of macroinvertebrates (impacted by high conductivity levels) and the stress this puts on fish by decreased food availability. This is clearly an additional justification for a conductivity criterion – fish losses reduce the natural resource quality of the stream, reduces recreational potential, increases the number of threatened and endangered species and possible extirpation.

Should other stressors be included in the above discussion? For example: poor habitat, lack of water depth diversity, high ammonia levels, high suspended solids, low DO ... all are stressful to fish either directly or indirectly. This weakens fish so that the effects of other stresses, such as the ionic imbalance from high conductivity levels, are likely enhanced. In other words, should other types of stress be taken into the analyses with conductivity because of the increased level of impact there might be?

- Of the six fish species listed as relatively tolerant of elevated conductivity, pg. G-4, lines 4-7, I believe that *Micropterus dolomieu* (smallmouth bass) perhaps shouldn't be included. It is generally intolerant of pollutants and poor water quality. Even *Lepomis cyanellus* (green sunfish) prefers somewhat good conditions, even though it can be found in eutrophic waters.
- Identification of fish to the species level is indeed the routine for fish sampling, unlike macroinvertebrates, and this does lend itself to species-level XC<sub>95</sub> values.
- The fish analyses used a combined data set for fish from portions of four contiguous ecoregions and seven states! Seven data sets collected between 1991 and 2009, 1657 sampling events across 1364 distinct sites, gives great spatial and temporal data. What an amazing quantity and variety of data! This extensive data base is difficult to find in the environmental arena. Kudos for bringing together such an excellent base from which to assess for a fish conductivity criterion.
- A concern: Reference sites were not identified in the dataset, however the document seems to imply that 134 sites, which were >90% forested, were likely such. What is the reason for not identifying, and assuring, that reference sites were included? Could the data not have been identified, perhaps as a separate grouping within the dataset? How can you be sure that there were adequate reference sites in the initial sample collections? Heavily forested does not necessarily assure that water quality will be of high quality. This exact situation happened with a monitoring program which I designed and implemented for the stream system running through Omaha, Nebraska. After careful searching, I found what appeared to be a minimally impacted small stream of which for most its length it flowed through rolling wooded hills. Although some small acreages, occasional houses, and further back, a new development were in among the hills and woods, the stream appeared to be minimally impacted and exhibited great fish and macroinvertebrate habitat. It appeared to be the perfect reference stream (there was no existing data available to use as a guide). Eventually, it became exceedingly apparent that I chose poorly as fecal coliform levels repeatedly were some of the highest of the 24 sampling sites in the system and some of the other parameters were also not appreciably better than any of the other sites.

- Description on pg. G-4, Lines 15 -28 is very thorough; provides the reader with a detailed understanding of the area, conditions, etc. In G-7, there is good use of excluding larger catchment areas, sites with low pH and high chloride levels, and sites which were too small to support fish. This strengthens the data analyses. The discussion of fish on pg. G-10 regarding the exclusion of sites where there is question of the presence or absence of a species is also good. I am surprised to read that brook trout (*Salvelinus fontinalis*) are stocked. While they are native, they are few in number and are especially sensitive to poor water quality conditions. I believe that I disagree that they should be included. I have concerns about counting any of the stocked fish species because of the possibility of affecting the XC<sub>95</sub> estimates. Stocking is a manmade “condition”, largely to improve recreational fishing; the expected life span is pretty much irrelevant and independent of the stream’s condition.
- The paragraph on pg. G-11, lines 9-15, is not fully clear to me. It would be helpful to have it explained a bit more fully. The following lines on that page, lines 16-31, are well done. I would say, however, if a sensitive taxon is found in a waterway in which it is unexpected (outside the distribution of that species), there is the possibility that perhaps the species range had not been accurately established originally. Or that it had expanded its range – either way, it would probably be best to check with local fish biologists before exclusion.

**14. Protection:** *Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.*

Comments: Yes, the analyses for the fish example criterion is protective of fish. The HC<sub>05</sub> of 392 uS/cm (95% CI 256-424 uS/cm) is appropriate. Good discussion in the Results, and as I observed in Question #13, many strong attributes accompany the fish analyses

### III. SPECIFIC OBSERVATIONS

Page	Line	Comment or Question
vii	5-4	Space needed between “of” and “survey”
xi	5-10	Delete the “and” before Kentucky
xvii	Continuing paragraph	Paired analyses is a strong statistical test. Exceptional number of field samples, sites and years of sampling!



Page	Line	Comment or Question
xvii and 2-3	10-12	Conductivity is described for eastern and western montane ecoregions but nothing is said about the Midwest – it's a major portion of the mid-section of the country and probably should be included.
xviii	Executive Summary	Very well done.
2-4	25	Figure 2-1 is located five pages away; could it be moved closer to pg.2-4?
2-3 to 2-7	Section 2.2.1	Thorough; excellent overview and foundational information
2-8	Table 2-1	Parentheses should encase 2012 in: Samarina (2007); Ruhl et al 2012.
2-11	Sections 2.2.3, 2.3, 2.5	Also excellent information; valuable for water quality staff to better understand the causes and mechanisms.
3-1	29-30	Could there be a bit more information with the “weighted CDF model”?
3-2	1-3	An accompanying short explanation of the statistical package R would be helpful.
3-10	Section 3.1.1.3.	Specific description of the sampling methods as well as assurances that adherence to standardized sampling techniques were observed, would be nice. Perhaps sampling details are in the Appendices.
3-16	8-15	Good recognition of the variance in sampling protocols among different agencies or monitoring groups. My concern is whether this variability can be “handled” by the process? The authors believe that it does.
3-23	Section 3.4	Well done.
3-27	10	Would not extirpation for the most sensitive crustacean occur before 960 uS/cm (of <960 uS/cm)? This is a high level of conductivity and mollusks are “canary” indicators of contaminants and water soluble stressors.
3-30	1-8	This was new information for me and find it very interesting. I did not know that Ephemeroptera can tolerate such low pH conditions if the conductivity is high. Good information; well done throughout all of Section 3.6.
3-32	Figure 3-8	An amazingly weak correlation! I might not have expected this, but it is clear.
3-34	7	“illiustrated” is misspelled.
4-1 to 4-27	Section 4	Figures and tables are very helpful.



Page	Line	Comment or Question
4-8 & 4-9	Figures 4-3, 4-4	These figures show large increases in conductivity in October. Seems that these higher values would affect the calculations for HC <sub>05</sub> and the HC <sub>05</sub> when simply looking at the figures. I understand the explanations given but am not 100% sure that they are complete enough for the less-trained in statistics.
4-27	4-11	Clarity of the use of the one day sampling/grab sample serving for CMEC and CCC.
5-4	Figure 5-1	Number of and distribution of sampling sites is exceptional.
5-6	12-15	Wouldn't the <200uS/cm Dec through June and >200 July through Oct provide support for the argument that there is a seasonal difference, thus calling for separation of data by seasons or seasonal weighting?
5-7	Figure 5-2	Seasonal variation is clearly shown for September. Difficult to understand how this wouldn't skew the results.
5-12	1 <sup>st</sup> paragraph	Personal Comment: Here in the Midwest we have distinct seasons, and water quality parameters often reflect this. Having unweighted, monthly/seasonal data is helpful to state agency staff who are trying to determine sources of pollutants and causal relationships. Determining sources of impairment are challenging and a clear understanding of what is happening each month (when there is monthly data available!), provides insight.
5-14	Figure 5-7	I concur with the acceptability of the hazardous concentration of 338 uS/cm but some interest groups may believe it is too stringent.
5-18	Line 7	Delete the second "for."
5-20	Figure 5-10	Delete "and" in the figure's title: "...southeastern Ohio into <u>and</u> Kentucky"
A-1	Figure A-1	A sentence or two describing LOWESS would be very helpful.
A-4	Figure A-4	Good to address other water quality parameters; informative.
A-5	1 <sup>st</sup> paragraph	My initial thoughts when reading this were that the confounders listed would have an effect on conductivity, and as I've stated in my response to one of the charge questions, I do believe that additional stressors can indirectly increase the damage done by high conductivity levels. Sorting it out, however, is an immensely difficult undertaking, requiring considerable data much uncertainty. However, in this document it was determined that confounders were not an issue.
A-7	7-8	"Removal of poor habitat samples from the data set had almost no effect on the SD model or HC <sub>05</sub> ." Based on the work of this study, this appears to be true. Unfortunately, if the removal of poor habitat doesn't affect conductivity, then those who oppose habitat restoration projects can use this as an argument in support of their position.

Page	Line	Comment or Question
A-11	1-7	Weighting, by its very purpose, brings a comparable 'status' to a data set with variable values and is a method by which calculations can be made. But care must be taken to correctly do the weighting to ensure correct representation of the data is seen in the results. I believe the authors have endeavored to do it well.
A-13	5	Not sure if this is recalculating to make data "fit" expectations. What was the RBP score used for the first calculations?
A-14	Section A.3.	Confidence intervals – there are some immensely large ranges.
B-9	1-5	Perhaps the RBP 130 score is not the correct level to use – could it have removed too many of poor and moderately poor sites?
B-9	17	Paired conductivity and biological data is a time-tested statistical test in environmental research; reliable and strong.
Appendix B	All figures	Well done; very helpful in conveying the relationships.
B-13	2-4	Of the 13 factors that were listed as being considered for having a causal relationship between conductivity and macroinvertebrates, some of them have had only minimal or no discussion in this document. Those are: nutrients, deposited sediments, selenium, settling ponds and dissolved oxygen. Selenium and metals were addressed in Appendix G.4.5.
Appendix C		Excellent data design, rationales, descriptions; Table C-1 very helpful.
C-4	Table "C-2"	The numbering of the tables is incorrect. It should be Table C-1 because it is a continuing of the table on the previous page.
C-5	Table "C-2"	Same as above
C-5	1-6	Helpful for understanding the information in the table.
C-7	4	"(see Table C-3)" should be: (see Table C-4). The numbering of the tables for the rest of the section is now 'off'.
C-8	Table "C-3"	Should be "Table C-2"
C-14	Top of page	Figure "C-3" should be Figure C-2. Very interesting geological information.
C-14	Bottom of page	Figure "C-4" should be Figure C-3.
C-15	Figure "C-5"	Should be Figure C-4. Strong relationship in the cumulative distribution between the Criterion data set and the Ohio data set; gives significant strength to the analyses.
C-16	Figure "C-6"	Should be Figure C-5. Very good illustration of distributions' overlap and the ranges overlap; Strong.
C-18	Figure "C-7"	Should be Figure C-6.

Page	Line	Comment or Question
C-19	4-6 and 22-23	Good points on looking at reasons for absence of sensitive species for evidence that the regions are different. However, in stating that “current conditions may not allow re-colonization,” means that habitat is poor, and this conflicts with the previous determination in the document that habitat quality doesn’t affect the analyses. Here it appears to factor-in.
C-21	6-7	The likelihood that “watersheds with >90% native vegetation are more likely to have low conductivity” are also likely to have better quality stream habitat. Indirectly this also supports the role of habitat.
C-22	Table C-10	Would have liked to see responses for 23-27, but this is an example of how there might be descriptions or verifications not clear or missing. Thanks to the authors for presenting it as it is.
C-24	Table “C-4”	Should be Table C-11. Very good table; informative and well presented.
C-26	2-3	“(see Figure C-7)” should be Figure C-6. “(see Figure C-8)” should be Figure C-7.
C-26	Figure “C-8”	Should be Figure C-7.
C-27	Figure “C-9”	Should be Figure C-8.
C-28	C.4.1	Appreciated the descriptions of the regions.
C-30	References	Brady, K: ... - overly bold underlining. Kahneman, D. - are there pages for the book?
F-3	Figure F-1	The tight correlations in the scatter plots are very good.
F-6	Table F-1	Excellent table; SO <sub>4</sub> + HCO <sub>3</sub> clearly significant.
F-21	Section F.5	Summary and Tables F-5 and F-6 are helpful. I wonder how the CCC of 160 mg/L and the CMEC at 300 mg/L compares with other regions around the country?
F-22	8	“(see Figure F-10)” should be Figure F-11.
F-23	Figure “F-10”	Should be Figure F-11.
F-24	Figure “F-11”	Should be Figure F-12. I would like to see more explanation of the bootstrapping method.
F-25	Figure “F-12”	Should be Figure F-13.
F-54	F.10	3 references should have underlining of the authors if the format is to be kept the same throughout: Barbour, MT..., Newman, MC..., and R Development Core Team
G-1	28-29	While I understand the need for minimum sample sizes of 500-800 macroinvertebrate sample and 800-1000 fish samples, I wonder if state agencies will be able to have that many in their databases for each ecoregion? Has there been any checking with other states to see if most can meet this?

Page	Line	Comment or Question
G-2; G-3	1-30; 1-13	Clear, accurate, and helpful discussion.
G-20	Figure G-8	Second to last line: “and 75-80 species evaluated.” The text on page G-18, line 29, said that it is 89, not 80. And in another, the number was 87.
G-20		Section 3.4 seems to be missing. Pages go from G.3.2. on pg. G-17 to G.4 on page G-20. If 3.4 is there, I didn’t see it.
G-25	Figure “G-7”	Should be Figure G-11.
G-27	G.4.6	Multivariant analysis for fish was interesting, especially the finding that catchment area and habitat significantly contributed to the model.
G-28	4 <sup>th</sup> line down	“Catchement” is misspelled: should be catchment.
G-34	G.6	The format of entries in the Appendix G’s references is not exactly the same as in the main reference section; To maintain the format here, Gerritsen, J.,... needs to have: a) semicolons b) initials follow the last name, c) uniformity in use of periods.
G-36	3, 7, 12	“Availble” is misspelled. Should be available.
G-42	Table G-7, title	“Ecoregions observed are the ecoregions where the species was collected in the combined data set” - needs to be bold
7-1	References	<p>Reference Section</p> <p>The entire section does not maintain one particular format. The following are problems:</p> <ol style="list-style-type: none"> <li>1) The initials on authors who are not the first and last author are not uniformly handled. In picking a uniform format, I suggest placing the initials in front of the surname. And periods following the initials.</li> <li>2) Parentheses around the year of publication or just a period? Some entries have parentheses, others do not. Some have periods, others not.</li> <li>3) A period after the journal name – or not?</li> <li>4) Titles in small or all capital letters?</li> <li>5) Listing of pages referenced in books – often missing.</li> <li>6) The agency’s name followed by its abbreviation in parentheses, or, the reverse?</li> </ol> <p>The following is the first author’s last name on every entry that I suggest be changed to meet a standard format and have one or more of the above problems. For me to re-write each faulty entry would be too time consuming.</p> <p>APHA Barbour Berra Bradley</p>

Page	Line	Comment or Question
		<p>Boelter  Brinck  Clark  Cormier (2010)  Dahm  Duncan  Dunlop (delete the second “Water Quality”)  Echols (2009 a)  Echols (2008b)  Efron  Entrekin  Evans (2008a) (2008b)  Evans (3001)  Farag  Fox  Godwin  Gregory  Griffith  Haluszczak  Harper  Hem  Higgins  Hill  Hille  Hitt  Hopkins  Hynes  Jackson (2007)  Jackson (2005)  Kaushal  (2005)  Kaushal (2013) – check to see if it is now published.  Kelly  Kennedy (2003), (2004), (2005)  Kimmel  Komnick  Lasier  Lefebvre, O. <u>and</u> R. Moletta  Likens (1970)  Merricks  Meyer  Mount  Mullins  Newman (2000), (2001)  Nelson</p>

Page	Line	Comment or Question
		<p>NYSDEC  Omernik (1987), (1995)  Paul, M.J. <u>and</u> J.L. Meyer  Pond (2004), (2010), (2008), (2014)  Posthuma  Remane  Sams – spell out USGS: U.S. Geological Survey  Scanlon – add “and” just prior to the last author.  Smithson  Soucek  Stauffer  Stubblefield  Suter (2007), (2001)  U.S. EPA (1985), (1987), (2000a, 2000b), (2003), (2006),(2009),  (2010) (2011a, 2011b, 2011c)  Van Dam – add “and” just prior to last author  Veil  Wallace – entomol. Needs capitalizing.  Werner – remove comma and add “and” between authors; Delete  “Wright et al. 1993”  Wood (2008)  Woods (2002), (1996)  Ziegler (2007), (2010)  Zielinski</p>



### Comments from Reviewer 3

## **Peer Review Comments on EPA's *Draft Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity***

**Reviewer 3**

### **I. GENERAL IMPRESSIONS**

I was impressed by the overall depth and breadth of this very well-prepared report by EPA (in my opinion it is one of the best technical reports, within my research areas, ever prepared by EPA). Obviously, this report has already gone through a rigorous internal EPA review (by many people who I know professionally) and by EPA contract support, as well as review by several other people who I also know and respect for their work. There is no question that the internal EPA technical workgroup contributed strongly to the report quality, again many people that I know professionally and respect.

Having worked with acid mine drainage, acid rain and numerous stream chemistry studies (as well as a few other lotic and lentic projects where conductivity was measured, besides a past life in estuarine and marine ecosystems measuring salinity), I am quite familiar with the strengths and vagaries of this very important measurement in both the field and laboratory. Also, I was pleased to see most of the key, but rather ancient, papers cited (e.g., the 1985 Hem paper), indicating that the literature review was excellent (although the key Hem 1982 paper was missing). However, there were also some recent, rather significant papers missing, but that may be due to the timing of the report preparation. No matter how hard you try, there are always supportive papers that may be missed in any literature review.

One concern was the redundancy of writing throughout the report (clarity was excellent for most sections), where general concepts appear to be often restated within some sections of the main report. I don't think that you always need to restate the obvious throughout every section of the main body of the report. However, with the potential wide array of future readers, some writing redundancy may be helpful. Here, my advice would be to have an outside professional editor (e.g., someone from Academic Press, Science, etc.) review the report structure and make recommendations to streamline this effort.

I like the Level III ecoregion approach, a method that I am using in some of my own work. My only concern here would be if there are applicable and robust data sets for each of the 85 Level III ecoregions. The report certainly uses a very robust, regional data base to develop criteria and to examine the statistical techniques needed to develop conductivity criteria within an ecoregion, and adjoining ecoregions. It will be really interesting to see how the States and Tribes respond to this report, as well as Congress

### **II. RESPONSE TO CHARGE QUESTIONS**

#### **Questions 1-3: Data Set Considerations**

**1. Ion Matrix Characterization:** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).*

Comments: I like the approach for using the ionic basis of:  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  to develop the initial conductivity criteria. This step alone eliminates any problems due to the potential effect of road salts, especially throughout the Appalachians and the Eastern Seaboard in general. It would be an interesting exercise to run the same analyses with no sample exclusion, and then do a comparison of XC<sub>95</sub> and HC<sub>05</sub> for only a few selected sensitive genera and some important benthic assemblages (e.g., EPT). I would assume that these would not be too time consuming, but may be worthwhile if there are ecoregions with lower sample sizes than the very rich data set employed in this report.

I would be a little concerned with any fall samples collected during an extreme drought period. If one assumes the normal two-component groundwater mixing model for eastern ecoregions, there is the possibility that a severe drought could result in over 95% of the stream flow coming from deep groundwater, and would represent an anomalous case for stream chemistry (successive years of drought may also be a very strong stressor on aquatic biota). It may be best to exclude any sample pairs (biota X conductivity) collected where gaged stream flows in a watershed, or a series of watersheds, dropped to below the 5<sup>th</sup> percentile of long-term flow records. Also, any exceptionally high-water events (greater than 99<sup>th</sup> percentile or perhaps 100-500 year storm events) may need to be considered if they occurred in the year before sampling. Benthic assemblage recovery (as cited in the report using the classic paper by Wallace, 1990) may take more than one year, depending on the species complex present in the stream and nearby refugia. Over my career, I learned quickly that there is no such thing as a normal year, and benthic and fish field collections need to be correlated with antecedent climatic conditions (e.g., temperature, flow, etc.).

Not being very familiar with the water chemistry of western streams, I believe it may be important to think seriously about any exclusionary criteria for these lotic systems. However, I know that some of the mid-western and western states have good data bases with which to run the same analyses as done for ecoregion 69.

**2. Catchment Size:** *All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate*

*anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: My personal opinion, and scientific bias, is to use only data from wadeable streams – this is the critical field design driver for stream assessment with EPA, and many of the eastern States and NGOs. EPA, along with many States, did a lot of work on developing such protocols to assure that there was robust physical, chemical and biological data collected in order to make non-biased estimates of many important parameters. Indeed, many key biotic and habitat metrics were developed based solely on wadeable streams. Also, 1<sup>st</sup> through 3<sup>rd</sup> order streams may constitute 70-90% of stream km in an ecoregion, with larger streams (4<sup>th</sup> to 12<sup>th</sup> order) representing less than 10-30% of stream km.

If one follows the River Continuum Theory, the 1<sup>st</sup> through 3<sup>rd</sup> (and perhaps some small 4<sup>th</sup>) order streams are where the real action is, and that the larger streams and rivers (large 4<sup>th</sup> to 5<sup>th</sup> and higher) start to reflect a major change in both ecological structure and function. OK, so one may collect some benthic organisms (genus may be the same, but probably different species) in the larger order streams that would also be found in lower order streams. However, stream processes in the larger order streams are so different I feel it would be difficult, and unjustifiable, to use this approach. Obviously, EPA would welcome this opportunity to be able to set conductivity criteria for large aquatic ecosystems (large stream and rivers), especially in light of the NPDES permits, etc.

*3. Seasonality: The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: First, see part of the response to question 1. I would be careful including data for extreme stream flow conditions, which may occur in spring (high flows), summer (possible hurricanes), and fall (drought). Care should also be taken to examine any unusual antecedent conditions within watersheds to be studied. In our regional work, we needed to delete a few 1<sup>st</sup> and 2<sup>nd</sup> order sites due to extreme high flow conditions in the previous year that affected two subwatersheds in our study area

#### **Questions 4-8: Case Studies: Example Criteria Calculations**

*4. Criterion Continuous Concentration (CCC): Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).*

Comments: I really like this approach, since these are well developed exposure-response relationships at the genus level, assuming that any species within the genus would share a similar response (well-known for many fish genera exposures to numerous stressors). The entire sequence

of CCC analysis and the derivation of the CCC for conductivity are very well-presented in Figure 3-2, as well as in the text. Also, the example in Figure 3-1 is good, giving the reader an example of how to derive the HC<sub>05</sub> of a genus sensitivity index – not a particularly easy concept to grasp unless one has some background in bioassay statistical techniques.

One analytical - statistical comment: There have been a series of papers in recent years by King and Baker who use the TITAN model to examine stressor relationships with biota. It may be beneficial to explore this model to estimate conductivity-response as a check on the CCC.

**5. Criteria Maximum Exposure Concentration (CMEC):** *The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?*

Comments: Similar to my comments for Question 4, the sequence for determining the CMEC is very well described in Figure 3-6. I feel that the derivation of both the CEC and the CMEC are very robust, in part because of the availability of rich ecoregion data sets. I like the fact that there is careful trimming of the data set, followed by examining for unequal variances and for estimating Type I errors (there are often models published that do not perform these simple tests).

**6. Duration:** *Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Comments: Description and justification are more than adequate to support both CCC and CMEC. I am always a little leery about a CCC (or any water quality criteria that is based on a yearly value), since Figure 3-7 does illustrate very well the potential for large yearly variations in stream conductivity. In one of my forested study sites, conductivity may range from 75-100  $\mu\text{S}/\text{cm}$  in the spring to over 600 – 700  $\mu\text{S}/\text{cm}$  in late summer - early fall due to the dynamics of stream flow and forest transpiration.

**7. Frequency:** *Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Comments: This section is well written and uses two classic papers (Niemi et al. and Wallace) to illustrate recovery rates in benthic organisms (insects primarily) from stressors. Recovery in stream fishes is not as clear since there may be multiple physical stressors that create long-term problems after water quality remediation (e.g., AMD), especially for lithophilic spawners. Generally, this section is highly supportive of CCC and CMEC.



**8. Alternate measurement endpoint: Is the example alternate measurement endpoint ( $\text{HCO}_3^- + \text{SO}_4^{2-}$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?**

Comments: Just a very general comment to start with in this response to question 8. In our water quality laboratory, we generally do a complete cation and anion scan since these are easy on an ion chromatograph. Although primarily interested in Ca and Mg, we also analyze for K and Na, and have found these to be important cations in some streams. The anion scan is important in that it also gives a few other ions that appear sometimes in our study streams, although we do nutrient scans on other instruments due to sensitivity and detection limits.

So, based on the discussion in Appendix F, I would be very comfortable with using the alternative measurement endpoint, but only as a last resort if the water quality data is not adequate for a data set (meaning no measurement of Ca, Mg or Cl). There is not much difference in the slopes of Figure F-1 (c) and F-1 (d), although there is less scatter in F-1 (d) with the addition of Cl. (Note: why wasn't a test for equality of slopes performed or did I miss it somewhere in this section?).

#### **Questions 9-12: Geographic Applicability**

**9. General: Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?**

#### **Comments:**

First Comment: I wonder if the general statement could be made that the process would be applicable for any Ecoregion III level imbedded within any Ecoregion II level. This seems logical to me, since the original development of ecoregions was designed to address similarities in geology (and other parameters) that would translate to similar, but not identical, stream ionic concentrations. In the examples in Section 3.6, the ecoregions are adjoining so there may be a very high probability that stream chemistries may be similar.

Second Comment: A good test may be to examine two Ecoregion III level watersheds that are not contiguous, or at a large distance from each other, or perhaps two watersheds close to each other and two distant from each other.

Sidebar: In regard to section 4.1.1, I calculated background conductivity using the Y-intercept method developed by Dodds (for estimating background nutrients in mid-western streams) for some 152 probability-based stream sites that we sampled in Ecoregion 69 over the years. My estimate of background conductivity was 82  $\mu\text{S}/\text{cm}$  and the estimate in the report was 80  $\mu\text{S}/\text{cm}$ . I was pleased that these two estimates were in close agreement, especially since the analytical



approaches were quite different. It is not that the Dodds' technique is so great (unless there is an adequate sample size), but similar background estimates indicate that the EPA approach for estimating background conductivity is consistent with other potentially useful statistical techniques.

***10. Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

Comments: Based on the analyses presented, I do not have any problem with the process. Indeed, the case study analyses in Sections 4.3 and 5.3 do support the criteria application. Basically, I feel that the parent data set is a training set, and the conductivity estimates outside of this set should be well within statistical bounds.

***11. Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

Comments: I think this section was well written, and that the data and analyses do support the application of conductivity criteria to other areas.

***12. Applicability to ephemeral streams: In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?***

Comments: My basic comment is that we are not doing enough to protect either intermittent or ephemeral streams, and I would recommend that EPA take a stronger stance on these important characteristics of the watershed. It is hard enough to protect 1<sup>st</sup> order streams in the United States, but trying to gain protection for zero order streams is almost impossible. Consequently, and where applicable, any paired analyses with conductivity and benthic organisms would be beneficial to support the importance of ephemeral streams.

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

**13. General:** *The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.*

**14. Protection:** *Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.*

Comments: Unless I am missing something, there is no Section 3.4 in Appendix G. So, there are many obvious differences between benthic and fish data in assessing conductivity effects. The benthic data is at the genus level – good enough, but fishes are easily identifiable (with a few exceptions) to the species level. However, fish sampling is more time consuming so not as many samples are collected in comparison to the benthic collections. There are a number of other considerations as well discussed on pages G-1 and G-2.

I liked the approach where fish data was lumped from four ecoregions, which resulted in a data base of over 1,437 observations. The clarity and validity of the modified method was adequate to derive the XC<sub>95</sub> and HC<sub>05</sub> values. I also liked the data filters that were employed for the analysis – these eliminated a lot of potential problems with the data analysis.

OK – here is where I am unhappy with the fish data. First, both rainbow trout and brown trout must be excluded from the data set. These are exotic, introduced species and even through there are established populations of these two species, that is not a good reason to include them. Many folks are trying hard to protect native species throughout the Appalachians, and including them as well as carp just does not make sense. I would follow the listings for introduced and exotic species, as found in Wiley and Hocutt, as the cut for potentially introduced species into an ecoregion. Also, just because the two trout species are recreationally important, that is not a good enough reason to include them in the analysis. If the work done by Tim King is valid, then there are only a few places where one needs to worry about brook trout introductions. After all, this is the native salmonid of the Appalachians.

I like Figure G-9 very much. First, it showed the species sensitivity distributions for many fish species. More importantly, it illustrated that even within a genus, there was wide variation in the response to conductivity, e.g., *Etheostoma* and *Cottus* sp. The derivation of the fish HC<sub>05</sub> is excellent, but the hazardous concentration of 392  $\mu\text{S}/\text{cm}$  seems a little high to me, but that may be a reflection of the species that are common in my research sites.

**14. Protection:** *Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.*

**Comments:** It appears the case study criteria for the benthic organisms would also be protective of fish populations. I don't think any additional data and analyses are needed. I look at this approach as a two-factor method where the benthic conductivity criteria would drive the fish protection, and perhaps *vice versa* in special situations.

**FINAL COMMENT:** I would like to have seen a little more done with assemblages, e.g., EPT, intolerants, tolerants, etc. However, the genus level for bugs and the species level for fish approaches are great, especially with the highly robust data sets

### III. SPECIFIC OBSERVATIONS

Page	Line	Comment or Question
Entire report	Entire report	Capitalize States where appropriate – eliminates the confusion between noun (States) and verb (state or states) forms. Perhaps also capitalize Tribe. Check foreword.
xiii		FORWORD should be FOREWORD
xiii	8, 14, 17	Capitalize States
2-4	15	Split ¶ after effluents (before Ionic)
2-10	Figure 2-1	Change black font to white on right side of all three blue blocks. One never uses black on blue.
2-17	18	Period after al (in et al. check entire document)
4-6	5	Figure 4.1 rather than 4.2
4-10	Figure 4-5	Cannot see data points!!!! Faint!!!!
5-11	Figure 5-5	Cannot see data points!!!! Faint!!!!

#### Comments from Reviewer 4

## Peer Review Comments on EPA's Draft Document:

### "Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity"

#### Reviewer 4

#### I. GENERAL IMPRESSIONS

Overall, this is a well written, scientifically sound paper which lays out the technical approach and underpinnings for deriving conductivity benchmarks for aquatic life for streams using field derived measures of water chemistry and ionic strength measures with co-currently collected measure of aquatic macroinvertebrate response at the genus level of taxonomy.

My major issues are related to the actual application of these results to protect aquatic life uses in State water quality standards and, particularly, under tiered aquatic life use frameworks. Application of this method in states, like Ohio, that have tiered aquatic life uses, could result in benchmarks that are overprotective of the baseline warmwater aquatic life use, but could also be under-protective of exceptional (EWH) uses. Fortunately, I think the methodology presented here can take these factors into account. For example, for "EWH" streams there may be a more restrictive suite of species that occur in these waters and exclusion of more tolerant forms could drive the XC95 a bit lower. In contrast for "WWH" streams, the most sensitive species may not occur frequently enough in those streams to "drive" the XC95 benchmarks and the more common sensitive species might result in a less stringent, but perhaps more attainable benchmark.

Another key issue that may influence the derivation of criteria with this method is the definition of reference conditions. I think some more discussion of reference conditions as in Stoddard et al. (2006) would be helpful, I will delve into these comments in more detail in the charge questions below.

I do think this paper provides a solid technical basis for deriving benchmarks using field derived SSDs and the derivation of HC05 and XC95 values. My comments focus on the need to deal with some of the application issues surrounding these benchmarks. The ability of a State to use this methodology will be related to the quality of their monitoring, assessment, and water quality standards programs (Yoder and Barbour 2009). Reference to the critical elements that monitoring programs should have to accomplish this methodology should be a part of this document. Again more detail will be provided below.

#### II. RESPONSE TO CHARGE QUESTIONS

##### Questions 1-3: Data Set Considerations

***1. Ion Matrix Characterization:*** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set*

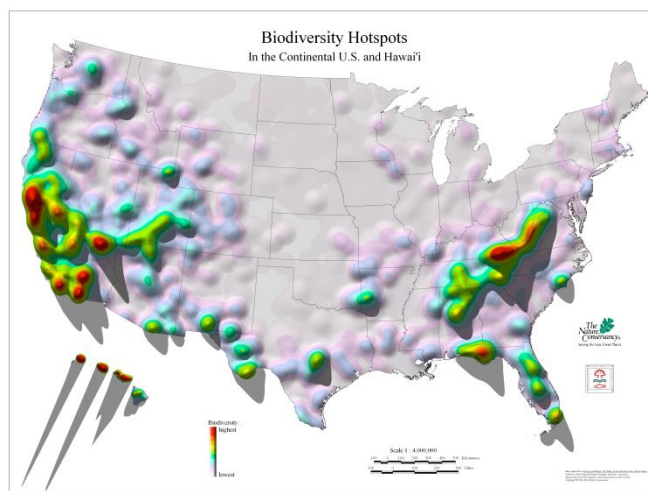


(e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[SO_4^{2-}] + [HCO_3^-] > [Cl^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).

Comments: I have no problem with removing sites from the analyses where different ionic mixtures are likely to confound results (e.g.,  $[Cl^-] > [SO_4^{2-}] + [HCO_3^-]$ ) or where other stressors (e.g.,  $pH < 6$ ) may also contribute to confounded results. I do have some question on whether there could some other confounding caused by: 1) natural variation in conductivity at “reference” sites, and 2) variation in conductivity along a gradient of sites that may be considered “reference” in the sense of “least impacted” conditions vs. “minimally disturbed” in the sense described in the paper by Stoddard et al. (2006).

It is clear that there are clearly natural biodiversity “hotspots” in the ecoregions examined here (see example Nature Conservancy map). Some of these hotspots may partly be remnants of where biodiversity has been minimally disturbed by human activities, but, some are where there is a combination of natural features (e.g., habitat, gradient, elevation, water chemistry) that combine to maximize biodiversity. My concern is that these natural “hotspots” may well be driving the XC<sub>95</sub>/HC<sub>05</sub> value particularly when aquatic life use potential is defined by a single aquatic life use, and therefore a single benchmark is derived. The effect of a single benchmark is that it may be under-protective of the most unique “hotspots” but overprotective of more typical habitats. I will address this comment more specifically below.

2. Catchment Size: All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas ( $>155 \text{ km}^2$ ) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.



My concern with this discussion is not so much with stream size as important variable, but other natural classification issues and some anthropogenic changes that might have occurred from human habitation and land disturbance that are not acute or readily controllable and are within a



definition of “least impacted” streams. For example, in the mountainous regions of the WAP ecoregion in West Virginia for example, the relief has led to land uses (e.g., forestry, park, light agricultural, low density residential) that result in more highly forested (>90%) reference conditions. Along the edge of the WAP ecoregion in Ohio for example, the relief is more variable and farming and some other land use changes are somewhat more intense. “Least impacted” reference sites are much less likely to be “>90% forested.” This broaches the important question of whether a single benchmark or multiple benchmarks to match tiered uses are more appropriate.

**3. Seasonality:** *The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: Since the analyses indicated the effects of seasonality are minor, I have no problem with how the paper dealt with this issue. Given the pattern in conductivity in some of the datasets where there are higher values in the late summer (e.g., August-September), a period that corresponds with typical lowest monthly flow periods, it might be of use to discuss how this might influence monitoring for compliance with any derived criteria. For example, the paper talked about a monthly weighting of conductivity values to determine the effect of seasonality on the criteria. If a State only collects data during a summer period (e.g., Aug/Sep) should the values be adjusted to the annual geometric mean to determine whether benchmarks are exceeded?

#### **Questions 4-8: Case Studies: Example Criteria Calculations**

**4. Criterion Continuous Concentration (CCC):** *Please comment on the clarity of the method to derive the  $XC_{95}$  and  $HC_{05}$  (Section 3.1, Deriving a CCC).*

Comments: I generally found the approach to derive the  $XC_{95}$  and  $HXC_{05}$  relatively easy to understand, with perhaps a more step-by-step on how to calculate the weighted CFD values. Was this done using Excel, R, or some other application?

**5. Criteria Maximum Exposure Concentration (CMEC):** *The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the  $HC_{05}$  values?*

Comments: This approach seems reasonable, however it seems that further empirical analyses of the consequences of this approach would be useful. For example, for sites that are achieving some biological benchmarks (e.g., IBI, ICI) what is the frequency that these are considered impaired based on the CCC and/or CMEC? Again my concern with single criteria rather than tiered criteria has some consequences with use of both of these benchmarks. An example of using tiered criteria and calculating CCCs and CMECs for both would be useful.

**6. Duration:** *Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Comments: States like Ohio and localities such as the MSDGC (collected by MBI) commonly collect biological data paired with one or more weekly continuous regimes of conductivity data (e.g., Datasonde collectors). It seems that some of these sites can be used to examine duration questions in more detail. Such datasets have hourly values of conductivity collected over 7-10 days, once or twice a summer.

**7. Frequency:** *Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?*

Comments: I think the discussion of the estimation of frequency is reasonable (not to be exceeded more than once in three years on average), but perhaps can be supplemented by some ambient analyses as another form of evidence. One suggestion might be to derive “biological stressor metrics” using the  $XC_{95}$  values. For example, I have used the most sensitive 15<sup>th</sup> percent of conductivity weighted mean values by taxa to determine taxa “sensitive to conductivity.” For each site one can then generate the number of conductivity sensitive taxa present which can then be used to provide evidence that the count of sensitive taxa varies with conductivity as predicted under various duration, frequency and magnitude scenarios. This can also be used to compare potential tiered use responses to conductivity.

**8. Alternate measurement endpoint:** *Is the example alternate measurement endpoint ( $[HCO_3^- + SO_4^{2-}]$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?*

Comments: The correlation between conductivity and the alternate measurement endpoint ( $[HCO_3^- + SO_4^{2-}]$ ) is so strong that I think most users (e.g., States) will focus on conductivity given its cost and ability to cheaply monitor it continuously. Because of this I did not analyze this as closely as some other parts of the report, but it seems to result in a similar type of benchmark.

## **Questions 9-12: Geographic Applicability**

**9. General:** *Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?*

**Comments:** To derive criteria as proposed, the process for assessing general geographic applicability of conductivity is fine. As I discussed above, derivation of benchmarks under a tiered series of aquatic life use may need to accommodate modifications to the derivation approach. The geographic applicability approach generally compares whether the range/variability in conductivity in background conductivity is similar between regions. I am not sure it addresses conditions where a subset of streams may have uniquely (and predictably) lower conductivity that need to be considered separately. Unfortunately, this is somewhat confounded with accurately identifying “background” conditions, particularly in the Ohio region of ecoregion 70. I think this paper would be well served to be placed within the conceptual framework of the Biological Condition Gradient framework (Davies and Jackson 2006) and the reference site framework of Stoddard et al. (2006). For example, Appendix D defines “Background conductivity as the range of ionic concentrations naturally occurring in the environment that has not been influenced by human activity.” Several paragraphs later, in weighing lines of evidence, it asks: “Are conductivity values at natural background (least-disturbed) sites similar in the new area compared to the original area?” As defined by Stoddard et al. (2006), least impacted is the best available physical, chemical and biological habitat conditions *given today’s state of the landscape*. With a naturally occurring measure such as conductivity, this definition can be important. Is the single benchmark or recommended criteria a reflection of “minimally disturbed” (site condition in the absence of significant human disturbance) or of least impacted conditions? Tiered uses allow a State to recognize that a subset of sites may approach minimally impacted and the associated criteria can form a baseline to protect that level of condition. Conversely, if a State has another class of sites with an appreciably higher level of *acceptable* development across the landscape and these sites are still considered least impacted (and these cannot be managed in a way to reduce the conductivity footprint), then different criteria for certain stressors may be applicable. For nonpoint sources of pollutant the CWA talks about controlling stressors that can be feasibly addressed with best management practices.

One way to more closely examine the influence of tiered aquatic life uses and tiered water quality criteria would be to construct a number of human disturbance indices or gradients (e.g., Bryce et al. 1999; Wang et al. 2008) and then relate them back to well-founded biological condition gradient exercises that classify sites into six ranges of biological condition based on definitions for ten components of biological condition (Davies and Jackson 2006). This has been done for many States. It may be that the species that comprise the upper tiers of the BCG (e.g., 1-2) could well be the ones that drive the selection of the XC<sub>95</sub> value, and absence of more tolerant taxa from reference sites would drive the conductivity benchmark lower. Conversely, these species may occur in too few reference sites at lower tiers that represent “least impacted” conditions (e.g., BCG tier 3-4).

Appendix C, in describing how to use weight-of-evidence to examine the geographic applicability, did not examine the different aquatic life use streams across this ecoregion (particularly EWH vs WWH). Table C-7 indicates that there were no reference sites; however, biocriteria for WWH streams in this ecoregion is based on the 25<sup>th</sup> percentile of “least impacted” reference sites (EWH based on 75<sup>th</sup> percentile of sites statewide). Analysis of conductivity values at reference sites and by aquatic life use would be important evidence for this analyses.

Base flow seemingly is an important variable not considered other than in a general way in this appendix. Elevated conductivity values at sites in Appendix C in August and September correspond with the lowest estimate monthly average flows by month in Ohio (USGS ungagged model output). Some more explicit consideration of local flow influences on ionic strength would be useful. It is my experience that local base flows in headwater streams can vary considerably within across this region.

***10. Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

Comments: The same comments I provided above apply to this charge question.

***11. Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

Comments: I think the background matching analysis is a sound approach for comparing applicability to another ecoregion. Again, I have the same caveat about potentially doing this in a tiered use framework.

***12. Applicability to ephemeral streams: In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?***



**Comments:** I would examine Ohio EPA's Primary Headwater Assessment data. It focused on streams generally less than about a square mile and includes ephemeral as well as perennial and interstitial streams. They have collected conductivity data and have sites in the WAP ecoregion as part of their studies. Ohio University also collected primary headwater stream data as part of a study in the vicinity of the Portsmouth nuclear facility that might be of use. In a neighboring ecoregion (Interior Plateau), MBI has data from around 100 or so streams around Hamilton Co., although many have urban impacts.

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

**13. General:** *The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.*

**Comments:** Again, I generally found no real problem with the statistical approach and found the modified bootstrap methodology reasonable. My main concerns are related to how the regions are combined, given biogeographical differences in fish distributions across ecoregions. The argument is made that, in a manner similar to SSDs generated for toxicity testing, it is not important that the species that make up those below the HC<sub>05</sub> do not occur in Ohio. I am not sure this is reasonable for a natural "stressor" such as conductivity. The benchmark is driven by coldwater and rare species that do not occur in Ohio. If this analysis was conducted with Ohio ecoregion 70 data alone, perhaps with a lower threshold of sample size, perhaps it is possible that other sensitive fish species would replace the most sensitive taxa in Appendix G. It is likely, however, that those sensitive taxa are inhabitants of the EWH tiered use in Ohio rather than the WWH use. Again, I think that some discussion of tiered uses is essential to how a State might apply this approach.

**14. Protection:** *Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.*

**Comments:** My experience with conductivity and other ionic strength parameters is that macroinvertebrates are generally more sensitive as a group than fish, but I think because of sample size considerations the most sensitive fish are often being excluded from the analyses here. The limited distribution of certain fish species (and macroinvertebrate genera) that are often excluded may in themselves be evidence that the aquatic life potential may vary with some natural as well as anthropogenic impacts. As monitoring programs mature, samples sizes in these areas are continually growing, so I think that State water quality standards programs need to be continually exploring their databases to refine aquatic life uses and the criteria designed to protect these uses. Modification to the approach for deriving a single ecoregion criterion for all streams requires an adequate monitoring program with robust critical program elements (Yoder and Barbour 2009) that provide the data and the capability to conduct the analyses described in this document. For example, a program needs to be capable of accurately classifying and controlling for natural

features that influence biological assemblages (e.g., stream gradient, size, elevation, base flow, etc.). Tiered uses require robust data with the ability to recognize the influence of anthropogenic influences on the landscape and the ability to address what stressors may or may not be feasibly controllable.

### III. SPECIFIC OBSERVATIONS

Page	Line	Comment or Question
Entire Report	--	As I mentioned in my general comments I think conductivity criteria should be considered in a tiered aquatic life use framework. There is natural variation in “background” conductivity due to variation in precipitation, base flow, etc., and within the range of “least impacted” to “minimally impacted” reference sites (as defined by Stoddard et al. 2006) there are variations due to human occupation of the landscape (e.g., agriculture, residential). My fear is that the criteria may not be stringent enough for the minimally impacted regions, and too stringent for land uses that State’s would not considered to be impaired, but rather to be consistent with the swimmable/fishable goals of the act. This is not a suggestion that least impaired would encompass mine-related acute impacts or other impacts that are feasibly controllable.
Glossary	--	“Background”– The definition of background I think is a bit “murky” given that later in text it includes both minimally impacted and least impacted reference sites. I would also add definitions of “least impacted” and “minimally impacted” reference sites and “tiered aquatic life uses.”
1-3	--	The document talks about how the protection of 95% of genera with this method is comparable with the protection of 95% of “species” in the lab toxicity approach. Some genera are more speciose than others. Is it truly similar? Not a major comment.
2-3	25-28	This supports the contention that “background” levels can vary depending on how reference sites are defined. Within an ecoregion there could be subwatersheds with lower or higher conductivity than neighboring watersheds. Because of this, the natural distribution and abundance of those taxa that are most sensitive to conductivity can vary. These differences can be due to natural or some level of anthropogenic impacts (not acute or controllable sources). A discussion of tiered uses, I think is warranted for a stressor such as conductivity or perhaps other “natural” stressors such as habitat.



Page	Line	Comment or Question
2-3; 2-12	29-30; 3-6	Because precipitation can influence conductivity (the variation in seasonal concentrations, that is higher in late summer, tends to be months when precipitation is lowest) it seems that measures of base flow may be important in resolving within regional background variation in conductivity. My experience from Ohio is that the Exceptional Warmwater Habitat streams often have high base flows. In any case, I think discussion of potentially tiering conductivity benchmarks should be discussed.
2-16	22-29	Along a gradient of stress, such as conductivity, the probability of capturing an individual of a genus decreases with increasing stress. How does sampling methodology potentially influence derivation of benchmarks. If a taxa is not collected when the sample size is small, there is some likelihood it is present when not represented in the catch. Is there a way to use methods that count many more organisms (e.g., Ohio EPA method can have abundance estimates greater than a 1000-2000) to determine the bias when using methods that only count 200 or 300 individuals. Thus what is considered “mortality” may be partly under sampling. Although the trend in taxa response with conductivity may be similar between methods, the actual benchmarks could change if “sensitive taxa” show up at somewhat higher conductivities.
3-5	20	What is considered “similar background conditions?” If there are two groups of sites, one centered on a conductivity of 150 and the other on 250 and both are considered background, is that similar enough? When do you decide that you have two groups of sites that might be within the same ecoregion but that differ enough in conductivity and taxa that two tiers are needed?
3-8	22-29	This is where the concept of tiered uses could be tested. If minimally impacted reference sites can be distinguished from “least impacted” then the general not observed at reference sites could differ. The BCG process that has been used in a number of states results in output that classifies sites into different BCG tiers with reference sites usually varying between tiers 1-4 depending on definition. Sites are rarely classified as BCG1, but data sets where sites identified as BCG tiers 1-2 could be compared to those defined as BCG tiers 3-4 to see how this might affect the derivation of criteria. It would certainly influence which sites occur or do not occur at reference sites.
3-17	18	Because of the assumption of using samples from both seasons to generate the criteria, this implies that looking for exceedences of conductivity should be based on a geometric mean value from monthly samples. For determining water quality violations of the criteria are we expected to take monthly samples including both spring and summer periods or is there a methodology to adjust the “expected” criteria if only summer samples are taken as is common

Page	Line	Comment or Question
		for many monitoring programs. Some more specific guidance on sampling for what would be considered violations or exceedences of criteria would be useful.
3-35	4-14	This is where I think some explicit definition of reference site conditions (minimally impacted distinguished from least impacted) is advisable (sensu Stoddard et al. 2006). In addition, an independent human disturbance gradient could be used to estimate the anthropogenic footprint.
3-35	23	Would an independent human disturbance gradient measure be useful here?
3-39	--	Given the “fuzziness” of deriving background conditions I think this can be difficult. Again the concept of tiered uses is important in this context and I think there is a need for a more explicit approach to distinguish between minimally impacted, least impacted and best attainable. For example a 90% forest threshold for a watershed may not be feasible in many areas and it is an important question whether conductivity levels are actually feasibly controllable in all cases.
4-17	5-19	It is easier to read if the X-axis of the plot is in actual conductivity units rather than as the log of conductivity.
4-24	--	Is it possible that this graph argues for the existence of two tiers of expectations? For example could there be unique, less common areas where conductivity is naturally very low and the most sensitive taxa more narrowly distributed and thus rare and less likely to meet a threshold of 25 sites for a genus to occur, and then more typical sites where more sensitive taxa are not found as frequently?
A-13	--	General Question: The section analyzed differences of HC <sub>05</sub> values when low habitat scores were eliminated as potentially confounding variables. To explore the different benchmarks that might occur under tiered uses, perhaps the RPH habitat could be used to establish “reference” cutoffs under a crude tiered use scenario (reference sites with habitat scores > 160 vs. > 180). Ideally this could be done with a State like Ohio where tiered uses (EWH vs. WWH) are clearly defined.
C-1	19	Here is an example where I think the concept of background needs to be better quantified. At a minimum it should be related to Stoddard et al. (2006) definition of minimally vs. least impacted conditions. Ideally some form of a human disturbance score can be calculated. There is scatted mention of > 90% forested as a reference benchmark, but that may be hard to find even in the WAP ecoregion of Ohio. For States to apply these benchmarks I think it argues for detail discussion of tiered uses and reference or background conditions.

Page	Line	Comment or Question
C-2	15-16	Clarify whether background is minimally disturbed or least impacted.
C-2	17-18	The concept of subregions and local variations in base flow, which may not be easily predictable needs to be discussed more. Base flow seems to be a very important influence on conductivity given the variation observed in conductivity by month. Conductivity is usually higher overall in late summer early fall when flow is a minimum, but base flow makes up the greatest % of stream flow. For very small headwater streams, base flows can vary substantially within a region depending on the complexity of groundwater systems and points where streams become gaining flows. In larger streams I think these likely average out within a region, but in small streams may be important and may result in difference in rare and sensitive macro taxa.
C-3	Table C-1	Regional properties – need to add consideration of base flow to this table. The sandstone aquifers in SE Ohio tend to have conductivities of 450-600 which indicates as the percent of flow that consists of groundwater increases the more likely that higher conductivity will occur. May also want to see Ohio primary headwater assessment data from Ohio (streams < 1 sq mi) that has conductivity data ( <a href="http://www.epa.ohio.gov/portals/35/wqs/headwaters/PHWH_Compendium.pdf">http://www.epa.ohio.gov/portals/35/wqs/headwaters/PHWH_Compendium.pdf</a> )
C-17	Table C-7	This table indicates that reference sites do not occur in Ohio, but Ohio does have reference sites. Also, would it be difficult to get forest cover for the Ohio sites? My guess is that few approach the 90% forest cover mentioned (page C-21, line 8) for further south. This has implications for attainability and setting feasible and controllable benchmarks.
C-17	Table C-10	Again reference sites are available for Ohio.
G-2	23-24	The comparison of species vs. genus level XC <sub>95</sub> values identified that genus values represented the more tolerant species in the genus. Why wouldn't that apply to the macroinvertebrate analyses and does this suggest that the lower conductivity sites in a region are not adequately protected? Would this be resolved with tiered uses and perhaps a lower threshold that would let more rare and sensitive species into a higher tier use?
G-7	4-5	Ohio fish sites are not all in sites I would call "perennial." Although very few of the sites dry completely, many small headwater sites can occur in what I would term interstitial streams that have periods where flows in riffles are subsurface, although permanent pools remain.
G-7	9	Again we have the 90% forested "benchmark" for reference with little discussion of what this means.

Page	Line	Comment or Question
G-20	4-6	It would be useful to see data on sites that missed the N=25 cutoff in this table to see if they characterize tiered uses or local high quality sites that may be important or some unique restricted distribution that might be important.
G-22	9	In addition to tiered uses on some states, most states characterize coldwater uses as unique from warmwater uses. Ohio for example identifies species considered as characteristic of its coldwater uses. How would removal of coldwater taxa change the HC <sub>05</sub> values?
G-23	26-31	If coldwater benchmarks were delineated separately from warmwater benchmarks, how would this affect this conclusion?

### References cited

Bryce, S. A., D.P. Larsen, R.M. Hughes and P.R. Kaufmann, 1999. Assessing relative risks to aquatic ecosystems: A Mid-Appalachian case study. *Journal of the American Water Resources Association*. 35: 1752-1688.

Davies, S.P. and S.K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16(4):1251-66.

Stoddard, J., P. Larsen, C. P. Hawkins, R. Johnson, and R. Norris. 2006. Setting expectations for the ecological condition of running waters: the concept of reference conditions. *Ecological Applications* 16:1267-1276.

Wang, L., T. Brenden, P. Seelbach, A. Cooper, D. Allan, R. Clark Jr. and M. Wiley. 2008. Landscape Based Identification of Human Disturbance Gradients and Reference Conditions for Michigan Streams. *Environ Monit Assess* (2008) 141:1–17.

Yoder, C.O. and M.T. Barbour. 2009. Critical technical elements of state bioassessment programs: a process to evaluate program rigor and comparability. *Environmental Monitoring and Assessment* 150(1-4): pp 31-42.

### Comments from Reviewer 5

## **Peer Review Comments on EPA's *Draft Recommended Field-based Method for States to Develop Ambient Aquatic Life Water Quality Criteria for Conductivity***

**Reviewer 5**

### **I. GENERAL IMPRESSIONS**

For the most part, the document is well written – although with some exceptions that I have noted below. I was impressed with the thoroughness of the presentation. I found most of the substantive information in the document to be accurate. I did find some errors, and I differ with some of the interpretations.

I find language used to express essential concepts as problematic, especially the manner in which the term “conductivity” is used to express what is more widely described as “specific conductance.” If EPA persists with its current use of the term “conductivity” within the context of the program proposed, the term “conductivity” would then have two different meanings: electrical conductivity (the raw measure) and the 25°C-temperature-corrected value. This result can only breed confusion. I strongly encourage EPA to use the words “conductivity” and “specific conductance” (SC) in accord with well-recognized and widely used precedents and practice.

Given the current status of scientific knowledge concerning major-ion effects on aquatic biota in the Appalachian coalfield, I see the primary method presented by the document and illustrated by Case Studies 1 and 2 as generally adequate, as a temporary measure, for describing specific conductance (SC) levels that will be protective of 95% of benthic macroinvertebrate taxa in Appalachian coalfield streams. I say “temporary” given the lack of scientific certainty concerning the precise nature of the stressor that is causing the effects that are being observed so widely. I expect that the stressor will be identified with greater certainty, eventually, and that will allow more precise targeting by regulatory and other management actions. I do have some technical concerns which are described in the responses. Many of my technical responses are focused on what I see as inadequate biological confirmation of results which are obtained from analysis of large datasets. While the conclusions derived from such analysis would likely be considered as adequate if expressed with appropriate caveats in the context of academic studies, these results are proposed for application to individual situations of widely varying circumstances as a regulatory program.

Clearly, aquatic communities are depressed in waters influenced by mining throughout the Appalachian coalfield. Given that numerous studies have found close associations between elevated SC/major ions from mining and alterations of benthic macroinvertebrate community metrics, it is reasonable to expect that some effect on water by surface coal mining is playing a major role. Given the number of studies that have found negative associations between elevated SC and benthic macroinvertebrate conditions in the Appalachian coalfield, and the lack of relevant studies that have failed to find such effect, release of SC/major ions has to be considered as prime suspect. However, the direct causal agent – i.e., the precise combination of ions and/or SC-associated stressor such as, perhaps, specific ion combinations or ratios, mining-induced



hydrologic alterations, or other unstudied factors – is not known. Hence, it is my view that any public policy actions taken should recognize that the science defining causation is not yet settled. Hence, I see the following statement from the document's forward as appropriate: "State and tribal decision makers would retain the discretion to adopt approaches on a case-by-case basis that differ from those described in this draft document, even if the method in this document is issued under CWA section 304(a)."

I have been conservative in my interpretations due to recognition that the document's content has the potential to become a regulatory program. Given the consequences of the potential regulatory actions that may be based on HC<sub>05</sub> values derived as described here, ensuring those values' validity across the full range of resources targeted for application requires additional biological confirmation. The HC<sub>05</sub> values are being defined by SC values associated with small numbers of taxa ("limit-defining taxa"). Those values should be checked by conducting additional analysis to determine if the limit-defining taxa occur throughout the resources being proposed for application. For example, the ecoregions used for the examples of Chapters 5 and 6 extend over considerable distances in the north-south direction, and climatic differences can be expected to occur throughout such ecoregions. Do the limit-defining taxa occur only within one portion of the ecoregion, or throughout? Such logic can also be applied across stream orders, and across other dimensions that define the extent of water resources proposed for application. Given the potential consequences of a regulatory program, I see the additional assurances that would be provided by such confirmations as prudent and essential.

I have concern with the document's attempt to nationalize an approach developed in Appalachia where issues concerning elevated SC are well studied. Given the lack of understanding that concerns the causal mechanism, given that ions at issue are released to waters due to both natural processes and anthropogenic activities, and given that relationships of aquatic biota to SC/major ions are not well documented in other areas of the country (at least to my knowledge), I am unable to reach the conclusion that the proposed method – and its reliance on either SC or [HCO<sub>3</sub><sup>-</sup> + SO<sub>4</sub><sup>2-</sup>] as a measurement endpoint – could be implemented without unanticipated problems in other areas of the country.

I reach my conclusions concerning the method's adequacy reluctantly due to several related concerns:

- Concern for the effect that a water quality criterion for (SC) would have communities throughout the Appalachian coalfield and the people who live there, given the historic and recent importance of coal mining as an economic activity that brings money into region. The economic and human effects of recent coal-mining declines in these communities are severe, and implementation of a ~300 µS/cm water quality criterion would continue that trend.
- Concern for "regulatory equity," or a lack thereof in this case. As I understand Clean Water Act implementation procedures elsewhere in the US, such as urban, agricultural, and residential areas: Regulatory procedures intended to enforce maintenance of 95% of reference taxa in local streams and rivers are not in place, as the multimetric indices that are commonly used for biomonitoring and bioassessment are developed on a

different basis.

I express these concerns with expectation that a ~300  $\mu\text{S}/\text{cm}$  criterion, if established as a firm limit in the Appalachian coalfield, would fail to incentivize further development and implementation of the mining and reclamation technologies that are intended to reduce mining environmental impacts and improve environmental restoration – the incentive would be to shut the mines down.

I also have concern for environmental quality in the Appalachian coalfields; that concern is informed by recognition that regional ecosystems are among the richest (biologically) and well-preserved non-tropical ecosystems on the face of this earth; and that the scales of mining operations and mining effects are large and growing.

With all of that said: I have reviewed the document objectively and have endeavored to provide my professional and technical opinions without bias.

## II. RESPONSE TO CHARGE QUESTIONS

### Questions 1-3: Data Set Considerations

***1. Ion Matrix Characterization:*** *The ionic composition of water samples represented in the Case Study datasets was dominated by the cations calcium ( $\text{Ca}^{2+}$ ) plus magnesium ( $\text{Mg}^{2+}$ ) and the anions bicarbonate ( $\text{HCO}_3^-$ ) plus sulfate ( $\text{SO}_4^{2-}$ ) ions (Sections 4.1.3, 5.1.3, and 6.1). The Case Study example criteria are derived for an ionic mixture dominated on a mass basis by  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$ . Please comment on when it is appropriate to remove samples from the data set (e.g., ionic mixtures not represented in the data set, or based on physiological rationales). Is it more appropriate to use all the data and note the conditions that are represented by the dataset used to derive the criterion? Please comment on adequacy of the discussions and data analyses provided prior to deriving the Case Study example criteria for  $[\text{SO}_4^{2-}] + [\text{HCO}_3^-] > [\text{Cl}^-]$  on a mass basis and estimating background conductivity to assess geographic applicability (e.g., are different or no data exclusion thresholds more appropriate?).*

Comments: Data from sites with elevated TDS/SC but with ionic composition that differs from the dominant ion matrix (i.e. dominated by  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ ,  $\text{SO}_4^{2-}$ ), should be excluded from the datasets used for the analysis, as described by the document. Scientific literature is clear in demonstrating that the ionic composition of TDS influences the SC/TDS concentration at which toxic effects are observed (Mount et al. 1997). Scientific literature is also clear in documenting that  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ , and  $\text{SO}_4^{2-}$  are the predominant dissolved ions in most Appalachian coal-surface-mine influenced waters (Bryant et al. 2002; Pond et al. 2008; Fritz et al. 2010; Timpano et al. 2011; Agouridis et al. 2012; Bernhardt et al. 2012; Lindberg et al. 2012; Wood and Williams 2013; Pond et al. 2014; Sena et al. 2014).

**2. Catchment Size:** *All data from the example criterion data set that met selection criteria were included in the analyses used to derive the Case Study example criteria regardless of stream size. The confounding analysis in the EPA Benchmark Report and additional analyses provided in Section 3.6.2 (Waterbody Type) of the current draft document indicated no scientific reason to exclude data from streams with large catchment areas (>155 km<sup>2</sup>) primarily because sensitive genera were documented in these large streams, background conductivity estimates were sufficiently similar, and the ionic mixture was the same (dominated by sulfate plus bicarbonate anions). Do the analyses and discussions provided in the aforementioned section provide adequate support for the decision to include all samples regardless of catchment size? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.*

Comments: My discussion below assumes that reference streams are low-order, small-drainage-area streams.

No, the document does not provide adequate basis for including all observations, regardless of stream size. There should be a stream size cutoff, and EPA should provide guidance on an appropriate cutoff. One factor in defining the stream-size range appropriate for the analysis concerns reference sites. As stated on page 2-1, line 23-24: “Genera that are not observed at reference sites ... are excluded from the data set.” Therefore, only streams of size classes where community compositions can be documented as being similar to those at reference sites should be included; or, only taxa found to be both occurring at reference sites and as characteristic of the higher-order streams should be included.

There is a large volume of scientific literature supporting the understanding that aquatic communities and community compositions vary by stream size (e.g. Vannote et al. 1980). Scientific literature documents the taxonomic differences that occur between in the river continuum which extends from headwater (low-order) streams and the higher-order streams commonly known as rivers. For example, Grubauch et al. (1996) refer to the “rapid faunal replacement” that occurs in the mid-order reaches of an Appalachian river continuum; and they cite other studies with similar findings.

The proposal to include both large-stream and small-stream (headwater stream, low-order stream) observations in the analysis dataset is not well supported by the logic in the paragraph starting on page 3-31, line 15.

The first argument cited by the paragraph concerns Ephemeroptera taxa in large streams and cites Appendix B of US EPA (2011a) which states that Ephemeroptera occur at lesser richness in large streams with elevated SC than in large streams with low SC. This fact, in and of itself, is peripheral to the logic proposed by this document which concerns frequencies of occurrence by reference-site taxa. Appendix B (US EPA 2011) does not document that the relevant Ephemeroptera taxa - those occurring in larger streams with low SC but not occurring in larger streams with high SC - are taxa that also occur at reference sites. Even if it did, that additional fact would not provide full support because it does not document that the taxa composition high-SC high-order streams are altered in a manner that exceeds the 5%-of-reference-taxa loss threshold. If both high- and low-order streams are to be included in the analysis dataset, only taxa observed as characteristic

of both high- and low-order streams should be considered in the analysis. If conducting such analysis, the finding that a given taxon occurring at reference sites is also characteristic of high-order streams, should be based on more than a single occurrence by following the logic of the so-called extirpation concentration defined as the 95<sup>th</sup> percentile of capture probability and not as the maximum SC for observed occurrence. This precaution is justified by these organisms' mobility.

The document (paragraph starting on page 3-31, line 15) also states that “conductivity and drainage area are very weakly correlated” within the areas studied. This fact is not of direct relevance to the argument that biological data from rivers and headwater streams should be mingled within datasets that are analyzed using the methods described. The use of “background” SC (i.e. 25<sup>th</sup> percentile) to approximate reference condition for large streams does not help the logic, in my view; “background” SC and “reference condition” are different concepts, with the “reference condition” concept as more restrictive.

The document also states that “Inclusion of the data from large streams did not significantly change the magnitude of the HC<sub>05</sub>”. That statement is supported by citing Suter et al. (2011), which is a conference presentation and not a peer-reviewed manuscript that is accessible to reviewers, potential regulatory commenters, etc.

Most importantly: The method proposed by this document is novel, as admitted by the authors. However, when applied to headwater streams in coalfield areas, it is being applied in a context where numerous studies have found altered benthic macroinvertebrate communities in low-order streams influenced by major ions discharged by coal surface mines; and no peer-reviewed studies I am aware of have found the opposite to be occurring. A comparable body of supporting science does not exist for the higher-order, high-drainage-area streams.

***3. Seasonality: The datasets used in Case Study I and II did not employ weighting to account for seasonal effects. While the vast majority of samples were taken once on an annual basis, further analyses indicated that the effects of seasonality on the example criteria were minor (Sections 4.1.3 and 5.1.3). Do the analyses employed for seasonal effects and corresponding results adequately support the decision not to weight for season? If not, please describe additional analyses and/or discussions needed or identify any shortcomings in the current analyses and/or discussions.***

Comments: I do not agree with the “not weight for season” decision. Research at Virginia Tech (Boehme 2013; Boehme et al. 2013) has demonstrated that composition of benthic macroinvertebrate samples from coalfield headwater streams varies seasonally, both in reference streams and in streams with elevated TDS originating from mining sources. Other research demonstrates seasonal differences in response by a multimetric index to contemporaneous SC in both reference streams and those affected by elevated SC/TDS (Timpano et al. 2011), meaning that community composition differs by season. Also, the document itself demonstrates clearly that SC in non-reference streams varies by season (Figures 4-2, 4-4, 5-2, and 5-4). Hence, I do not see scientific support for analysis using methods described in the document of data sets that mingle samples from different seasons without a seasonality check, such as a check to determine if limit-defining taxa are seasonal.

Section 3.14 describes a seasonality check procedure that compares spring, summer, and “all year” samples. That section defines spring as March – June. Elsewhere, the document describes “summer” as July – October, an unorthodox definition of that season. Does that definition of “summer” also apply in Section 3.14, and in the Case Study 1 and 2 seasonal analyses (Figures A-7 and B-7)? Seasonal definitions should be stated clearly.

Seasonal HC<sub>05</sub> values were developed for Case Studies 1 and 2; spring and summer HC<sub>05</sub> values are similar for Case Study 1 (Figure A-7) but not for Case Study 2 (Figure B-7). On page 5-12 (Case Study 2), the document states: “In the final assessment, due to the similarity at the low end of the genus sensitivity distribution (SD) between the spring HC<sub>05</sub> and the HC<sub>05</sub> based on the full data set, the example ecoregional criteria were derived using all available data, regardless of the time of year they were collected.” Based on Figure B-7, I do not see the seasonal HC<sub>05</sub> values as similar.

In conclusion, I see no justification for a procedure that would mingle data from all seasons with no seasonality check or adjustment for the data’s seasonal distribution. The Case Study 2 results justify the need for consideration of season. The fact that both community composition and water quality vary by season demonstrate that seasons should be considered separately in HC<sub>05</sub> development.

#### **Questions 4-8: Case Studies: Example Criteria Calculations**

***4. Criterion Continuous Concentration (CCC): Please comment on the clarity of the method to derive the XC<sub>95</sub> and HC<sub>05</sub> (Section 3.1, Deriving a CCC).***

Comments: The methods for deriving the values are described clearly, especially when viewed in association with the examples presented.

However, it is not quite clear what the CCC is intended to be within the context of a potential regulatory program. The CMEC is described as the maximum concentration likely to occur at a site where water quality satisfies the CCC 90% of the time, yet the CCC is also described elsewhere as a geometric mean. Which is it?

***5. Criteria Maximum Exposure Concentration (CMEC): The CMEC is the maximum concentration that occurs while meeting the CCC 90% of the time. Does the analysis to derive this maximum exposure concentration (using the subset of data available with temporal resolution requirements described in Section 3.2, Deriving a CMEC), characterize the maximum concentration that will result in meeting the CCC 90% of the time, and is it reasonable to expect it to be a protective upper limit for sites in the data set? What are the strengths and weaknesses of the approach described in Section 3.2 to derive upper limits for the HC<sub>05</sub> values?***

Comments: The logic for the CMEC derivation (Section 3-2) is not presented. Where did this equation come from, and where is the supporting logic? Has the validity of the proposed approach been checked using laboratory bioassays, or with any other method that uses measured data? If so or if not, that should be stated clearly.



The assumption underlying the CMEC calculation appears to be that the CMEC is defined as a maximum concentration that is likely to occur at a site that satisfies the CCC (estimated as the HC<sub>05</sub>) 90% of the time. If one wishes to estimate a maximum concentration that is likely to occur at a site that satisfies the HC<sub>05</sub> 90% of the time, one must know the temporal distribution for the target variable – SC in this case. It appears that the CMEC equation has been derived assuming that SC will vary in time independently and as a normal distribution. Has this been demonstrated with field data? Others have noted that water quality data rarely vary normally and are often autocorrelated (Helsel and Hirsch 2002, see Chapter 12 for temporal analysis).

The proposed site selection procedure for the CMEC derivation is not adequate. The proposed procedure requires:

- At least 6 samples over a given year.
- A minimum of one sample in the spring (low conductivity, March–June), and one sample in the summer (high conductivity, July–October) are included to capture temporal variability.

Desirable changes are:

- To remove the specific date designations from the second bullet, if the document goes forward with an intent for national application. Certainly, both high-concentration and low-concentration periods should be represented; but these periods may vary by time of year among regions, and among years (based on climate variability) for any given region.
- To add an additional criterion: that remaining samples should be evenly distributed throughout the year. If remaining samples are clustered within a given time of year, they will not be representative of the SC variability that occurs throughout the year; and, hence, would not be suitable for estimating a CMEC using statistical procedures.

Also concerning CMEC: What are the units for the Y axes for Figure 4-9 and 5-9? Presumably, the Y axis (standard deviation) is expressed as log<sub>10</sub> SC, is that right? Whatever it is, it should be stated either in the axis label or in the figure caption. Also: If I understand the axis correctly, those numbers look quite low to me – I suggest they be checked.

Also, the CMEC concept is not clearly defined by the document. For example, page xviii (Executive Summary) states “Below the CMEC, sites are expected to meet the CCC 90% of the time; i.e., a conductivity level that is protective of acutely toxic exposures for 95% of macroinvertebrate genera.” This sentence is not written correctly. Similarly, the Glossary defines the CMEC as “In this document, the CMEC is the conductivity level at which the CCC is met 90% of the time.” I think I understand what is meant by these sentences, but the language is not clearly stated.

As an overall comment: I find that logic that underlies the CMEC as thinly supported, considering that its purpose is regulatory program development. The logic being applied here is statistical, not biological; and no biological data are presented as confirmation of results derived from statistical analyses.



**6. Duration: Please comment on the adequacy of the description and justification supporting the duration of the CCC (one year) and CMEC (one day) (see Section 3.3, Estimation of Criteria Duration)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?**

Comments: Section 3.3 discusses studies that are relevant to the duration question, but none of those studies address the question directly. I am not aware of relevant studies other than those discussed by the document.

Answering this question would require continuous monitoring of water quality in association frequent benthic macroinvertebrate sampling, such as the data described by Boehme (2013), Boehme et al. (2013), and Timpano et al. (2013); but I am not aware of analyses by these or other authors that address this question directly.

**7. Frequency: Please comment on the adequacy of the description and justification supporting the estimation of frequency (not to be exceeded more than once in three years on average) (see Section 3.4, Estimation of Criteria Frequency)? What additional key published studies or publicly available scientific reports exist that may be useful in this discussion?**

Comments: Comment similar to the above response to Question 6.

**8. Alternate measurement endpoint: Is the example alternate measurement endpoint ( $\text{HCO}_3^- + \text{SO}_4^{2-}$ ) clear and adequately supported (Appendix F)? If not, please provide a discussion of additional data or analyses needed to support the alternative measurement endpoint. What are the benefits and weaknesses, if any, of using only two anions to describe the measurement endpoint given that ionic regulation in freshwater organisms is affected by the relative amounts of individual ions (i.e., the ionic composition)?**

Comments: The alternate endpoint is clear and is supported by the scientific information that is available at this time. However, additional investigations are warranted as a means of thoroughly documenting the appropriateness of  $[\text{HCO}_3^- + \text{SO}_4^{2-}]$  as a biotic condition indicator that would provide information other than that which is provided by SC.

It is clear that  $\text{HCO}_3^-$  and  $\text{SO}_4^{2-}$  are the two dominant anions in most Appalachian coal-surface-mine influenced waters (Pond et al. 2008; Timpano et al. 2011; Agouridas et al. 2012; Pond et al. 2014; Sena et al. 2014). Since numerous studies have found elevated SC to be closely associated with benthic macroinvertebrate community alterations and taxa losses, it seems quite reasonable to use  $[\text{HCO}_3^- + \text{SO}_4^{2-}]$  as a measurement endpoint – although no more reasonable than use of SC itself. However, I see this relationship as a reflection of the geochemical processes that drive ion release from the mine spoils and not necessarily as a causative indicator. For that matter, the sum of  $\text{Ca}^{++}$  and  $\text{Mg}^{++}$ , which are typically the two dominant cations, could also be used as a measurement endpoint to the same effect.

I do not see support for an argument that  $[\text{HCO}_3^- + \text{SO}_4^{2-}]$  would be a “better” endpoint than SC; I do not presume it to be more or less representative as an indicator of the “actual toxicant” because the actual toxicant or toxicants is/are unknown.

I find the role of  $\text{HCO}_3^-$  in the observed phenomena to be quite puzzling.  $\text{HCO}_3^-$  is often elevated (relative to ecoregion 69 reference levels) in an adjacent ecoregion (Ridge and Valley, 67) due to natural conditions. Therefore, why would similar  $\text{HCO}_3^-$  levels contribute to benthic macroinvertebrate impairments in the coalfields? Are the taxa that different? Is it possible that the ratio of  $\text{HCO}_3^-$  to other ions present acts as an ecotoxicological influence? The fact that Mount et al. (1997) found  $\text{HCO}_3^-$  to be more directly associated with lab-test organism toxicity than most of the other ions studied does support a potential ecotoxicological role for  $\text{HCO}_3^-$ . However, Mount et al. (1997) also found  $\text{Mg}^{2+}$  to be associated with those toxicities; and scientific literature (e.g. Pond et al. 2008 and 2014, and other studies) demonstrates that  $\text{Mg}^{2+}$  is also quite elevated in mining-influenced high-SC streams; and  $\text{Mg}^{2+}/\text{Ca}^{2+}$  ratios are often altered in mining.

I suggest that EPA investigate benthic macroinvertebrate status in mine-influenced  $\text{SC} > 300 \mu\text{S}/\text{cm}$  waters where  $\text{SO}_4^{2-}$  concentrations are quite low and  $\text{HCO}_3^-$  is the predominant anion to determine if biotic condition in such waters is consistent with expectations based on studies to date. The current proposed document would apply  $\text{HC}_{05}$  levels as criteria in such waters but it is not clear that such inclusion is warranted. Such waters have not been represented in any of the existing studies that associate SC levels with biological effects. Mine-spoil leaching studies (Agouridis et al. 2012; Daniels et al. 2013; Sena et al. 2014) indicate that  $\text{SO}_4^{2-}$  concentrations decline with progressive leachings more rapidly than SC/TDS, suggesting that  $\text{HCO}_3^-$  remains as an important solution component and becomes the dominant anion. Hence, one would expect effluents from aging mine-spoil fills constructed with non-pyritic spoils to approach a condition: SC/TDS remains elevated and  $\text{HCO}_3^-$  concentrations are also elevated (relative to reference) but  $\text{SO}_4^{2-}$  concentrations have declined substantially from the elevated levels that characterize leachates from fresh mine spoils to a concentration much lower than  $[\text{HCO}_3^-]$ . To my knowledge, biological effects of such waters have not been studied. If my understanding of mine spoil geochemistry is correct: Frequencies of occurrence by such waters will increase with time as the existing stock of mine-spoil fills that have been constructed throughout central Appalachia age and their leachate chemistries change.

#### **Questions 9-12: Geographic Applicability**

***9. General: Is the process clearly described for assessing geographic applicability of conductivity criteria to a new area (Section 3.6, Assessing Geographic and Waterbody Applicability)? If not, please provide suggested additional description or clarifications. Is the process a reasonable application of the recommendations made by the SAB for geographic extrapolation (see Section 3.6 and Appendix D)? Do the discussions and data analyses (to determine similarity of ionic matrix composition and estimated background conductivity) provided in these sections adequately support applicability of existing criteria to a new area with a similar ionic signature?***

Comments: For the most part, the process is clearly described generally, especially when the Chapter 3 description is viewed in association with the Case Study 3 example. However, the process is not fully supported as a reasonable process for regulatory development as described. Certainly, the described process might be used by resource managers in the “new” ecoregion to inform management decisions, but regulatory development would be a different and more serious application.

The process described depends upon a “matching” of background SCs for two regions. The document’s glossary defines the term “background conductivity” as representing the “conductivity for a region that occurs naturally and not as the result of human activity” so that is clear.

A deficiency in the Geographic Applicability description concerns the process for determining a background SC when using a distribution of water quality data from a given region. As noted by the document: “the 25<sup>th</sup> centile is conventionally used ... However, when land cover modification (or other anthropogenic disturbance) is pervasive, selection of a centile lower than 25% may be justifiable.” The fact that the 25<sup>th</sup> percentile is based on assumptions and not on rigorous analysis should be noted. The document references US EPA (2000a), but US EPA (2000a) only asserts that a percentile within the range of 5<sup>th</sup> to 25<sup>th</sup> percentile can be used to represent a reference value without citing supporting studies or analyses. As stated by US EPA (2000a), page 4-8: “Both the 75<sup>th</sup> percentile for reference streams and the 5<sup>th</sup> to 25<sup>th</sup> percentile from a representative sample distribution are only recommendations. The actual distribution of the observations should be the major determinant of the threshold point chosen.” As far as I can tell, the decision concerning which centile to select as a “background SC” indicator is being left to judgment, yet the outcome of this process could be significant as a determinant of the “new” ecoregion’s HC<sub>05</sub> according to the process described here. A decision which centile to select can make a big difference in the resulting background estimate. In regional databases I have available for analysis, the 25<sup>th</sup> percentiles of SC distributions differ from the 5<sup>th</sup> percentiles by multiples ranging from 3 to 10.

One of the authors of this document has published a peer-reviewed study (Griffith 2014) that derives 25<sup>th</sup> percentiles SC distributions for ecoregions from throughout the US; that study describes a method summarized in the abstract as “followed EPA methods to estimate reference values” but does not describe the resulting values as “background”. In fact, the author states that “Much discussion exists in the literature as to whether estimates like mine are true estimates of background or at least current reference conditions ...”.

The process for assessing geographic applicability of conductivity criteria to a new area is not fully supported. What is missing is a biological confirmation. Are benthic macroinvertebrate communities for the two ecoregions in streams of similar sizes comprised of similar taxa? Do the limit-defining taxa also occur in the “new” ecoregion? A method for evaluating biological data as a means of answering such questions should be described. If background conductivity, background ionic signature, and benthic macroinvertebrate taxa for the two ecoregions are all similar and the limit-defining taxa are present in both ecoregions, it would be reasonable to consider applying an HC<sub>05</sub> value developed for one ecoregion to another.

The underlying assumption of the Geographic Applicability process, as described, is as follows: If background SC estimates for two regions are similar, then sensitivity of regional taxa to elevated SC will be similar as well. The document cites no studies to support the validity of that assumption.

***10. Geographic applicability to a new area within an ecoregion: Please comment regarding the clarity of the process described for assessing geographic applicability of field-based conductivity criteria to locations within the same ecoregion that are outside the geographic bounds of the parent data sets (see Section 3.6). Do the Case Study analyses (Sections 4.3 and 5.3) adequately***

***support the application of the derived example criteria within those ecoregions? If not, please describe why and any additional data and analyses needed.***

Comments: Answer is similar to that provided to Question 9 above: The assumption here is that taxa comprising benthic macroinvertebrate communities within the smaller area are similar to those that occur within the larger area. This validity of this assumption should be verified before extending criteria developed in one area to another, regardless of ecoregion boundaries.

***11. Geographic applicability to a new area in another ecoregion: Please comment regarding the clarity of the applicability analysis for the background-matching approach described in Section 3.6.3 and illustrated in Section 6. Do the data and analyses adequately support the application of the example criteria to other areas? If not, please describe why and any additional data and analyses needed.***

Comments: The background matching concept is clearly described, generally, but certain details are murky.

As mentioned above, the process for defining a “background SC” should be more explicit – if it is to be used as an essential component of a criterion definition process as described by the document.

One detail in the Case Study 3 example is not clear. Multiple data sets were used for Case Study 1 and for Case Study 2. A background SC estimate ( $\pm$  CI) was derived for each of these datasets. Then, Case Study 3 uses a single background SC estimate from each of the two case studies for the “matching” analysis but it does not state a clear rationale for selection. More specifically: the value described as the “the example Criterion data set” (as described on page 4-21) was used to represent Case Study 1 (94  $\mu$ S/cm; 95% CI 86–101  $\mu$ S/cm); however, the background SC estimated for the corresponding data set (“example Criterion derivation data set” as described on page 5-22) was not used to represent Case Study 2; rather, the “WABbase data set, probability sample subset” (147  $\mu$ S/cm; 95% CI 136–159  $\mu$ S/cm) was used to represent Case Study 2. What is the rationale for deciding which background SC estimate to use in procedures such as the Case Study 3 example?

Also, it is not clear to me how the 95% CIs are derived for the background SC estimates. Maybe that procedure is described somewhere in the document but I am not finding it.

In my view, the underlying rationale for use of the background matching approach alone to establish HC<sub>05</sub> values is not clearly supported (as stated above in response to question 9). Justification of the process would require biological confirmation; no process for biological confirmation is described.

***12. Applicability to ephemeral streams: In their 2011 review of the EPA Benchmark Report, the SAB indicated that because the data used to derive the benchmark were collected from perennial streams, the empirical relationship between conductivity and genera occurrence likely would be applicable to perennial and intermittent streams, but not to ephemeral streams. In preparing the current draft document, EPA found several publications that indicate that some aquatic organisms on which the Case Study example criteria are based do occur in ephemeral streams***



*and that these organisms are critical to these headwater systems. EPA also believes it appropriate to include ephemeral waters as applicable water bodies for field-based conductivity criteria in order to ensure protection of aquatic communities in downstream intermittent or perennial waters. Therefore, EPA considers the field-based method applicable to all types of flowing waters, including perennial, intermittent, and ephemeral streams. Do you believe that this recommendation is well supported by Section 3.6.2 (Waterbody Type), including the publications it cites? Are you aware of any additional published studies or publically available scientific reports or data (e.g., paired chemical and biological sampling) relevant to this issue?*

Comments: I answer this question with two caveats: (1) I am not trained in aquatic ecology and have no professional experience dealing with the subject of this question, and (2) I am well aware of legal issues concerning “Waters of the United States,” and that the role of ephemeral streams within that framework is at issue; in answering this question, I take no position on that issue.

Because I lack experience with the precise issue, I reviewed several articles on the topic including some cited by the document and some not. In my reading of the literature, it became clear that there is significant overlap among taxa residing in “temporary” streams (as ephemeral streams are often called in these studies) and those residing in permanent streams. Therefore, it is reasonable to expect that HC<sub>05</sub> levels derived from analysis of biological data collected from low-order permanent streams will be protective of most taxa occurring in ephemeral streams. However, because some taxa occurring in low-order permanent streams do not occur in ephemeral streams, and vice versa, it is not reasonable to expect that the HC<sub>05</sub> levels derived from low-order permanent streams to achieve the goals of the document’s HC<sub>05</sub> derivation method exactly. Hence, it is my view that a process for applying HC<sub>05</sub> levels derived from permanent streams to ephemeral streams would require biological confirmation. One method of biological confirmation could be to verify the presence of the permanent streams’ limit-defining taxa within the ephemeral streams; other methods are also possible.

Articles reviewed to reach this opinion include DeJong et al. (2013), del Rosario and Resh (2000), Delucchi (1988), Feminella (1996), Grubb (2010), Price et al. (2003), Stout and Wallace (2003), and Williams (1996):

**Questions 13-14: Supporting Information: Field-based HC<sub>05</sub> for Fish in Appalachian Streams (Appendix G)**

*13. General: The method used to derive the fish HC<sub>05</sub> generally followed the same field-based method used to derive the macroinvertebrate HC<sub>05</sub> described in the Analysis Plan (Section 3) and in the original EPA Benchmark Report. However, different data sets were used in the fish analysis (Appendix G, Section 2), and some modifications to the method were required to account for differences between fish and macroinvertebrate natural history; e.g., modification to the boot-strapped statistical approach used to characterize uncertainty in the fish XC<sub>95</sub> and HC<sub>05</sub> values (Appendix G, Section 3.4). Please comment on the sufficiency of the data set and the clarity and validity of the modified method to derive the fish XC<sub>95</sub> and HC<sub>05</sub> values.*

Comments: I have no professional expertise or activities that concern fish. I can see that the data analysis procedure is similar. However, my scientific knowledge of fish and of environmental

characteristics that influence their condition and behavior is insufficient to allow my informed opinion about this part of the document.

**14. Protection: Do the analyses for fish (Appendix G) demonstrate that the Case Study example criteria (based on macroinvertebrate data) are protective of fish in those areas? If not, please describe why and any additional data and analyses needed.**

Comments: Same answer as for number 13

### III. SPECIFIC OBSERVATIONS

Page	Line	Comment or Question
Entire report	--	<p><b>Comments Concerning Potential Regulatory Programs that Apply SC = ~300 <math>\mu</math>S/cm as Criterion:</b></p> <p>I offer this comment in the event that EPA does, eventually, issue an SC criteria document; and if such document would be accompanied by regulatory implementation guidelines or guidance to the states.</p> <p>Significant research concerning major ion release by mine spoil fills has been conducted at Virginia Tech, University of Kentucky, West Virginia University, and elsewhere; much has been learned about Appalachian mine spoils and their release of major ions when exposed to environmental processes. Results of these studies are described in a variety of publications including Orndorff et al. (2010), Agouridis et al. (2012), Daniels et al. (2013), Odenheimer (2013), Odenheimer et al. (2013), Evans et al. (2014), Sena (2014), and Sena et al. (2014).</p> <p>These results lead me to conclude that a regulatory program that restricts water discharges to &lt;300 <math>\mu</math>S/cm throughout the mining period is likely to have the effect of serving as an effective prohibition on Appalachian surface mining, as it is unlikely that this level can be achieved on most or all mine sites during the active mining process. It is possible that advanced spoil management and handling methods can be developed that would enable &lt;300 <math>\mu</math>S/cm level to be achieved after mining and reclamation are complete on mine sites that have adequate low SC/TDS spoil materials available for use in constructing soils and hydrologic media. Such practice will require that high SC/TDS materials be placed in non-hydrologic locations, such as within or beneath highly compacted spoil zones; and constriction of functional hydrologic media above those compacted zones. Because materials giving rise to low SC/TDS drainage waters are often highly weathered, such practice would be consistent with associated practices required to restore forest plant communities.</p>
		<p><b>Terminology: Conductivity / Specific Conductance:</b></p>



Page	Line	Comment or Question
report		<p>Throughout, including page xix, Glossary; text under “Conductivity”, and page 2-12, line 22: “specific conductivity” is not a well-recognized term. “Specific conductance” is commonly used in scientific literature to describe the electrical conductivity of water at 25°C:</p> <ul style="list-style-type: none"> <li>• see US EPA, Analytical Methods and Laboratories; <a href="http://water.epa.gov/scitech/methods/">http://water.epa.gov/scitech/methods/</a>, Method 120.1, Conductance (Specific Conductance, umhos at 25°C).</li> <li>• see Hem J.D. 1989. Page 3-18, “Specific Electrical Conductance”.</li> <li>• Consider US EPA Storet Code 00095 (“SPECIFIC CONDUCTANCE (UMHOS/CM @ 25C)”.</li> <li>• Consider also US EPA SW-846, Method 9050A, Specific Conductance.</li> </ul> <p>The text states (page 2-13, lines 11-13) “The term “specific conductivity” indicates the measurement has been standardized to 25°C, a reference temperature (Wetzel 2001).” This may be true, but the Wetzel (2001) reference is far less well known than the references I have cited above. I am not aware of other instances where the term “specific conductivity” is used and accepted. In my experience, the term “specific conductance” is used more commonly to express this concept.</p> <p>Also: Operating instructions for hand-held meters, such as those commonly used by field personnel to monitor water quality often use the term “conductivity” as a measure of the raw electrical conductivity value; and “specific conductance” as the temperature-standardized value. Operators of these meters often have the choice of setting the readout for “conductivity” or “specific conductance.” In my discussions with personnel representing agencies that supervise water quality databases and in my review of such databases themselves, I have observed considerable difficulties when the word “conductivity” is used to represent a water quality variable. Such difficulties include lack of knowledge by certain agency personnel concerning whether the variable is intended to represent actual electrical conductivity of temperature-standardized conductivity (specific conductance); and apparent mingling of records representing both types of measurements under the “conductivity” heading. I strongly encourage EPA to follow what is a routine and well-recognized scientific practice: To reserve the word “conductivity” for raw electrical conductivity measures; and to use the “specific conductance” for temperature-standardized measures. EPA’s current practice (as represented by this document)</p>

Page	Line	Comment or Question
		<p>assigns two different meanings to the word “conductivity.” Perhaps document authors have adopted this practice in response to usage of “conductivity” in Pond et al. (2008)? I call attention, however, to Pond et al. (2014), which uses the term “specific conductance” to designate the temperature-standardized value, as per current scientific practice and most other EPA documents that I am familiar with.</p> <p>In my comments, I have used the term “specific conductance” (SC) to represent a 25°C-standardized value.</p>
Entire report	--	<p><b>Terminology: Extirpation Concentrations:</b> In my view, this term is being used inappropriately.</p> <ul style="list-style-type: none"> <li>Webster defines the term “extirpate” to mean “to destroy completely.” Other dictionaries have similar meanings. That is not an appropriate term here because of the way in which the document quantifies the term: 5% of observations occur at concentrations higher than the so-called extirpation concentration (expressed as XC<sub>95</sub>). The capture probability figures within the document (Figs. 3-1 and 3-4) indicate that individuals of certain taxa are being observed at concentrations &gt;2x XC<sub>95</sub>.</li> <li>Any water quality measure is highly variable in time. The method described here recommends “measurements of the agent(s) should be paired in space and time with biological sampling”. If measurements were timed to acquire samples at the point in time within any given stream where SC is at its highest point during a given genera’s life cycle, the term “extirpation concentration” would be more justifiable – but such targeting is not described by the method presented.</li> </ul>

Page	Line	Comment or Question
Entire report	--	<b>Terminology: Field-based Method.</b> In my view, a term such as “field data analysis method” would be more appropriate because the method described includes no actual field activities data collection; it relies solely on secondary data that have been obtained for other purposes.
Entire report	--	<b>General Comments: Box Plots:</b> <ul style="list-style-type: none"> <li>There are a number of box plots that show distributions of SC and related ions among months throughout the document (e.g. Figs 4-2, 4-3, 4-4, etc.). Given that most of the annual data are distributed quite unevenly among months, I would suggest displaying numbers of observations used to generate each monthly box plot.</li> <li>There are a number of box plots throughout the document. Suggest stating at some point quantities represented by the box plots (I presume median, 25<sup>th</sup>, and 75<sup>th</sup> percentiles for the boxes; 90<sup>th</sup> and 10<sup>th</sup> percentiles for the tails? Are all observations lying outside of the 10<sup>th</sup> and 90<sup>th</sup> percentiles represented as data points, or only some?) Suggest stating the nature of box plot representations explicitly at some point in the manuscript. The caption for the first box plot would be a logical place to do this.</li> </ul>
xxi	--	Glossary: Suggest that the term “Reference Site” be added to the glossary, given the importance of Reference Sites to the proposed method (e.g. Section 3.1.1.2.5. Exclusion of disturbance or pollution-dependent genera: “Genera that are not observed at reference sites or are estuarine or marine organisms are excluded from the data set.”)
2-8	Table 2-1	Table should be annotated to communicate that these are examples only (i.e. this is not a comprehensive or exhaustive list.).
2-15	1-2	“charged particles” are not equivalent to “dissolved ions”.
2-15	28	“Physiological Mechanisms”: My scientific background and training does not enable me to evaluate this section.
2-19	17-20	“Freshwater insects are among the most sensitive organisms relative to other taxa, including zooplankton, fish and amphibians (see Appendix G of this report; Kennedy et al. 2004; Echols et al. 2009b; Lazorchak et al. 2011; Consbrock et al. 2011; Williams et al. 2011).” The statement is poorly supported by the references provided. Lazorchak et al. (2011), Consbrock et al. (2011), and Williams et al. (2011) are citations of conference presentations; hence, supporting documentation is not available to the public or to reviewers and, hence, are inappropriate, in my view, as a means of providing scientific support for a statement with this level of significance in a (potential) regulatory document. Kennedy et al. (2004) compared sensitivity of <i>Isonychia bicolor</i> , a mayfly, to only one other taxon, <i>Ceriodaphnia dubia</i> , which does not typically inhabit the flowing waters where <i>Isonychia</i> are generally found.

Page	Line	Comment or Question
		Echols et al. (2009b) also worked with <i>Isonychia bicolor</i> ; they compared laboratory-derived toxicity values for <i>Isonychia</i> with comparable values obtained from the literature for other species (Table 1), some of which were aquatic insects; and the aquatic insect species were not the most sensitive for most measures. Appendix G derives a species sensitivity distribution (SSD) for fish (Figure G-12); visual comparison of this distribution to the benthic macroinvertebrate SSDs (Figures 4-7 and 5-7) indicates fish as more sensitive throughout most of the SC range. I am not saying the statement in question is in error as I have not looked into that topic with depth. I am saying that statement is of great significance relative to the regulatory program proposed by this document; and the statement is poorly supported as currently presented.
3-5	Section 3.1.1.2	Selection and Adequacy of Data Sets: An additional selection criterion should be that observed SC levels should be well distributed over the population of streams used for the analysis, when those streams are stratified using measured characteristics.
3-6	13-14	“As a general rule of thumb, the minimum sample size to estimate an XC <sub>95</sub> using this field-based method is 25 observations of the genus in the region.” Suggest that this rule of thumb be investigated further to determine if minimum number of observations should be expressed alternatively as a fraction of dataset size. If an SSD dataset were to include 500 samples, 25 observations would constitute 5% of the dataset; but if the SSD dataset were to include 2500 samples, the 25 observations would constitute only 1% of the dataset. Does 25 observations of a taxon remain as an adequate number as dataset size increases?
4-11	7-9	“Samples collected from the WVDEP-identified reference sites indicate that conductivity levels are generally low and similar throughout the year, although slightly higher in summer/fall months of August, September, and October ...” My interpretation of “slightly higher” is not consistent with its use in this sentence. Figures 4-2 and 4-4 indicates that mean SC during the 3 months listed is >2x the mean SC during most other months.
5-12	1-3	Same comment as for page 4-11, lines 7-9.
7-1 and forward	--	Reference formatting is inconsistent.
Entire report	--	Seasonal definitions are not clear. For example, page 4-11, lines 7-9 refer to August, September, and October as “summer/fall months” while page 3-19, line 5 refers to the July-October period as “summer.”

## References cited in

- Agouridis, C., P. Angel, T. Taylor, C. Barton, R. Warner, X. Yu, C. Wood. 2012. Water quality characteristics of discharge from reforested loose dumped mine spoil in eastern Kentucky. *Journal of Environmental Quality* 41:454–468.
- Bernhardt E.S., B.D. Lutz, R.S. King, J.P. Fay, C.E. Carter, A.M. Helton, D. Campagna, J. Amos. 2012. How many mountains can we mine? Assessing the regional degradation of central Appalachian rivers by surface coal mining. *Environmental Science and Technology* 46:8115--8122.
- Boehme E.A. 2013. Temporal dynamics of benthic macroinvertebrate communities and their response to elevated specific conductance in headwater streams of the Appalachian coalfields. M.S. Thesis, Virginia Tech.
- Boehme E.A., S.H. Schoenholtz, C.E. Zipper, D.J. Soucek, A.J. Timpano. 2013. Benthic macroinvertebrate community temporal dynamics and their response to elevated specific conductance in Appalachian coalfield headwater streams. P. 7-22 in: 2013 Powell River Project Research and Education Program Reports. Virginia Tech. [http://www.prp.cses.vt.edu/Reports\\_13/Reports\\_13.html](http://www.prp.cses.vt.edu/Reports_13/Reports_13.html)
- Bryant, G., S. McPhilliamy, H. Childers. 2002. A Survey of the Water Quality of Streams in the Primary Region of the Mountaintop/Valley Fill Coal Mining, October 1999 to January 2001, USEPA Region III, Wheeling, WV.
- Daniels W., Z. Orndorff, M. Eick, C.E. Zipper. Predicting TDS release from Appalachian mine spoils. p. 275-285. In: J.R.Craynon (ed.), *Environmental Considerations in Energy Production*. Society for Mining, Metallurgy, and Exploration. Englewood, CO.
- De Jong, G.D. and S.P. Canton. 2013. Presence of long-lived taxa and hydrologic permanence. *Journal of Freshwater Ecology* 28(2): 277-282.
- del Rosario RB, Resh VH (2000) Invertebrates in intermittent and perennial streams: is the hyporheic zone a refuge from drying? *J N Am Benthol Soc* 19:680–696
- Delucchi, C. M. 1988. Comparison of community structure among streams with different temporal flow regimes. *Canadian Journal of Zoology* 66: 579-586.
- Evans D.M., C.E. Zipper, P.F. Donovan, W.L. Daniels. 2014. Long-term trends of specific conductance in waters discharged by coal-mine valley fills in central Appalachia, USA. *Journal of the American Water Resources Association* 50: DOI: 10.1111/jawr.12198
- Feminella, J.W. 1996. Comparison of benthic macroinvertebrate assemblages in small streams along a gradient of flow permanence. *Journal of the North American Benthological Society* 15: 651-669.
- Fritz K.M., S. Fulton, B.R. Johnson, C.D. Barton, J.D. Jack, D.A. Word, & R.A. Burke. 2010. Structural and functional characteristics of natural and constructed channels draining a reclaimed mountaintop removal and valley fill coal mine. *Journal of the North American Benthological Society*. 29:673-689.
- Griffith M.B. 2014. Natural variation and current reference for specific conductivity and major ions in wadeable streams of the coterminous U.S. *Freshwater Sciences* 33: 1-17.



- Grubaugh J.W., J.B. Wallace, E.S. Houston. 1996. Longitudinal changes of macroinvertebrate communities along an Appalachian stream continuum. *Can J Fish Aquat Sci* 53:896–909
- Grubbs S.A. 2010. Influence of flow permanence on headwater macroinvertebrate communities in a Cumberland Plateau watershed, USA. *Aquatic Ecology*. 45: 185-195.
- Helsel D.R., R.M. Hirsch. 2002. Statistical Methods in Water Resources. U.S. Geological Survey. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation Chapter A3.
- Hem J.D. 1989. Study and Interpretation of the Chemical Characteristics of Natural Water. U.S. Geological Survey, Water Supply Paper 2254.
- Lindberg, T. T., E. S. Bernhardt, R. Bier, A. M. Helton, R. B. Merola, A. Vengosh, and R. T. Di Giulio. 2011. Cumulative impacts of mountaintop mining on an Appalachian watershed. *Proceedings of the National Academy of Sciences* 108:20929-20934, with online supporting information.
- Mount, D. R., J. M. Gulley, J. R. Hockett, T. D. Garrison, & J. M. Evans. 1997. Statistical models to predict the toxicity of major ions to *Ceriodaphnia dubia*, *Daphnia magna*, and fathead minnows (*Pimephales promelas*). *Environmental Toxicology and Chemistry* 16:2009-2019.
- Odenheimer J.L. 2013. Determining a Total Dissolved Solids Release Index from Overburden in Appalachian Coal Fields. M.S. Thesis, West Virginia University.
- Odenheimer J., J. Skousen, L.M. McDonald. 2013. Predicting total dissolved solids release from overburden in Appalachian coal fields. In: J.R. Craynon (ed.). *Environmental Considerations in Energy Production*. Society for Mining, Metallurgy, and Exploration. Englewood, CO.
- Orndorff Z.W., W.L. Daniels WL, M. Beck, M.J. Eick. 2010. Leaching potentials of coal spoil and refuse: Acid-base interactions and electrical conductivity pp 736–766. In: Barnhisel RI (ed.), *Proc Am Soc Min Reclam Ann Meetings*, Pittsburgh, PA. 5–11 Jun. 2010. Amer Soc Mining & Rec
- Pond, G.J., M.E. Passmore, F.A. Borsuk, L. Reynolds, & C.J. Rose. 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus level macroinvertebrate bioassessment tools. *Journal of the North American Benthological Society* 27:717-737.
- Pond, G.J., M.E. Passmore, N.D. Pointon, J.K. Felbinger, C.A. Walker, K.J.G. Krock, J.B. Fulton, W.L. Nash. 2014. Long-term impacts on macroinvertebrates downstream of reclaimed mountaintop mining valley fills in central Appalachia. *Environmental Management*. DOI 10.1007/s00267-014-0319-6
- Price K., A. Suski, J. McGarvie, B. Beasley, J.S. Richardson. 2003. Communities of aquatic insects of old-growth and clearcut coastal headwater streams of varying flow persistence. *Can J For Res* 33:1416–1432
- Sena K.L. 2014. Influence of Spoil Type on Afforestation Success and Hydrochemical Function on a Surface Coal Mine in Eastern Kentucky. M.S. thesis, University of Kentucky.
- Sena K., C. Barton, P. Angel, C. Agouridis, R. Warner. 2014. Influence of spoil type on chemistry and hydrology of interflow on a surface coal mine in the eastern US coalfield. *Water, Air, & Soil Pollution* 225: 1-14.



- Stout B., J.B. Wallace. 2003. A Survey of Eight Major Aquatic Insect Orders Associated with Small Headwater Streams Subject to Valley Fills from Mountaintop Mining. Appendix in Mountaintop Mining/Valley Fills in Appalachia. Final Programmatic Environmental Impact Statement. U.S. Environmental Protection Agency, Philadelphia, PA
- Timpano A.J., S.H. Schoenholtz, D.J. Soucek, C.E. Zipper. 2014. Salinity as a limiting factor for biological condition in mining influenced central Appalachian headwater streams. *Journal of the American Water Resources Association*. DOI: 10.1111/jawr.12247
- Timpano A.J., D. Soucek, S. Schoenholtz, C. Zipper. May 2013. Continuous conductivity monitoring for predicting macroinvertebrate community structure in coal mining-influenced streams. Society for Freshwater Science 2013 Annual Meeting, 19-23 May, Jacksonville, Florida. Abstract ID 7546.
- Timpano A.J., S. Schoenholtz, C. Zipper, D. Soucek. 2011. Levels of dissolved solids associated with aquatic life effects in headwater streams of Virginia's Central Appalachian coalfield region. Final report prepared for Virginia Department of Environmental Quality; Virginia Department of Mines, Minerals, and Energy; and Powell River Project. April 2011.
- U.S. EPA (Environmental Protection Agency). (2000a) Nutrient criteria technical guidance manual: Rivers and streams. Office of Water, Washington, DC. EPA/822/B-00/002.
- Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, C.E. Cushing. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130-137.
- Williams D.D. 1996. Environmental constraints in temporary fresh waters and their consequences for the insect fauna. *J N Am Benthol Soc* 15:634–6