GEI Consultants

Geotechnical
Environmental
Water Resources
Ecological

INTERIM DRAFT REPORT

Technical Review:

A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams

Submitted to:
National Mining Association
101 Constitution Ave NW Suite 500 East
Washington, DC 20001

Submitted by:
GEI Consultants, Inc.
Ecological Division
4601 DTC Boulevard, Suite 900
Denver, CO 80237

July 12, 2010
Project 091380

Table of Contents

List of Figures

List of Tables

Executive Summary

On behalf of the National Mining Association (NMA), GEI Consultants, Inc. (GEI) has conducted a technical review of the U.S. Environmental Protection Agency's (EPA) external review draft of *A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams*. EPA is proposing that the correlation between conductivity and benthic macroinvertebrate community structure in ecoregions 69 and 70 in West Virginia is strong enough that an aquatic life "benchmark" can be derived. They are also proposing that they can use species sensitivity distribution (SSD) based methods, similar to those used to derive numeric ambient water quality criteria for protection of aquatic life and their uses. Based on these two assumptions, EPA used field data from stream benthic macroinvertebrate surveys to derive a proposed aquatic life benchmark of 300 µS/cm conductivity that would be applied to a limited set of specific waters in the Appalachian Region that are dominated by salts of sulfate (SO_4^{2-}) and bicarbonate (HCO_3^-) at circum-neutral to mildly alkaline pH.

We believe there are a number of factual, methodological, and conceptual issues with their proposed benchmark.

First, multiple stressor-response profiles are exhibited by the genera used by EPA to derive the conductivity benchmark and, thus, do not represent an internally consistent dataset from which to derive a regulatory benchmark using an SSD approach. This suggests either that different invertebrate genera exhibit fundamentally different responses to elevated salinity, or that factors other than conductivity are more closely and functionally related to the capture probability of individual genera across the study region. Therefore, use of an SSD of XC_{95} values based on mixed stressor-respond profiles from assumed field distributions is a fundamental flaw in their derivation of a regulatory benchmark.

Second, there are insufficient data from the physiological and toxicology literature to rigorously support EPA's conclusion that "conductivities in the region of concern reach levels that are sufficient to cause effects on stream communities" (p. 52). Although elevated salinity can clearly induce adverse effects on aquatic invertebrates, the taxonomic patterns of sensitivity are not yet clearly defined. Toxicity to ions associated with salinity also varies strongly as a function of specific ion composition and can be mitigated under conditions of elevated hardness. In fact, criteria based on individual ions—rather than those based on composite variables such as conductivity—have already been considered in other states as a preferable regulatory approach that best fits the available scientific information.

Third, the confounding factors analysis in Appendix B of the EPA benchmark document should not take presumed impacts from conductivity *as a given*. Rather, a confounding factors analysis should also include rigorous and independent tests of the primary hypothesis and first determine whether conductivity is indeed the best predictor of biological impairment that is causally related in such a way as to justify the proposed benchmark value. EPA's confounding factor analysis would benefit from a closer evaluation of other factors which could provide alternative explanations for patterns in macroinvertebrate community structure relative to MTM/VF activities:

- *Habitat*: EPA's assertion that habitat presented little potential for confounding in their derivation of the conductivity benchmark is flawed. First, RBP habitat scores are not the most rigorous measure of habitat quality for benthic macroinvertebrates. Second, RBP habitat scores are correlated with both conductivity and the biological response (i.e., HC_{05}). Third, analysis of confounding factors focused almost exclusively on the relationship with Ephemeroptera, to the exclusion of the rest of the benthic macroinvertebrate community.

- *Confounding factors analysis limited to Ephemeroptera*: Relationships between all potential stressors (in addition to habitat) and Ephemeroptera were generally cited as reasons to reject the stressors as potential confounders. There is a clear need to include similar analyses from other members of the entire invertebrate community to conclusively reject additional environmental factors as potential confounding stressors.

- *Influence of rare taxa*: EPA did not sufficiently demonstrate that the rare taxa were rare due to conductivity or any other water quality effect, and not from general rarity itself.

Fourth, we do not agree that the EPA's presumed 95% protection level for the conductivity benchmark is ecologically relevant with respect to changes in functional groupings of macroinvertebrate genera. Evaluation of trends in macroinvertebrate community structure and function relative to conductivity found few observed changes in the proportional abundance of functional feeding groups within the regional pool of taxa until conductivity levels exceeded approximately 2,500 µS/cm to 5,000 µS/cm.

Fifth, our preliminary data analyses from the WABbase dataset indicate that conductivity alone is not the most appropriate parameter when trying to explain the variation observed among the Central Appalachian macroinvertebrate communities with respect to water quality and physical habitat. Rather, ionic composition, substrate, and channel features may be the most appropriate stressor variables to consider. These analyses also indicate that total taxa and percent EPT abundance are the key response variables to consider when evaluating factors that shape the macroinvertebrate community, as opposed to a singular focus on Ephemeroptera. Additionally, total suspended solids, dissolved oxygen, and fecal coliforms appear to be key variables to consider when evaluating these stream sites, as they are strong indicators of other anthropogenic disturbances in the watersheds.

We conclude that the relationship between conductivity and changes in benthic macroinvertebrate community structure is neither strong nor reliable enough to warrant derivation of a regulatory benchmark at this time. While correlations may exist between elevated conductivity and the capture probability of select invertebrate genera, there is insufficient evidence to conclude that elevated concentrations of ions related to salinity are responsible for losses of presumed sensitive taxa. For the most part, this lack of evidence is because EPA did not rigorously or independently test the primary hypothesis that elevated salinity (as measured by conductivity) was the best predictor of changes in macroinvertebrate community structure in West Virginia streams associated with MTM/VF activities. Rather, most of the analysis conducted by EPA takes it as a given that conductivity is the best predictor. Furthermore, insufficient laboratory studies are available to confirm either the causal mechanisms or conductivity thresholds that would confirm the proposed benchmark of 300 µS/cm under the specific ion composition of streams in this region. For similar reasons, Illinois, Indiana, and Iowa have rejected the use of TDS or conductivity-based criteria in lieu of criteria for individual ions such as sulfate or chloride.

We also conclude that the use of a SSD of XC_{95} values based on mixed stressor-response profiles from assumed field distributions is a fundamentally flawed method for derivation of a regulatory benchmark. Additional study is needed to confirm or refute the hypothesis of the conductivity relationship to aquatic life, both through use of additional statistical hypothesis testing with the existing dataset, and through additional study of West Virginia streams associated with MTM/VF activities.

Therefore, we believe it is inappropriate and inadvisable to adopt this conductivity benchmark until or unless such additional study is conducted. To adopt this benchmark without the additional study runs a significant risk of forcing mining operations to expend significant financial resources to reduce conductivity from MTM/VF outfalls, with little confidence that this would achieve the desired goal of preventing extirpation of sensitive genera.

1.0 Introduction

It has been recently proposed that mountaintop mining and valley fill (MTM/VF) activities in West Virginia lead to increases in the conductivity of surface waters located immediately downstream and these increases in conductivity are related to adverse changes in the structure of benthic macroinvertebrate communities (Pond et al. 2008). In particular, reduced abundances of mayflies (the aquatic insect order Ephemeroptera) were considered to be most closely related to elevated water conductivity. The relationships identified in Pond et al. (2008) were based purely on statistical correlations between water quality characteristics and benthic macroinvertebrate community structure and do not represent a formal or mechanistic test of the hypothesis that conductivity (or chemical parameters detected by the composite measure of conductivity) is the primary cause of changes in the macroinvertebrate communities downstream of MTM/VF activities. This and other potentially confounding issues challenging this conclusion were summarized in earlier GEI analyses (GEI 2009a,b).

The U.S. Environmental Protection Agency (EPA) is now proposing that the correlation between conductivity and benthic macroinvertebrate community structure is strong enough that an aquatic life "benchmark" can be derived (EPA 2010) and that the relationship is strong enough that they can use methods similar to those used to derive numeric ambient water quality criteria (AWQC) for protection of aquatic life and their uses (Stephan et al. 1985, hereafter referred to as the "1985 Guidelines"). In this context, aquatic life criteria (or "benchmarks" in the case of the conductivity proposal) represent concentrations of chemicals that, if not exceeded, would ensure protection of aquatic communities at levels set forth in the Clean Water Act (CWA). The draft conductivity benchmark that is the subject of this review was released in March 2010 as an external review draft (EPA 2010). It is anticipated that this, or a similar version of the benchmark document, will also soon be the subject of an external peer review by EPA's Science Advisory Board (SAB) sometime this year.

On behalf of the National Mining Association (NMA), GEI Consultants, Inc. (GEI) has prepared this technical review of the external review draft of *A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams* (EPA 2010). This report uses field data from stream benthic macroinvertebrate surveys to derive a proposed aquatic life benchmark for conductivity that may be applied to waters in the Appalachian Region that, according to EPA, are dominated by salts of sulfate (SO_4^{2-}) and bicarbonate (HCO_3^-) at circum-neutral to mildly alkaline pH. While the EPA states that this conductivity benchmark was derived using a method modeled after the 1985 Guidelines, the use of field benthic macroinvertebrate community data as opposed to individual species laboratory toxicity data represents a significant technical departure from this guidance. Given its potential regulatory implications, this aquatic life benchmark for conductivity must be carefully reviewed to determine whether it represents a scientifically plausible and reliable means of ensuring aquatic life protection.

This technical memorandum summarizes the results of GEI's technical review of the EPA conductivity benchmark document. The primary scope of this review was to evaluate the overall technical basis of how the conductivity benchmark was derived, with a particular focus on evidence presented by EPA in support of the mechanistic plausibility of using conductivity as the basis of deriving a field-based aquatic life benchmark, and the extent to which confounding factors other than conductivity were addressed. GEI's review also presents a preliminary independent statistical evaluation of the ecological factors most likely associated with patterns in benthic macroinvertebrate community structure in West Virginia headwater streams associated with MTM/VF activities. This analysis utilizes the same raw data and field sites used in EPA (2010) to derive the conductivity benchmark, but uses different sets of statistical tools to evaluate the potential impacts of MTM/VF activities.

GEI's technical review consists of the following main elements:

- Summary of the proposed conductivity benchmark and its technical basis

- Summary of GEI's primary concerns with the scientific plausibility and reliability of the proposed conductivity benchmark, including:

    - Diversity of stressor-response profiles and the use of species-sensitivity distribution methods for benchmark derivation
    - Plausibility of physiological mechanisms proposed as causes of extirpation of sensitive taxa
    - Analysis of confounding factors other than conductivity
    - Ecological relevance of the protection level intended from the proposed conductivity benchmark

- Independent statistical evaluation of ecological factors most closely associated with patterns in benthic macroinvertebrate community structure

2.0 Proposed Conductivity Benchmark

2.1 Summary of Proposed Benchmark

EPA (2010) used field data to derive an aquatic life benchmark for conductivity that is intended to be applied only to a limited set of waters in the Appalachian Region that are dominated by salts of SO_4^{2-} and HCO_3^- at circum-neutral to mildly alkaline pH (see benchmark definition and limitations below). It is derived by a method modeled on the EPA's standard methodology for deriving AWQC, except that the methodology was substantially altered for use of field data. Field data were used because EPA stated that sufficient and appropriate laboratory data were not available and "high quality" field data were available to relate conductivity to effects on aquatic life.

The method used in EPA (2010) has the appearance of being based on the 1985 Guidelines primarily because it used the 5^{th} percentile of a species sensitivity distribution (SSD) as the basis for mathematical derivation of the benchmark value. An SSD represents the response of aquatic life as a distribution with respect to exposure and is a widely used statistical approach for derivation of regulatory aquatic life criteria worldwide. It is implicitly assumed that if the exposure level is kept below the 5^{th} percentile of the SSD, at least 95% of tested aquatic species (or their surrogates) will be protected. In this respect, EPA's data analysis followed the standard methodology in aggregating species to genera and using interpolation to estimate the percentile.

However, the method used in EPA (2010) differs significantly from the 1985 Guidelines in that the points in the SSDs are so-called extirpation concentrations (XCs) rather than median lethal concentrations (LC_{50}s) or chronic values from exposure to a single chemical. The XC is defined by EPA as the level of exposure above which a genus is "effectively absent" from waterbodies in a region. For this benchmark value, the 95^{th} percentile of the distribution of a calculated "probability of occurrence" of a genus with respect to conductivity was used as a 95^{th} percentile extirpation concentration (XC_{95}). Hence, this aquatic life benchmark for conductivity is expected to avoid the local extirpation of 95% of native species (based on the hazardous concentration [HC_{05}] – the 5^{th} percentile of the SSD) in surface waters that include neutral to alkaline effluents containing a mixture of dissolved ions "dominated by salts of SO_4^{2-} and HCO_3^-."

The chronic aquatic life benchmark value for conductivity derived using all-year data from West Virginia was calculated to be 300 µS/cm. According to EPA (2010), this benchmark is only applicable to parts of West Virginia and Kentucky. They expect it to be applicable to the same ecoregions in Ohio, Pennsylvania, Tennessee, and Maryland, but data from those states have not been analyzed. EPA states that this benchmark could also be appropriate for other nearby regions such as Ecoregion 67, but has only been validated for use in Ecoregions

68, 69, and 70 at this time. However, EPA further states that this level may not apply when the relative concentrations of dissolved ions are not dominated by salts of SO_4^{-2} and HCO_3^{-}.

2.2 Analysis of Causal Mechanisms and Confounding Factors

Because numeric aquatic life criteria are based on laboratory toxicity tests from single chemical exposures, the causes of biological impairment (i.e., toxicity) are generally clear, test results are repeatable, and confounding factors are minimized or eliminated under controlled laboratory conditions. However, associations between biological patterns as a function of one or more chemical stressors in the field are not necessarily causal, nor are they free from other factors that may confound or obscure the presumed association. Therefore, EPA conducted a causal assessment (Appendix A of EPA 2010) based on epidemiological approaches (e.g., Hill 1965) and EPA guidance for conducting stressor identification and diagnosis (EPA 2000; www.epa.gov/caddis). From these assessments, EPA concluded that the available evidence indicated that salts, as measured by conductivity, are a common cause of impairment in aquatic macroinvertebrates in the region of concern (i.e., Ecoregions 68 and 69 of West Virginia).

EPA also conducted a confounding factors assessment (Appendix B of EPA 2010) to evaluate the extent to which variables that may co-occur with conductivity might limit or alter their ability to quantify the effects of conductivity (i.e., derive a quantitative benchmark). A weight of evidence approach was used to evaluate each confounding factor and, to the extent possible, test whether removal of confounding factors might alter the ultimate derivation of the conductivity benchmark. EPA concluded that "the effect of confounders was found to be minimal and manageable," and that only the elimination of sites with pH < 6 was needed to remove this potentially significant confounding factor.

3.0 Primary Technical Concerns

The EPA (2010) conductivity benchmark represents a fundamentally different application of an SSD approach for derivation of a regulatory benchmark. In particular, the proposed conductivity benchmark is based on field surveys and correlations between the stressor and biological response in uncontrolled field environments, with multiple species present and all possible biotic (predation/competition/etc.) and abiotic (temperature/flow/season/etc.) interactions occurring. Most other regulatory thresholds (including AWQC) are derived using laboratory data on individual species in which the relationship between the stressor and the biological response are directly manipulated and studied in a controlled manner following prescribed protocols.

As summarized in EPA (2010), there may be some advantages to using a field-based approach. Because it is based on biological surveys, it may be more directly relevant to the streams where the benchmark may be applied and represent the actual aquatic life use in these streams. Another advantage is that field-based biological measurements of whole communities integrate the effects of all life stages and ecological interactions (although only for benthic macroinvertebrates). Further, the data represent actual exposure conditions in the region, the actual temporal variation in exposure, and the actual mixture of ions that contribute to salinity as measured by conductivity.

However, there are several disadvantages to the field-based approach that greatly limit the scientific reliability of using this approach to set specific regulatory thresholds for a composite water quality measurement such as conductivity. The primary disadvantages of using field data result from the fact that exposures are not controlled, so the causal nature of the relationship between conductivity and associated biological responses are very difficult to evaluate. As discussed below, EPA's arguments supporting the mechanistic plausibility of conductivity as the (virtually only) cause of "impairment" are relatively weak, and so cast considerable doubt on the overall reliability of its calculated conductivity benchmark.

Furthermore, any chemical or biological variables that are correlated with conductivity or the biotic response may confound the presumed relationship between conductivity and biological impairment. To address this, EPA (2010) conducted a relatively formal analysis of confounding factors—which was an improvement over the purely statistical approach taken by Pond et al. (2008). EPA concluded that although plausible confounding factors likely exist, their influence is not *strong enough* to prevent use of the conductivity benchmark as presented in this document. We do not agree that all of the confounding factors are so easily dismissed. In fact, we believe many of the confounding factors require a more in-depth analysis to evaluate whether or not conductivity alone is in fact a strong enough indicator of adverse changes in biological communities to allow for its use in derivation of a regulatory benchmark.

Below we summarize primary technical concerns with the proposed EPA conductivity benchmark, not presented in any particular order of priority. These concerns are focused around four primary questions:

- What is the diversity of stressor-response profiles exhibited in EPA's SSD of XC_{95} values, how might this influence the benchmark value derived from this SSD, and is this a valid approach?
- Is the underlying assumed mechanism for impairment—toxicity from ions associated with salinity—mechanistically plausible and is the proposed benchmark value consistent with thresholds obtained in toxicity tests?
- Is the confounding factors analysis convincing, i.e., do we agree that factors correlated with conductivity do not substantially confound or obscure biological relationships with conductivity?
- What is the ecological relevance of the proposed benchmark value?

3.1 Diversity of Stressor-Response Profiles

Multiple stressor-response profiles are exhibited by the genera used in EPA (2010) to derive the conductivity benchmark and, thus, do not represent an *internally consistent* dataset from which to derive a regulatory benchmark. Three types of stressor-responses are recognized by EPA (2010) as exemplified in their Figure 5 (p. 30), where the caddisfly *Lepidostoma* exhibits a declining probability of observation with increasing conductivity, the stonefly *Diploperla* exhibits a "bell-curve" or optimal level of conductivity, and the caddisfly *Cheumatopsyche* exhibits an increasing probability of observation with increasing conductivity. In addition to these three stressor-response profiles, a fourth type not recognized by EPA is characterized by basically no response or a bimodal response to conductivity. These are exemplified in the empidid flies *Clinocera* and *Chelifera* (Figure D-1; EPA 2010).

It is understood that there is some subjectivity in the interpretation of the shape of the individual stressor-response profiles. However, the increasing, optimum, and decreasing stressor-response profiles are each well represented in the dataset used in the SSD to derive the benchmark (Table 1), with less than half of the taxa in EPA's final dataset exhibiting the more "typical" stressor-response profile of decreasing abundance with increasing levels of the presumed stressor. Specifically, there are 66 genera (44%) in the database which exhibit a decreasing stressor-response profile (i.e., exhibit decreasing probability of capture with increasing conductivity)—this would be the expected "dose-response" for all taxa if conductivity is the primary stressor, as EPA postulates. Yet, there are 41 genera (27%) that exhibit an optimum conductivity level (i.e., capture probability decreases at both low and high conductivity). And finally, there are 30 genera (20%) that exhibit an increasing stressor-response profile (i.e., capture probability increases with increasing conductivity). Interestingly, 14 genera (9%) exhibit no response or a bimodal response (i.e., capture probability highest at both low and high conductivity). Thus, 56% of the taxa used by EPA to generate their SSD do not actually show the classic dose-response of decreasing probability of capture with increasing conductivity—yet were still used by EPA in their SSD calculations.

Table 1: Number of genera with identified stressor-response profiles.

Response	All	Decreasing	Optimum	No Response/Bimodal	Increasing
Number of Genera	151	66 (44%)	41 (27%)	14 (9%)	30 (20%)

The EPA dataset includes organisms from numerous taxonomic orders. The best represented groups include Ephemeroptera, Plecoptera, Trichoptera, Coleoptera, and Diptera. Each of these was represented by multiple stressor-response profiles, although the Ephemeroptera and Plecoptera were only represented by the decreasing and optimum profiles (Table 2). Interestingly, the Plecoptera and Trichoptera tended to have higher proportions of decreasing stressor-response profiles than the Ephemeroptera, with 76 and 63%, respectively.

Table 2: Number of genera in selected insect orders with identified stressor-response profiles.

Taxa	All	Decreasing	Optimum	No Response/Bimodal	Increasing
EPHEMEROPTERA					
Number of Genera	25	12 (48%)	13 (52%)	0	0
PLECOPTERA					
Number of Genera	17	13 (76%)	4 (24%)	0	0
COLEOPTERA					
Number of Genera	11	2 (18%)	5 (45%)	1 (9%)	3 (27%)
TRICHOPTERA					
Number of Genera	16	10 (63%)	1 (6%)	1 (6%)	4 (25%)
DIPTERA					
Number of Genera	64	22 (34%)	14 (22%)	11 (17%)	17 (27%)

Mixed stressor-response profiles by individual genera strongly indicate that conductivity is not a dominant or singular stressor that can be shown to limit the entire macroinvertebrate community above a single fixed threshold or benchmark value. This suggests that either different invertebrate genera exhibit fundamentally different responses to elevated salinity, or factors other than conductivity are more closely and functionally related to the capture probability of individual genera across the study region. Therefore, the inclusion of all taxa from the dataset, regardless of their stressor-response profile to conductivity, may be inappropriate for the derivation of a benchmark based on an SSD approach.

Furthermore, for 20% of the total genera that exhibit increasing stressor-response profiles, the proposed benchmark of 300 µS/cm could actually be interpreted as harmful—i.e., not protective. The proposed benchmark "is intended to prevent the extirpation of 95% of invertebrate genera" (EPA 2010; p. 20), and yet many waters with conductivity below this benchmark would be associated with very low capture probabilities (i.e., "extirpation" in the context of the proposed EPA benchmark) for taxa with increasing stressor-response profiles. EPA assigned "greater than" XC_{95} values for these taxa, which would imply the negative effects of salinity would not occur until conductivity values exceeded those measured in the West Virginia dataset. An alternative hypothesis is that extirpation for these species would occur at low conductivities, so the benchmark would clearly not be "protective" with respect to their presence in stream sites in the region.

From an analysis of the cumulative distribution plots in Appendix D, the following taxa, from multiple orders, exhibit increasing stressor-response profiles and so are poorly represented in waters with low conductivity: *Corydalus* (Megaloptera), *Atherix*, *Phaenopsectra*, *Dicrotendipes* (all Diptera), *Dubiraphia* (Coleoptera), *Chironomus* (Diptera), *Physella* (Mollusca), *Ochrotrichia* (Trichoptera), and *Krenopelopia* (Diptera). EPA (2010) Table 1 (p. 24) indicated that the minimum conductivity value measured in streams used for derivation of the conductivity benchmark was 15.4 µS/cm, and the 25th percentile of conductivity values was 153 µS/cm. Additionally, EPA's (2010) Figure A-1 (p. 64) showed a large number of sites had conductivity values below which the above-listed taxa would not be expected to be found. Therefore, these taxa would likely not be present if conductivities were limited to < 300 µS/cm. Thus, it is not appropriate to say that the proposed benchmark would prevent extirpation of 95% of taxa when 20% of the genera in the dataset exhibit increasing stressor-response profiles and would potentially be absent if conductivity were restricted to those levels.

3.2 Comparison of Derivation Method to Typical Aquatic Life Criteria

Although EPA (2010) states that methods for derivation of the conductivity benchmark are "based upon" the 1985 Guidelines for derivation of AWQC, the only similarity is that both derive the benchmark/criteria concentration as the 5th percentile of species sensitivities using an SSD. The SSDs used for derivation of AWQC are based on laboratory toxicity data from studies in which the stressor (e.g., toxic chemicals) is empirically and unambiguously manipulated in studies that follow standardized and scientifically valid protocols for individual species. The biological endpoints used (based on survival, growth, or reproduction) are derived based on relatively uniform and consistent stressor-response profiles generated from the laboratory toxicity tests. In other words, even though individual organism "sensitivities" to a given chemical will differ from one another, they all must demonstrate similar kinds of monotonically increasing adverse effects in response to increasing chemical exposure concentrations (i.e., each test must exhibit a consistent "dose-response") to be included in the SSD. To help ensure that the total range of chemical sensitivities of organisms likely to be encountered in a broad range of field conditions is represented, a minimum database of eight specific types of aquatic genera (i.e., the "eight family rule") is required before an AWQC can be derived (Stephan et al. 1985).

In contrast, the conductivity benchmark is derived based on a large number of macroinvertebrate genera, but does not represent the total range of aquatic organisms that inhabit these ecosystems. The 1985 Guidelines require data from fish, planktonic crustaceans, and aquatic plants or algae to ensure protection of all aquatic life, so a benchmark based only on benthic macroinvertebrates will not necessarily represent a concentration that is protective of the entire aquatic ecosystem.

Furthermore, as stated in Section 3.1, the SSD used to derive the conductivity benchmark is not based upon a consistent set of stressor-response profiles. We believe that there is a fundamental problem associated with using SSDs to derive an HC_{05} level on the basis of such a mixed set of stressor-response profiles. An SSD is intended to depict the distribution of sensitivity across numerous genera to a single stressor (Posthuma et al. 2002). Individual species responses are typically scaled and ranked to represent the cumulative probability of responding to a given level or concentration of that stressor. However, if some of the genera on the SSD are responding to the stressor in fundamentally different ways (see Section 3.1), then it is inappropriate to include them in the same SSD. Additionally, since the derivation of this benchmark is based solely on field-derived data, the same stressor (conductivity) may not be accurately depicted in the XC_{95}s and SSD. For example, it may be that those genera with narrow optima or increasing stressor-response profiles are, in fact, responding more strongly to something other than conductivity within the ranges of conductivity being observed. Further, it is very clear for those genera that have no response or a bimodal response that their distribution is reflecting something other than conductivity. Therefore, the use of an SSD of XC_{95} values based on mixed stressor-respond profiles from assumed field distributions is a fundamentally flawed method for derivation of a regulatory benchmark.

3.3 Physiological Mechanisms Causing Extirpation

Physiological stress from inorganic ions related to salinity was cited by EPA as one of the most plausible mechanistic reasons supporting use of a conductivity benchmark for macroinvertebrate community impairment. In Appendix A of EPA (2010), salinity is regarded as the mechanistically plausible, primary cause of macroinvertebrate community impairment. Salinity (and its resulting empirical measure, conductivity) is a property of water that represents the total concentration of dissolved mineral salts or “major ions”, including Na^+, Ca^{2+}, Mg^{2+}, K^+, Cl^-, HCO_3^-, CO_3^{2-}, and SO_4^{2-}. One of the basic premises of EPA (2010) is that elevated concentrations of these ions causes physiological stress in macroinvertebrates, ultimately leading to the extirpation of the most sensitive species in waters with conductivity levels that exceed the proposed benchmark of 300 µS/cm.

In this section, we review relevant portions of Appendix A of EPA (2010) with respect to our own understanding of the relevant scientific literature. First, we review Section A.2.3, *Interaction and Physiological Mechanisms*, which discussed the importance of disruption of osmoregulation and ionic homeostasis in aquatic organisms exposed to increased concentrations of inorganic salts. We then review Section A.2.4, *Specific Alteration*, which included relevant information concerning the sensitivity or tolerance of some taxa to conductivity. Finally, we review Section A.2.5, *Sufficiency*, which examined laboratory tests of ionic mixtures, mine discharges, and ambient waters in valley fill regions.

3.3.1 Interaction and Physiological Mechanisms (Section A.2.3 of EPA 2010)

3.3.1.1 Effects of salinity on osmoregulation and ionic homeostasis of aquatic organisms

The importance of osmoregulatory mechanisms in maintaining ionic balance (i.e., homeostasis) within all freshwater invertebrates is well documented. Most of the documents cited by EPA in this section are animal physiology and biochemistry books that adequately address the topic. As we discuss below, however, the specific comparative mechanisms of osmoregulatory disruption in different taxonomic groups which might be used to explain differential tolerance to salinity are not yet well understood. Ultimately, we suggest there is a need for further investigation of sub-lethal responses to high salinity exposure for a wide variety of sensitive and tolerant freshwater invertebrate taxa to confirm the mechanistic reasons which might explain taxonomic patterns of sensitivity.

Our review provided additional literature not cited by EPA, which discussed active ion absorption by specialized body structures, particularly in snails, mussels, leeches, dragonfly nymphs, crayfish, and some dipteran larvae (Smith 2001). In the insects, these structures include individual chloride cells, fields of chloride cells (known as chloride epithelia), and other absorptive structures on papillae or within the gut system. Individual chloride cells are present in some members of the Ephemeroptera, Plecoptera, and Hemiptera, while chloride epithelia are present in some members of the Trichoptera, Odonata, and Diptera (Komnick 1977). In addition, some Diptera, Trichoptera, and Coleoptera have intestinal or papillar ion absorption sites (Komnick 1977). Osmoregulation by aquatic invertebrates may also depend upon integument permeability, which varies by taxonomic group (Pennack 1978, cited in Pillard et al. 1999) and age, because older organisms may have thicker and less permeable surfaces (Pillard et al. 1999).

EPA (2010) acknowledged the existence of numerous physiological mechanisms involved in the toxicity of high conductivity waters. EPA (2010) mentioned mortality as one of the effects of elevated salinity, citing Kefford et al. (2003, 2005a). We confirmed the accuracy of the first citation; however, Kefford and Nugegoda (2005) (as the citation appeared in the EPA references section, page 54, not Kefford et al. 2005a which is an incorrect citation) reported sublethal effects (growth and reproduction), and not mortality, from elevated salinity to the fresh water snail *Physa acuta*. EPA (2010) also described sublethal effects from salinity, such as reduced growth, reproduction, and hatching success. EPA (2010) cited Clark et al. (2004), but this study showed salinity had opposite effects on two mosquito species, indicating differing or inconsistent physiological responses. A negative effect on growth rate due to increased salinity was observed in *Aedes aegypti*, while increased salinity had a positive effect with increased pupal mass in *Ochlerotatus taeniorhynchus* (Clark et al. 2004).

Several studies suggest tolerance of Diptera and Crustacea to elevated salinity. Dipterans are the most diverse group of insects in the aquatic environment (40% of insect taxa), and they possess morphological adaptations, such the presence of papillae, for regulation of salt concentration (Thorp and Covich 2001). Furthermore, because some dipteran species are capable of hypo-osmotic regulation, they are the only insect order that has successfully colonized highly saline waters (Hart et al. 1991). Some amino acids and carbohydrates in the hemolymph of the mosquito *Culex tarsalis* have an osmoregulatory role, allowing them to adapt to water with increasing salinity (Bradley and Garret 1986, cited in Natochin and Parnova 1987).

Freshwater crustaceans are relatively competent osmoregulators (Thorp and Covich 2001). The extreme tolerance of some freshwater decapods (*Amarinus lacustris*, *Paratya australiensis*, and *Caridina nilotica*) to salinity may be phylogenetically derived (Kefford et al. 2003). Adaptations in decapods include the epicuticular layer of the gill laminae, with a high selectivity to Cl^- and OH^- over all other ions normally present in freshwater and in the hemolymph (Avenet and Lignon 1985). Additionally, cladocerans frequently demonstrate very refined physiological adaptations to elevated salinity, as effective as those of decapods or even teleost fish (Aladin and Potts 1995).

Some studies reported negative effects of elevated salinity or conductivity on some Plecoptera, Nematoda, Oligochaeta, and Hydracarina. Salinity tolerance and osmotic stress were evaluated on the nymphs of the stonefly *Paragnetina media*, where mortality reached 80% in 1.2% NaCl (382 mOsm/L) and survivors after the 72-h exposure had slightly hyperosmotic hemolymph when compared to the medium (Kapoor 1979). Piscart et al. (2006) observed changes in the distribution of macroinvertebrate life history traits (i.e., salinity preferences, maximum size, life cycle duration, reproduction, potential generations per year, respiration, dispersal, and feeding habits) among sites with varying salinity in France. For example, taxa with multivoltine life cycles, asexual reproduction, ovoviviparity, and filter-feeding traits were more frequent at sites with higher salinity levels. The authors concluded that salinity promotes more generalist and permanently aquatic taxa and the reduction of specialized, semi-aquatic taxa (Piscart et al. 2006).

Although several studies (Hassell et al. 2006; Kefford and Nugegoda 2005; Kefford et al., 2004, 2006, 2007) evaluated the effects of elevated salinity on other types of aquatic macroinvertebrates, the authors do not claim conclusive support of negative effects, but rather recommend further testing. In addition, EPA (2010) cited Zalizniak et al. (2007) as stating that reduced population density occurs over generations after elevated conductivity exposure. Based on the journal and article title from the Literature Cited section of EPA (2010), this citation should have been Zalizniak et al. (2009). We were unable to find any statement in Zalizniak et al. (2009) that supports this conclusion.

In summary, osmoregulatory mechanisms for maintaining ionic balance in freshwater invertebrates are well documented. However, the specific comparative mechanisms of osmoregulatory disruption in each of these taxonomic groups which explain differential tolerance to salinity are not yet well understood. There is a need for further investigation of sub-lethal responses to high salinity exposure for a wide variety of sensitive and tolerant freshwater invertebrate taxa to confirm the mechanistic reasons which might explain taxonomic patterns of sensitivity.

3.3.2 Specific Alteration (Section A.2.4 of EPA 2010)

3.3.2.1 Sensitivity and tolerance of specific genera to salinity

In Section A.2.4 of EPA (2010), strong, relevant, and consistent evidence supporting the specific effects of elevated conductivity on benthic invertebrates, particularly Ephemeroptera, is purported to exist. However, as discussed above, there is a lack of physiological studies to explain the specific mechanisms of ion toxicity and the reported higher sensitivity of Ephemeroptera to salinity compared to other macroinvertebrates. As noted above, some studies have found increased sensitivity of Ephemeroptera, as well as Plecoptera, and Trichoptera, taxa to higher salinity levels (Kefford et al 2003, 2004); however, there is considerable variability even within these orders, and little is known about the physiological mechanisms that drive the proposed sensitivity of these taxa.

EPA (2010) relied on relatively few studies that evaluated empirical relationships between field occurrence of Ephemeroptera and water chemistry. For example, Pond (2010) (incorrectly cited in EPA 2010 as Pond 2009) evaluated data from 92 headwater streams in the Appalachian Mountains of Kentucky to explore and describe regional patterns of diversity and distribution of lotic Ephemeroptera in relation to two stressors: coal mining and rural residential land uses. Although Pond (2010) demonstrated a strong correlation between low population densities and taxa richness of mayflies and specific conductance in regions of MTM/VF, the study also suggested that other sources of toxicity to mayflies, including exposure to heavy metals, nutrients, organic waste due to bacterial infestation, and a mixture of potentially harmful chemicals, could also have existed. Therefore, Pond (2010) suggested using specific conductance data, in conjunction with a human disturbance metric, to predict mayfly abundance and richness. EPA (2010) also cited Pond et al. (2008), which concluded that MTM/VF causes downstream biological degradation, given the changes on landscape, hydrology, and potential toxicants discharged. However, Pond et al. (2008) also recognized that additional studies are needed to test ambient downstream waters and synthesized waters that would mimic the ionic components of downstream of mines but would not contain any other toxicants (e.g., metals).

In summary, the evidence cited by EPA in support of the specific alteration of presumed sensitive taxa (e.g., mayflies) to elevated conductivity is based only on correlations between field abundance and water chemistry from Pond (2010) and not from experimental studies. In general, there is a lack of physiological or other laboratory studies to explain and/or confirm the sensitivity of the Ephemeroptera, Plecoptera, and Trichoptera taxa to increased salinity. Therefore, it is difficult to confirm from these data whether the presumed effects from conductivity are strong, relevant, consistent, and of high quality.

3.3.3 Sufficiency (Section A.2.5 of EPA 2010)

3.3.3.1 Laboratory tests of defined ion mixtures

In Section A.2.5, EPA (2010) evaluated evidence that laboratory-based exposure to salinity would cause adverse effects to invertebrates (especially mayflies) at concentrations near or above the proposed conductivity benchmark. There are substantial differences in the toxicity of major ion salts; therefore, it would be expected that there would be differing toxicity in waters of different ionic composition (Mount et al. 1997). For example, Pillard et al. (1999) found that K^{+}, Mg^{2+}, and HCO^{3-} are the most acutely toxic ions to freshwater organisms; however, ion toxicity is not just a function of the total concentration of any one ion, but also of the balance or ratios between individual cations and anions in any given aqueous solution. This was demonstrated by Mount et al. (1997), which found that the most toxic combination of salts was a 1:1 mixture of K_2SO_4 and $KHCO_3$. The LC_{50} values reported for this ion combination were 390 mg/L for *Ceriodaphnia dubia* and 720 mg/L for *Pimephales promelas* (Mount et al. 1997). EPA (2010) reported that each of these LC_{50}s for *C. dubia* and *P. promelas* corresponds to 438 µS/cm and 1,082 µS/cm, respectively, although the basis of their conversion of ionic concentrations to conductivity is not clear.

Toxicity studies in Mount et al. (1997) were used to derive a salinity/toxicity relationship (STR) model to predict the acute toxicity of specific combinations of major ions related to salinity. EPA (2009; with the same result summarized in EPA 2010) used the STR model to suggest that salt mixtures in some streams below MTM/VF would cause acute lethality in *C. dubia*. However, the STR analysis provided in EPA (2009) does not appear to be entirely correct. EPA (2009) stated that more than 75% mortality is predicted for *C. dubia* using maximum concentrations for ions reported downstream of MTM/VF in Pond et al. (2008); however, one portion of the STR equations listed in EPA (2009), which were apparently used for these calculations, is incorrect[1]. To evaluate this further, we used the maximum ion concentrations from Pond et al. (2008) as inputs to the correct version of the STR model, and found that the predicted mortality for *C. dubia* is actually 57.4%. Regardless, a salt mixture based on the maximum values from a large dataset does not necessarily represent the salt mixture of an actual site or water sample. Therefore, toxicity predictions from a "mixture" based on the maximum concentration of each ion has limited environmental relevance.

[1] The equation for *C. dubia* 48-hr mortality in Mount et al. (1997) is in the form of the regression constant, 8.83 *plus* the remaining equation terms, whereas EPA (2009) shows this same equation in their footnote 11 as 8.83 *times* the remaining equation terms.

A more relevant approach would be to use the STR model to predict toxicity of actual water sample chemistries from the dataset used to derive the conductivity benchmark. Therefore, for sites in which such data were available, the concentrations of ions needed to run the STR model were compiled from the WVDEP data for all sites used in derivation of the conductivity benchmark. Notably, potassium concentrations were not available for any of these sites, so STR model runs were conducted at the mean "mined" site concentration (9.9 mg K/L) as reported in Pond et al. (2008) to be conservative. These data were then used to predict 48-hr LC_{50} values for *C. dubia* using the STR model (Mount et al. 1997), and plotted against conductivity for the same water samples (Figure 1). Additional STR model runs were also conducted at mean unmined potassium concentrations, but results did not differ substantially, and so are not presented here.

STR model predictions from the natural water chemistries demonstrated a consistent pattern of decreased percent survival when plotted against conductivity, but significant (i.e., < 90% survivorship) mortality only occurred as conductivity values exceeded 1,000 µS/cm (Figure 1). It should be cautioned that the STR model may not accurately predict the toxicity of ions in these mining impacted natural waters given that empirical effluent and simulated effluent tests cited below suggest chronic LOEC concentrations for *C. dubia* that are three to four times higher than the acute toxicity predictions from the STR model. Additional study is needed to determine the full extent to which the STR model accurately represents chronic toxicity to sensitive organisms.

Other studies not cited in EPA (2010) investigated the toxicity of various ion mixtures to freshwater invertebrates in laboratory waters. Soucek and Kennedy (2005) found that increasing chloride concentrations reduced the toxicity of sulfate to *Hyalella azteca*, and increasing water hardness ameliorated sulfate toxicity to both *H. azteca* and *C. dubia*. Further studies on the relationship between chloride and sulfate showed that increasing chloride reduced sulfate toxicity over the 5-25 mg/L chloride range, but resulted in increasing mortality over the 25-500 mg/L range (Soucek 2007). In addition, it was determined that the STR model does not account for the protective effect of elevated hardness on TDS toxicity (Soucek 2007). As a result of this and similar studies, Soucek (2007) stated, "Clearly, any attempt at water quality standard development, whether based on TDS, conductivity, sodium, or sulfate, should incorporate the fact that the water quality parameters like hardness and chloride strongly regulate the toxicity of high TDS solutions." Therefore, any attempts to use conductivity to evaluate the toxicity of specific water chemistries related to elevated conductivity must be interpreted carefully to ensure that the potentially confounding factors of hardness and chloride have been accounted for.

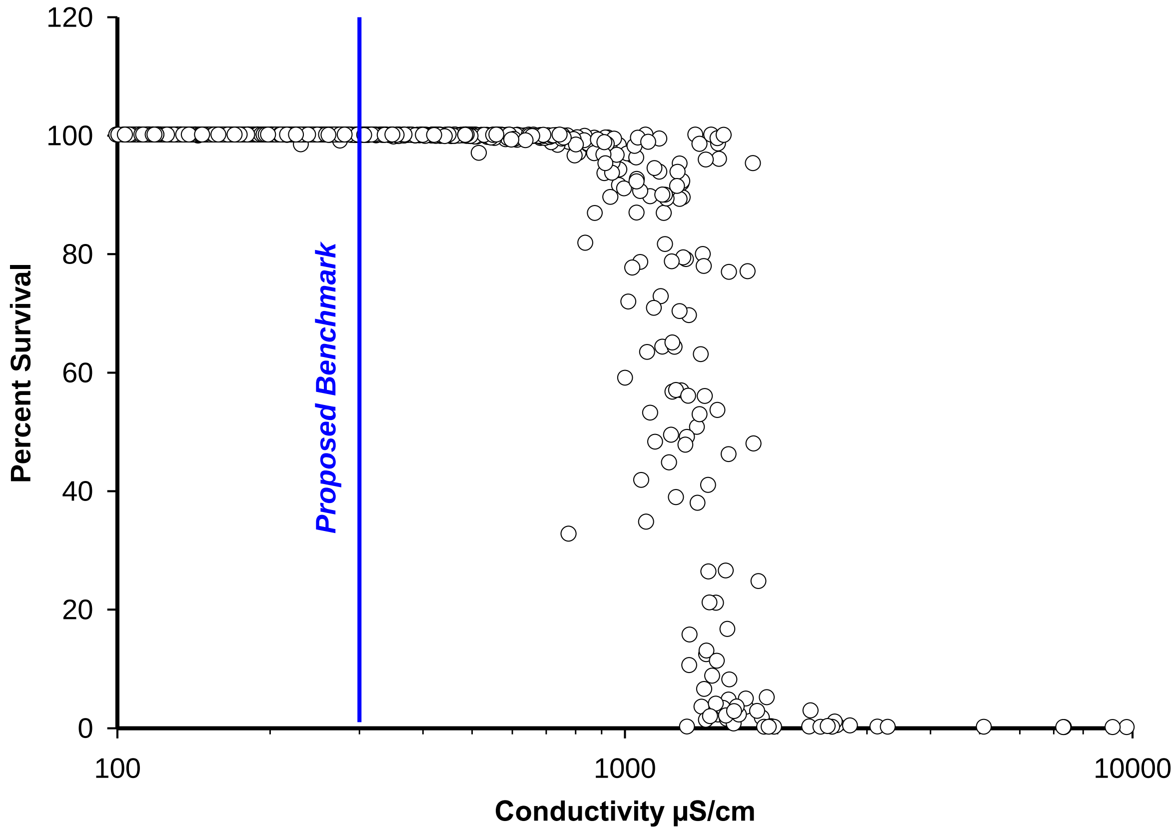

Figure 1: Predicted acute toxicity of natural waters to *Ceriodaphnia dubia* according to the STR model (Mount et al. 1997).

3.3.3.2 Laboratory tests of mine discharges

To evaluate the potential toxicity of coal mine effluents, EPA (2010) cited the Kennedy et al. (2004) study, in which *Isonychia bicolor* mayflies were exposed to simulated coal mine effluent in 7-day tests. The study reported lowest observed effect concentrations (LOECs) of 1,562 µS/cm, 966 µS/cm, and 987 µS/cm for mayfly survival at 20°C. These values bracketed the 95th percentile extirpation concentration (XC_{95}) of 1,177 µS/cm, calculated in EPA (2010) for the genus *Isonychia*. However, EPA (2010) failed to include all of the information from Kennedy et al. (2004). In the bioassay test conducted at 12°C on *I. bicolor*, the LOEC for survival was 4,973 µS/cm, which is substantially higher than the calculated XC_{95}. This difference in toxicity can possibly be attributed to the temperature the organisms were accustomed to in the natural environment prior to use in the test, since the organisms used in both the 20°C and 12°C tests were from an over-wintering cohort. Given that the proposed conductivity benchmark is intended to be applied for all times of the year, including tests from all temperatures is important. Additionally, EPA (2010) implied that the *Isonychia* tests were conducted on coal mine discharge waters, when they were actually conducted on simulated effluent.

Effects on *C. dubia* survival and reproduction in 7-day tests were also determined in the Kennedy et al. (2004) study on coal mine effluent and simulated effluent. In the effluent tests, no significant effects on survival were observed at conductivity levels up to 4,730 µS/cm and no effects on reproduction were observed at conductivity levels up to 3,254 µS/cm (Kennedy et al. 2004). In simulated effluent, no effects on survival occurred up to 4,530 µS/cm and no effects on reproduction occurred at up to 3,730 µS/cm. Similarly, chronic effects in simulated effluent tests for which toxicity was most likely attributable to sodium or sulfate (Kennedy et al. 2005) were observed at 3,200 µS/cm in very hard waters (792 mg/L), but as low as 2,000 µS/cm in soft waters (88 mg/L). Kennedy et al. (2003) also conducted 7-day tests on coal processing effluent using *C. dubia*. In duplicate tests, the chronic LOECs for *C. dubia* mortality were 4,730 µS/cm and 6,040 µS/cm, and the LOECs for reproduction were 2,910 µS/cm and 3,710 µS/cm.

Echols et al. (2009) conducted tests on the effects of coal processing effluent on *Isonychia* sp. and *C. dubia*. EPA (2010) provided the LOEC values for *Isonychia* survival, which ranged from 1,508 µS/cm to 4,101 µS/cm; however, *C. dubia* tests conducted in the same study resulted in broadly overlapping LOEC values ranging from 2,132 µS/cm to 4,240 µS/cm (Echols et al. 2009). These data indicate that *Isonychia* and *C. dubia* had similar sensitivities to high conductivity waters, which is contrary to Kennedy et al. (2003, 2004). The variability seen in the tests is not entirely unexpected; salinity tolerance for most species appears to be variable and may fluctuate, depending on abiotic factors such as temperature, pH, and dissolved oxygen (Pillard et al. 1999). Since not all of these factors were reported in Echols et al. (2009), we cannot confirm whether they might have been responsible for some of the variability between tests.

In Echols et al. (2009), three mayfly bioassays and five *C. dubia* bioassays were run on the coal processing effluent; it is unclear if the studies were concurrent. LOEC values from the three mayfly bioassays ranged from 1,508 µS/cm to 4,101 µS/cm. The final mayfly bioassay exhibited the lowest LOEC, and EPA (2010) attributed this to the dominance of sodium in that test. However, Echols et al. (2009) speculated that the lower LOEC may have been because the mayflies used in that test were from a summer cohort and may have been more sensitive. EPA (2010) also stated that Echols et al. (2009) attributed toxicity to mayflies in all of the studies to salinity because the effluent contained no detectable toxic metals, except for selenium (8.5 µg/L). However, the only effluent that was chemically analyzed (and could therefore support or refute such a claim) was the effluent from the first mayfly bioassay test, which had resulted in the highest LOEC (4,101 µS/cm), and in which there was a poor correlation between survival and conductivity for the first seven days, with improved correlations by day 14 (Echols et al. 2009). The effluent from the two other mayfly studies was not analyzed; therefore, it is unknown if there were any metals or unknown toxicants in the effluent.

Woodward et al. (1985) studied the effects of spent shale leachate on the mayfly *Hexagenia bilineata*. Although this is a different water type than the other studies presented here, it is a multi-ion water with high conductivity: measured ions in this water included B, Ca, K, Li, Mg, Mo, Na, Sr, F, Cl, NO_3, and SO_4. Woodward et al. (1985) reported that magnesium and sulfate represented 81% of the total ions in the leachate. In this study, mayflies were exposed to various dilutions of concentrated leachate and organism survival and growth, i.e., length, were determined at 15 and 30 days. The LOEC values for survival for 15 and 30 days were 2,950 µS/cm and 1,800 µS/cm, respectively (Woodward et al. 1985). There were no significant effects on length in either of the tests. These effect concentrations are comparable to the other mayfly studies previously described.

Some additional conclusions made by the authors in these studies were not reported in EPA (2010). For example, Kennedy et al. (2004) concluded that conductivity levels up to 900 µS/cm appeared to be safe for sensitive benthic invertebrates, based on survival of *Isonychia* in their studies, instream mayfly distributions, and endpoints from previous research. It was further concluded that reductions in the mayfly populations would likely occur between 1,500 to 2,000 µS/cm. Echols et al. (2009) determined that impairment occurred around 1,400 mg/L of total dissolved solids (TDS), which is approximately equivalent to a conductivity level of 2,333 µS/cm.

Therefore, the high levels of variability seen in the results of all of the studies described above further suggest that toxicity cannot easily be predicted from conductivity or TDS concentrations alone. Rather, the toxicity of major ions related to salinity can vary widely as a result of the concentrations and combinations of the ions present (Chapman et al. 2000), as well as other factors not easily compared between tests (Pillard et al. 1999). Short-term chronic toxicity as a function of conductivity in these tests suggested that although mayflies may indeed be somewhat more sensitive than *C. dubia*, effect levels sometimes overlapped broadly and were highly variable between tests. Therefore, the available toxicity test data do not necessarily agree with EPA's assertion that toxicity is observed at conductivities that are similar to the XC_{95} values for sensitive taxa.

3.3.3.3 Laboratory tests of ambient waters

Merricks et al. (2007) conducted acute bioassay tests on ambient water from streams below valley fills in West Virginia using *C. dubia*. EPA (2010) stated that LC_{50} values were established for *C. dubia* for some but not all of the waters from Lavender Fork. Three of the eight Lavender Fork sites did have LC_{50} values ranging from 1,763 µS/cm to 2,184 µS/cm; however, EPA (2010) did not mention that 19 other sites were tested, with conductivity levels ranging from 923 µS/cm to 2,720 µS/cm, and only one of the 19 tests resulted in significant effects on *C. dubia*. EPA (2010) concluded that these tests had low relevance to the conductivity benchmark and would underestimate toxicity in the field, due to the test species, duration, and endpoint. However, the data in Merricks et al. (2007) demonstrate that toxicity in waters below valley fills, whether acute or chronic, is highly variable and cannot be easily predicted based on conductivity alone. This supports our premise that generic measures of ionic concentration, such as TDS or conductivity, are inadequate for assessing the true potential toxicity of major ions present in waterbodies (Mount et al. 1997, Pillard et al. 1999, Goodfellow et al. 2000).

In addition, the USGS Columbia Environmental Research Center (CERC) is currently conducting U.S. EPA Project No. DW-14-922510010 "Toxicity of Total Dissolved Solids to Appalachian Aquatic Invertebrates" (Kemble 2010). These studies are using reconstituted waters that simulate water chemistry from several locations in West Virginia to more directly evaluate the sensitivity of mayflies and other aquatic invertebrates relative to conductivity. In the project summary for fourth quarter 2009 (October 1 to December 31, 2009), USGS reported the results of tests conducted using three reconstituted waters (Board Tree, Upper Dempsey, Winding Shoals) and a control water on *H. azteca*, *Lampsilis siliquoidea* (28-day tests), *C. dubia* (7-day test), and mayflies (14-day test using *Hexagenia* spp., likely a mixture of *H. rigida* and *H. limbata*). All species had acceptable control survival except for mayflies; the researchers reported that mayflies "do not do well" after 14 days in exposures without a sediment substrate. However, on day 8 of the mayfly bioassay tests, control survival was satisfactory at 88% survival; therefore, day 8 data were used for comparison to controls. There were no observed effects on *Hexagenia* or *C. dubia* survival in the Board Tree tests, with conductivity levels ranging from 579 µS/cm to 2,386 µS/cm. In the Upper Dempsey test on *Hexagenia*, there were effects on survival at 961 µS/cm, and for *C. dubia*, effects were seen at 1,817 µS/cm. In the Winding Shoals test, effects on *Hexagenia* survival were observed at 798 µS/cm, but no effects on *C. dubia* survival were seen in conductivities up to 1,828 µS/cm. These preliminary data indicate that in some ionic mixtures, mayflies appear to be slightly more sensitive than *C. dubia*, but in others they exhibit similar sensitivities.

3.3.4 Conclusions – Plausibility of Physiological Mechanisms for Extirpation

In summary, there are insufficient data from the physiological and toxicology literature to rigorously support EPA's conclusion that "conductivities in the region of concern reach levels that are sufficient to cause effects on stream communities" (EPA 2010, p. 52). First, although elevated salinity can clearly induce adverse effects on aquatic invertebrates, the taxonomic patterns of sensitivity are not yet clearly defined. Although laboratory toxicity data exposing mayflies to actual or simulated mining effluents suggest they may be somewhat more sensitive than the most sensitive surrogate test species, *C. dubia*, effect concentrations are highly variable and, in some studies, overlap between species. Toxicity to ions associated with salinity (e.g., sulfate) also varies strongly as a function of specific ion composition and can be mitigated under conditions of elevated hardness. Additional study is needed to confirm the relative sensitivity of macroinvertebrate communities to elevated salinity and the extent to which other water quality variables and major ion composition will influence the consistency of these results. Until such relevant studies are conducted, it is premature to suggest that a quantitative conductivity benchmark is an accurate and direct reflection of ions related to salinity, even if restricting the benchmark to waters in which conductivity is dominated by sulfate and bicarbonate.

It is noteworthy that three other states, Illinois Indiana and Iowa, have all rejected conductivity or TDS-based aquatic life standards in lieu of numeric standards for sulfate and chloride that also depend on water hardness. For Iowa, the current final rules (http://www.iowadnr.gov/water/standards/chloride.html) indicate that the existing scientific data support the importance of individual ions over composite variables such as TDS because "chloride and sulfate are better indicators than integral parameters such as TDS, conductivity, and salinity for water quality protection." (IDNR 2009). Similarly, the Illinois EPA proposed a numeric sulfate standard to replace TDS standards for the same technical reasons , and which was also ultimately approved by EPA (Norwest Co., 2010). Indiana proposed essentially the same sulfate and chloride criteria equations, which were also approved by EPA because "the TDS standard currently in place is inappropriate. By definition TDS is a measure of all dissolved solids, yet we know that the toxicity of TDS is exerted by its individual components" (EPA 2008). Therefore, the available scientific information does not support development of regulatory thresholds based on composite variables such as conductivity or TDS, but rather the development of individual numeric criteria for specific ions.

3.4 Confounding Factors Analysis

The confounding factors analysis in Appendix B of EPA (2010) uses a weight of evidence approach to evaluate whether environmental factors other than conductivity could substantially interfere with or otherwise bias the presumed relationships between conductivity and biological impairment in West Virginia streams. However, EPA's goal was not to eliminate confounding variables, nor was it an attempt to independently test the hypothesis that conductivity was the best predictor of biological impairment. As stated on page 69 of EPA (2010), "This assessment of confounding takes the result of the causal assessment as a given (*emphasis added*) and attempts to determine whether any of the known potential confounders interfere with estimating the effects of conductivity to a significant degree." Furthermore, the confounding factors analysis was based entirely on patterns related to mayfly abundance "(b)ecause the sensitive genera are primarily Ephemeroptera and the endpoint effect is extirpation of 5% of genera…" (EPA 2010, p. 69).

We agree that it is an important and relevant exercise to evaluate the potential influence of confounding factors on the primary factor(s) presumed to be the strongest predictor(s) of biological response and also causally related to the response. As discussed in Section 3.3, we do not agree that sufficient evidence exists to determine that conductivity is necessarily causally related to extirpation of "sensitive" species at the concentration represented by the proposed benchmark. Nor do we necessarily agree that conductivity is the single or best predictor of patterns in macroinvertebrate community structure related to MTM/VF activities, especially as manifested by mayfly abundances (Section 4.0). Therefore, we also do not agree that a confounding factor analysis should take it *as a given* that these are the only or primary relationships that require evaluation. Rather, we contend that a confounding factors analysis should also include rigorous and independent tests of the primary hypothesis, and determine whether conductivity is indeed the best predictor of biological impairment that is causally related in such a way as to justify the proposed benchmark value. Indeed, the causal assessment in Appendix A does not present or evaluate potential causal factors other than conductivity, so the overall analysis presented by EPA (2010) does not appear to thoroughly test alternative hypotheses to that of conductivity.

The comments below present observations on a selection of confounding factors addressed by EPA in Appendix B that we suggest represent factors that may correlate with and potentially confound conductivity relationships. We also suggest that these may represent factors that could be as or more important to benthic macroinvertebrate community structure than conductivity and, hence, require a more formal analysis to determine whether they represent viable alternatives to the hypothesis that conductivity is the primary factor responsible for impairment.

3.4.1 Evidence Rejecting Habitat Differences as Possible Cause of Extirpation

The assertion in EPA (2010) that habitat presented little potential for confounding in their derivation of the conductivity benchmark needs considerable additional scrutiny. There are three clear problems with this assertion.

First, Rapid Bioassessment Protocol (RBP) habitat scores may not be the most rigorous measure of habitat quality. Rates of mayfly presence were nearly identical between poor quality and high quality habitat at low conductivity levels in the contingency table (Table B-8 of EPA 2010), indicating that RBP habitat scores are not the best predictor of habitat quality for mayflies. This may be because RBP habitat scores are more directed toward identification of fish habitat and they are influenced by a significant level of subjectivity, even if the method itself results in some level of quantification.

Second, the RBP habitat scores were correlated with conductivity and the biological response , i.e., the HC_{05} (Section B.4.1. of EPA 2010). This in itself should suggest that habitat may be a significant confounding factor. Because RBP habitat scores do not appear to tell the whole story, a more detailed analysis of habitat quality and its relationship to the benthic macroinvertebrate community needs to be conducted before EPA can conclusively state that "low RBP was judged to have little effect on the derivation of the 5^{th} percentile hazardous concentration (HC_{05}) for conductivity" (Section B.4.1. of EPA 2010).

Third, as noted below in Section 3.4.2, the analysis of the potential confounding factors in EPA (2010) focused almost exclusively on the response of Ephemeroptera to conductivity levels, to the exclusion of the rest of the benthic macroinvertebrate community. The Ephemeroptera are represented by 25 genera in the database, which is only 16.5% of the total number of genera. Furthermore, while some genera of Ephemeroptera do appear to be sensitive, they do not appear to be the most sensitive genera. Based on the XC_{95} calculations, *Remenus* (a stonefly) and *Lepidostoma* (a caddisfly) are ranked more sensitive than the most sensitive mayfly genus, *Cinygmula*. Because mayflies are not the most sensitive organisms in the database, the other ordinal taxa should be investigated to determine their response to conductivity across gradients of habitat quality. Furthermore, because of the variety of stressor-response profiles exhibited by all of the genera in the database, it would be more informative and conclusive to analyze the response of a representative subset of genera representing multiple stressor-response profiles, not just the mayflies, to habitat variables and conductivity.

Even if the RBP habitat scores can appropriately be eliminated as a potential confounding stressor, EPA has not sufficiently demonstrated that habitat (by RBP scores or by a more detailed analysis of habitat quality) can be eliminated as a potential confounding factor to the rest of the benthic macroinvertebrate community.

3.4.2 Rejection of Confounding Factors Based Almost Exclusively on Ephemeroptera (Mayflies)

In EPA (2010), an attempt was made to reject as many potential confounding stressors as possible. However, from a community ecology standpoint, lack of correlation between number or presence of Ephemeroptera genera and each potential confounding stressor was most often cited in Appendix B as being a strong reason to reject most potential stressors as confounding influences with conductivity. As noted above (Table 2), Ephemeroptera genera represented only 16.5% of the dataset, yet it does not appear that the other taxonomic orders or the entire invertebrate assemblage were similarly tested to determine their relationships to the potential confounding stressors. Because several taxonomic groups were present in the community, and because the Ephemeroptera genera differed in their responses to conductivity (Table 2), we suggest it is inappropriate to focus only on Ephemeroptera in elimination of potential confounding stressors.

Furthermore, it also does not appear that individual genera were examined to determine their stressor-response to the other potential confounding stressors to eliminate those stressors definitively. If such testing had been conducted for representative genera in the database, it may have been found that many of the genera (particularly those with optimum or increasing stressor-response patterns) were in fact responding to some of the other potentially confounding stressors.

Relationships between all potential stressors and Ephemeroptera were generally cited as reasons to reject the stressors as potential confounders in the analysis that ultimately relates to the entire aquatic benthic community. There is a clear need to include similar analyses from the other invertebrate orders and the entire invertebrate community to conclusively reject the stressors as potential confounding stressors.

3.4.3 Natural Rarity as a Reason for Low Capture Probability

The original, full WVBDEP's Watershed Assessment Branch Data Base (WABbase) used by EPA included 559 taxa, of which 498 were identified to genus; the others were identified to family or were “slashed” taxa (e.g., *Leucrocuta/Nixe*). Because EPA (2010) restricted the database with several filters (e.g., particular ecoregions, pH, months, years, watershed size, sulfate dominance instead of chloride, etc.), 328 aquatic macroinvertebrate genera from West Virginia were excluded from analysis.

EPA (2010) attempted to control for the effect of rare taxa by including only those taxa that had been collected in at least one reference site and at least 30 general sites; therefore, 18 additional genera were excluded from analysis because they were never found at a reference site (Table 3, p. 26). There were 2,145 samples represented in the total dataset, based on Table 2 on page 26. According to Figure 2 (p. 28), there were 97 reference samples from 70 individual sites used in EPA (2010), although page 7 of EPA (2010) said there were 70 reference sites. Therefore, if a genus had a collection probability of at least 1.0% in the reference sites and at least 1.4% in the general sites, it was considered to be common enough to include in the SSD. The number of occurrences of each genus in the reference samples was provided in Appendix C of EPA (2010).

Even though the number of taxa included would necessarily be constrained (Table 3), it may have been more appropriate for EPA to have controlled for the effects of rare taxa by including in their SSD only those genera that had a high capture probability in the reference sites. Such an approach would be analogous to a laboratory study in which mortality in the control is a major determinant of the validity of a study. In discussing criteria development, the 1985 Guidelines stated that “data should usually be rejected if they are from . . . tests in which too many organisms in the control treatment died or showed signs of stress of disease. . . .” Many laboratory studies are rejected for inclusion in a criterion calculation because mortality in the control exceeded a certain percentage. Although “too many organisms” was not specifically defined in the 1985 Guidelines, many criteria we are familiar with used cutoffs near 20% mortality (or 80% survival) in the controls. EPA (2010) considered a 1% collection probability in reference sites to be acceptable, but a 1% survival rate in a laboratory test would clearly not be acceptable.

Table 3:
Number of genera available for SSD calculation based on capture probability in reference samples.

Capture Probability in the Reference Samples	Number of Genera Included
All	151
>1%	138
>5%	100
>10%	75
>20%	49
>50%	14

The most sensitive taxon in the database, the stonefly *Remenus*, had a calculated XC_{95} of 101 µS/cm. This taxon was found in three of the reference samples (3%) and 35 of the general samples (1.6%). Even though there were 38 data points (excluding non-detects) used to derive a stressor-response relationship, it is clear that the genus is rare even in reference streams where conductivity levels are low. It cannot be clearly demonstrated that the relationship observed in *Remenus* of a decreasing capture probability with increasing conductivity is accurate when there is only a 3% probability of capture in reference streams. (Using the laboratory control analogy, this is similar to only 3% survival in the control—a result that would strongly invalidate a study.) Furthermore, *Remenus* is so rare that nothing is known about the biology of the nymphs, except limited information on timing of emergence from collection records (Stewart and Stark 2002). There is not sufficient information for EPA (2010) to assume that high conductivity levels are responsible for the rarity of *Remenus* when it is naturally rare in the general population.

A total of 72 taxa (48%) had a higher capture probability in the general sites than in the reference sites (Figure 2). Of those taxa, the difference between the two probabilities was less than a full percentage point in many taxa; however, 27 taxa had a capture probability at least 5% higher in the general sites than in the reference sites. The largest difference was in the chironomid genus complex *Cricotopus/Orthocladius*, which was not used in the calculations. The second largest difference was in *Stenelmis*, which had a capture probability of 51% in the general sites, but only a 13% capture probability in the reference sites (difference = 38%).

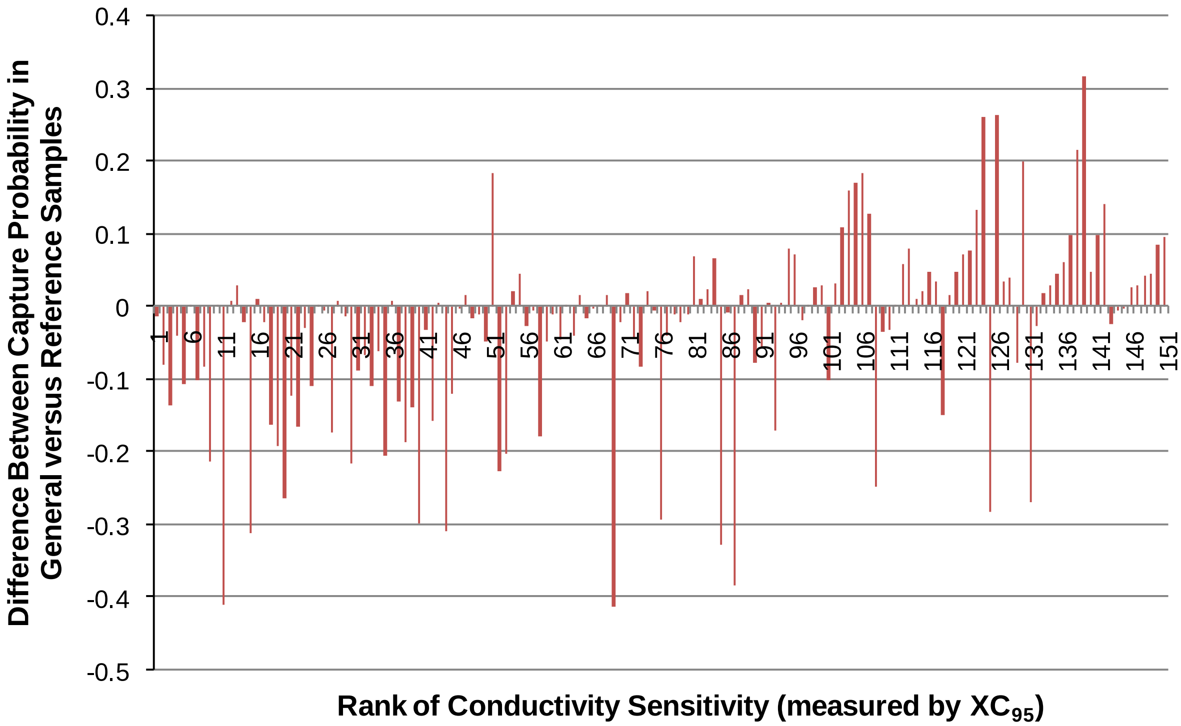

Figure 2: Difference between capture probability in general versus reference samples, ranked by sensitivity to conductivity. Positive values indicate that the genus had a higher capture probability in general samples than in reference samples.

A plausible argument against excluding rare taxa from the SSD would be that the taxon is rare because of the stressor. However, this argument would not be valid if the taxon is naturally rare, a phenomenon that could be analyzed using its capture probability in reference sites. EPA (2010) did not sufficiently demonstrate that the rare taxa were rare due to conductivity or any other water quality effect, and not from general rarity itself.

3.5 Ecological Relevance of Presumed Impairment as a Function of Conductivity

The ultimate protection goal of EPA's proposed conductivity benchmark is to determine a conductivity level that, if not exceeded, would prevent extirpation of 95% of the aquatic macroinvertebrate genera. This is similar to the protection goals of numeric criteria for protection of aquatic life and their uses (1985 Guidelines). However, it is worth evaluating the relevance of this protection goal when the criterion or benchmark is derived from a very large number of genera, as is the case with the proposed conductivity benchmark. For even the most data rich numeric criteria, far less than 100—and often less than 20—genera are used to derive the acute and chronic criteria. However, with the conductivity benchmark, 151 genera are used and yet only a few mayfly taxa are truly considered indicative of "sensitivity" to conductivity that requires protection. In addition, even if one accepts the role of conductivity in being directly correlated with and responsible for loss of taxa at elevated conductivity, is this 95% protection level ecologically relevant, i.e., do communities in the presence of elevated conductivity lose important ecological functionality?

To address these questions, we evaluated trends in macroinvertebrate community structure and function relative to conductivity from the data presented in EPA (2010). Numerous functional feeding groups (FFGs) were represented in the dataset used for EPA (2010), including filter-collectors, gather-collectors, omnivores, predators, scrapers, shredders, and one piercer. Gather-collectors and predators were the most abundant FFGs. Filter-collectors, gather-collectors, predators, and shredders were each represented by genera with each of the identified stressor-response profiles (Table 4). Scrapers were represented by genera in each of the stressor-response profiles except for the profile that increased with respect to increasing conductivity values.

Table 4: Number of genera in particular functional feeding groups in identified stressor-response profiles. Piercers and omnivores not included due to low numbers of taxa.

Functional Feeding Group	All	Decreasing	Optimum	No Response/Bimodal	Increasing
FILTER-COLLECTORS					
Number of genera	15	4 (27%)	5 (33%)	2 (13%)	4 (27%)
GATHER-COLLECTORS					
Number of genera	57	27 (47%)	12 (21%)	6 (11%)	12 (21%)
PREDATORS					
Number of genera	46	22 (48%)	11 (24%)	4 (9%)	9 (20%)
SCRAPERS					
Number of genera	17	6 (35%)	10 (59%)	1 (6%)	0
SHREDDERS					
Number of genera	12	7 (58%)	2 (17%)	1 (8%)	2 (12%)

We conducted an analysis to determine what changes would occur in the trophic group balance within the regional taxa pool at various conductivity levels, based on the XC_{95} value for each genus. For example, if the conductivity value was 400 µS/cm, it was assumed for this analysis that all genera with an XC_{95} value less than 400 µS/cm would be extirpated. The trophic balance of the remaining taxa available from the regional taxa pool was then analyzed (Figure 3).

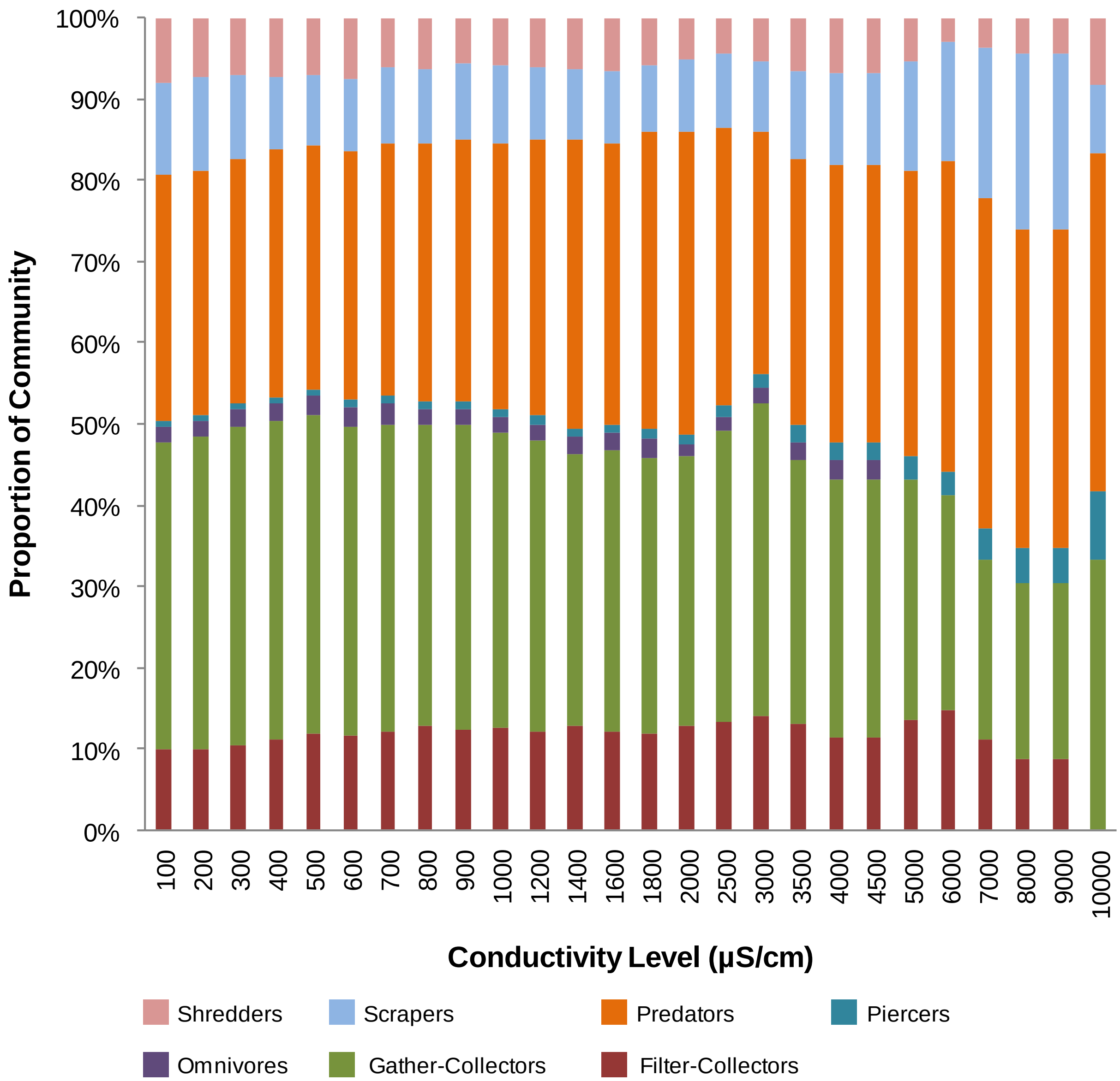

Figure 3: Proportion of generic richness by functional feeding group within the regional taxa pool at conductivity levels. All genera with an XC_{95} less than the conductivity level are considered to be unavailable. Note that the x-axis is not evenly divided.

There are few observed changes in the proportional abundance of FFGs within the regional pool of taxa at conductivity levels below approximately 2,500 µS/cm to 5,000 µS/cm. Excluding the omnivores and piercers, which were poorly represented in the first place, the first major FFG to undergo extirpation of all its member genera is the filter-collectors, when conductivity values exceeded 10,000 µS/cm. This indicates that the functional aspect of the stream community may not change due to conductivity levels above 300 µS/cm, since genera from all FFGs remain available in the regional taxa pool.

4.0 Independent Analysis of Factors Shaping Macroinvertebrate Communities

The EPA (2010) conductivity benchmark analysis appears to use the results of a single study (Pond et al. 2008) to presuppose as a "given" that conductivity is one of the strongest factors associated with the structure of macroinvertebrate communities in Central Appalachian streams, while disregarding many other factors that may influence community composition. The WVDEP database (WABbase) used by EPA provides an opportunity to examine other possible factors that may shape macroinvertebrate community composition. Therefore, we have conducted a preliminary independent analysis based on a data mining approach that considers all of the available information and strives to elucidate key water quality and physical parameters that are most strongly associated with biotic responses.

We used the EPA dataset that was originally extracted from the WABbase (http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=496202) as provided by EPA (http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=220171). This data set included results for 3,286 sampling events representing 3,121 unique Station ID codes. The dataset contains a variety of variables that present site-specific information regarding regional landscape, water quality, and aquatic habitat conditions as well as macroinvertebrate community composition. Following EPA's rationale for excluding samples (EPA 2010, Section 2), the dataset used in our analysis contained 2,152 sampling events representing 2,073 unique Station ID codes. Our attempt to follow EPA's exclusion process resulted in an additional seven sites being included in the data subset. This is different from the EPA's final data which contained 2,145 sample events, but when summary statistics for measured water quality parameters as presented in EPA 2010 Table 1 are compared, the datasets appear to be nearly identical. Notably, of the 2,152 sampling events selected from the EPA data set, approximately 43% of the sampling events are missing ion or metal chemistry results, and this does not include selenium or dissolved manganese which were analyzed infrequently.

Our data-mining analysis of the subset we generated from EPA's data-exclusion rationale (2,152 events) was based on an integrated approach to identify factors that best describe the observed variability between and among sites, and strongly correlate with each other, rather than trying to establish causal relationships. In the absence of a rigorous study design conducted under controlled experimental conditions, it is more important to identify data relationships rather than trying to establish cause-effect relationships.

Our integrated data analysis approach relies on a series of statistical analyses that reduce the total number of parameters to a more ecologically meaningful subset of variables, with respect to the available data. The original dataset was initially subdivided into independent stressor and dependent response variables. Independent stressor variables in a stream ecosystem include chemical and physical habitat variables such as metal and ion concentrations in the water column and the percent substrate composition (Paulson et al. 2001). Dependent response variables were selected to represent the biological component of the stream such as macroinvertebrate density or taxa richness. The independent variables generally represent a mix of both quantitative (e.g., major ion or metal concentrations) and qualitative variables (e.g., embeddedness), as well as composite variables (e.g. RBP score, conductivity). Thus understanding the general categories of each variable also helped reduce the overall list of variables.

The integrated analysis follows a series of statistical procedures (Paulson et al. 2001), as presented below, to identify key variables that can be used to characterize water quality, aquatic habitat, and macroinvertebrate communities. Due to the time constraints and data acquisition issues, our preliminary analyses have progressed through step 2b, although it is our intent to complete all of the analytical steps in this method.

1. Apply basic statistics
    a. Descriptive statistics and data plots
    b. Normalize data as needed to meet statistical assumptions
    c. Compile correlation matrices
2. Identify key stressor and response variables
    a. Principle Components Analysis (PCA)
    b. All possible regressions (APR)
    c. Chi-square Automatic Interaction Detection (CHAID)
3. Rank variables according to relative influence
    a. Develop matrix of key independent stressor variables and relationships found in Step 2.
    b. Repeat Steps 2 and 3 until the two most influential independent stressor variables are identified for each dependent response variable.
4. Fit equation to describe interactions between stressor and response variables
    a. Use three-dimensional modeling program to identify non-linear relationships
5. Repeat cycle if more variables are needed
    a. Use the residuals from multivariate regression as the response variables and repeat Steps 2 – 4.

Using the data subset, basic statistical procedures (e.g., Spearman rank correlation, scatter and box plots) were used to evaluate the characteristics of both stressor and response variables, as well as relationships between the two variable types. All variables were evaluated for approximation of a normal distribution using Shapiro-Wilkes normality test and Q-Q probability plots. When appropriate, variables were transformed and reevaluated for fit with an expected normal distribution. A logarithm base10 transformation (log) was used for water quality variables, and macroinvertebrate density, while the arcsine-square root transformation was used for variables reported as percentages (e.g., percent fines and percent

Ephemeroptera). The water quality variables—temperature and pH, as well as the physical habitat and macroinvertebrate variables such as embeddedness and genera-based metrics—did not require transformation. Two macroinvertebrate metrics (Trichoptera taxa and percent Trichoptera) were calculated based on subtraction of reported Ephemeroptera and Plecoptera metrics from summary EPT results provided in the data subset. Using the basic summary statistics, as well as professional judgment, the entire list of variables was initially reduced to a smaller subset of variables that we believed to be the most ecologically relevant when evaluating factors that explain the variability observed between sites, in terms of macroinvertebrate communities in Central Appalachian streams (Table 5).

Table 5: List of independent stressor and dependent response variables used in the integrated analysis.

Independent Stressor Variables		Dependent Response Variables
Water Quality	**Physical Habitat**	**Macroinvertebrate**
Temperature	Bank stabilization	Clinger taxa, genera
Dissolved oxygen	Bank vegetation	Ephemeroptera, genera
Alkalinity	Undisturbed vegetation	EPT, genera
pH	Channel alteration	HBI, genera
Chloride	Channel flow	Intolerant taxa, genera
Sulfate	Riffle sinuosity	Plecoptera taxa, genera
Total aluminum	Embeddedness	Trichoptera taxa, genera*
Total calcium	Sediment deposition	Total taxa, genera
Total iron	Epifaunal substrate	Density
Total magnesium	Velocity of pool	Percent Chironomidae
Total manganese	Percent fines	Percent Ephemeroptera
Total suspended solids	Percent sand	Percent Ephemeroptera minus Baetidae
Total phosphorus	Percent silt	Percent EPT
Nitrate – Nitrite nitrogen		Percent EPT minus Cheumatopsyche
Fecal coliforms		Percent EPT minus Cheumatopsyche and Baetidae
		Percent Hydropsyche
		Percent Orthocladiinae
		Percent Plecoptera
		Percent Trichoptera*
		Percent Simuliidae
		Percent dominant 5 taxa, genera

* Calculated metric.

It is important to note that composite type variables are often not very useful when evaluating biological responses to environmental stressors. For example, the RBP score for aquatic habitat evaluation may appear to strongly correlate with select biotic responses, yet this index provides little insight into the environmental characteristics that may be influencing biotic communities, because it is comprised of many metrics. To the extent possible, we have excluded such composite independent stressor variables in our preliminary data analyses, including conductivity and hardness, because they provide little information above and beyond the individual variables when trying to isolate water quality factors that may be most strongly associated with a biotic response.

4.1 Principal Component Analysis—Water Quality

PCA was used to identify variables that best explain the variability observed between sites and how those variables relate to one another, as well as whether one variable could be used as a surrogate for other variables within each grouping (water quality, physical habitat, macroinvertebrate). When such variables are replaced with a surrogate that explains the same amount of variation, the power of the statistic to identify relationships is maximized (Paulson et al. 2001). An iterative process was used for the PCA analyses, such that all variables from each grouping were loaded into separate PCA models. This initially created three distinct groupings, two for stressor variables and one for response variables. The PCA extraction method was based on a correlation matrix with a varimax rotated solution, pairwise deletion of missing values, and extracted eigenvalues greater than 1.0. The rotated component matrix for each variable grouping was examined with variables exhibiting coefficients greater than 0.6 being considered a significant part of the component. If the component contained multiple significant variables, the Spearman rank correlation values for those variables were also evaluated. If variables were highly correlated (i.e., > 0.6 or < -0.6) with each other, the variable with the largest component coefficient (i.e., heavily weighted) was selected. Up to five components were examined with the most heavily weighted or unique variables (either positive or negative) being selected for inclusion in a subsequent PCA model (Table 6).

The goal of this type of evaluation is to understand how the water quality variables "move together" (i.e., were positively or negatively correlated with one another) and to select variables that may be a surrogate for other variables. For example, in the first component, the log transformed variables for total magnesium, sulfate, and total calcium weighted the most heavily. This weighting and movement (all positive) of the variables along the first component was to be expected, based on the chemical relationship between all of these variables and their Spearman Rank correlation values. In the second component, the log transformed variables for total iron, total aluminum, and manganese were weighted the most heavily with all variables showing positive movement with each other. In the third component, fecal coliforms, pH, and alkalinity revealed the strongest weighting coefficients. Temperature and dissolved oxygen were key variables in the fourth component, and moved in opposite directions as is to be expected, while the nutrients—total phosphorus and nitrate-nitrite—were the most heavily weighted variables in the fifth component. In combination, these variables with the greatest weighting best explained the variability observed between sites in the database.

Table 6: Rotated component matrix for selected water quality variables.

Variable	Component				
	1	2	3	4	5
temp	0.254	-0.092	0.444	**0.620**	-0.035
pH	0.329	-0.303	**0.695**	-0.243	0.056
log_do	-0.033	-0.044	0.064	**-0.882**	-0.045
log_alk	0.556	-0.185	**0.641**	0.065	0.004
log_fecal	0.074	0.299	**0.651**	0.145	-0.074
log_mg_tot	**0.934**	0.067	0.166	0.069	-0.008
log_sulfate	**0.914**	0.031	0.109	0.035	0.020
log_tp	-0.130	0.064	0.157	0.220	**0.745**
log_al_tot	-0.064	**0.815**	-0.036	-0.121	0.087
log_ca_tot	**0.861**	0.067	0.322	0.120	-0.055
log_chloride	0.468	-0.015	0.485	0.080	0.241
log_fe_tot	0.119	**0.856**	0.011	0.100	-0.068
log_mn_tot	0.450	**0.601**	-0.172	0.401	-0.038
log_no23	0.243	-0.071	-0.230	-0.185	**0.738**
log_tss	-0.133	0.548	0.148	-0.097	0.511

Bolded values denote which variables are considered the most heavily weighted part of the component.

The selected variables within the first five components accounted for a total of 72% of the variation observed among sample sites with respect to the water quality variables contained within the WVDEP/EPA dataset (Table 7). Metal parameters such as calcium, sulfate, and magnesium along with parameters that characterize overall ionic strength explained approximately 38% of the variation among sample sites with respect to water quality.

Table 7: Total variance explained by the initial PCA for water quality variables.

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.474	23.158	23.158
2	2.305	15.364	38.522
3	2.026	13.506	52.029
4	1.552	10.344	62.373
5	1.448	9.654	72.026

For our preliminary analysis, the following variables were selected to be surrogates for other less heavily weighted variables in each component and to be included in a subsequent PCA analysis:

1. total magnesium
2. total iron
3. pH
4. fecal coliforms
5. dissolved oxygen
6. total phosphorus
7. total suspended solids

We selected TSS even though it that did not initially meet our original selection criteria. Based on its relative moderate loading in two of the five components, as well as its relationship to geological and hydrological underpinnings within the watersheds, we believed this to be an important variable that may influence macroinvertebrate communities.

The seven selected water quality variables were subsequently loaded into a second PCA model, with the same evaluative process being performed on the rotated component matrix. The rotated component matrix converged in the first two components, with the first component comprised of the log transformed variables—total magnesium (0.800), pH (0.692) and fecal coliform (0.638). In the second component, the log transformed variables for total iron (0.698) and dissolved oxygen (0.661) weighted the most heavily, while total suspended solids (0.780) and total phosphorus (0.743) was considered part of the third component.

The final water quality variables that were selected to be included in the overall PCA model evaluating relationships between water quality, habitat and macroinvertebrate variables were:

1. total magnesium—also surrogate for Ca, SO_4, pH
2. fecal coliforms
3. total iron—also surrogate for Al and Mn
4. dissolved oxygen—also surrogate for temperature, and
5. total suspended solids—also surrogate for TP

4.2 Principal Component Analysis—Physical Habitat

The iterative PCA process described above was also performed using the independent physical habitat stressor variables. The initial PCA model using physical habitat characteristics extracted four components with the first component being comprised of sediment deposition (0.832), embeddedness (0.735), riffle sinuosity (0.675) and epifaunal substrate (0.643), all of which exemplify substrate quality in these watersheds. The second component included undisturbed vegetation (0.855), bank vegetation (0.833), and channel alteration (0.755) all of which are characteristic of riparian habitat. The third component included the arcsine-square root transformation for percent fines (0.950), percent sand (0.844), and percent silt (0.678) which characterize substrate composition. The fourth component only included channel flow which had a weighting coefficient of 0.810. These four components accounted for a total of the 66% of the variation observed among sample sites with respect to physical habitat conditions. The first component accounted for approximately 20% of the variation in physical habitat observed among sample sites.

From our initial analysis, we selected the following physical habitat variables to be included in a subsequent PCA analysis:

1. sediment deposition
2. undisturbed vegetation
3. percent fines, and
4. channel flow

The second physical habitat PCA extracted two components with sediment deposition (0.795) and percent fines (-0.769) weighted heavily and in opposite directions in the first component, even though they are not strongly correlated (Spearman, -0.376). Channel flow (0.909) weighted heavily in the second component. All three variables were selected to be included in the overall PCA model evaluating relationships between water quality, habitat and macroinvertebrate variables.

4.3 Principal Component Analysis—Macroinvertebrates

The initial macroinvertebrate PCA model resulted in four components being extracted (Table 8) with the first component comprised of the arcsine-square root transformations for the percent EPT variable and its derivatives, along with percent Ephemeroptera and its derivatives, percent Chironomidae, and the genera-based HBI. Even though the genera based HBI is not too informative from the standpoint of identifying key macroinvertebrate response variables, it is informative from a community health perspective. The second component loaded the genera-based metrics for total taxa, clinger taxa, EPT taxa and its derivatives Ephemeroptera and Trichoptera taxa, as well as intolerant taxa and arcsine-square root transformed percent dominant 5 taxa (negative loading). The third component was comprised of the arcsine-square root transformed percent Trichoptera, percent Hydropsyche, and the genera-based Trichoptera taxa, all of which characterize the caddisfly assemblage. The fourth component only included the log transformed macroinvertebrate density variable. All four components explained a total of 76% of the variation observed in sample sites with respect to the macroinvertebrate metrics contained in the WVDEP/EPA data set. The first component which was mainly comprised of EPT metrics and a Chironomidae metric accounted for approximately 31% of the variation among sample sites with respect to macroinvertebrates.

Table 8: Rotated component matrix for selected macroinvertebrate variables.

	Component			
Variable	1	**2**	3	4
G_Clinger Taxa	0.265	**0.836**	0.261	0.176
G_Ephemeroptera Taxa	0.424	**0.622**	-0.355	0.323
G_EPT	0.502	**0.819**	-0.085	0.048
G_HBI	**-0.722**	-0.373	0.053	0.359
G_IntolTaxa	0.459	**0.753**	-0.157	-0.215
G_PlecopteraTaxa	0.457	**0.603**	-0.316	-0.332
G_TricopteraTaxa	0.165	0.532	**0.649**	0.036
G_Tot_Taxa	-0.007	**0.922**	0.006	0.015
log_density	0.162	0.104	0.063	**0.603**
arcsin_pct_chiron	**-0.790**	0.010	-0.260	-0.122
arcsin_pct_ephem	**0.637**	0.202	-0.383	0.471
arcsin_pct_ephembaet	**0.891**	0.255	-0.025	-0.200
G_Clinger Taxa	0.265	**0.836**	0.261	0.176

Variable	Component			
	1	**2**	3	4
G_Ephemeroptera Taxa	0.424	**0.622**	-0.355	0.323
arcsin_pct_ept	**0.906**	0.145	0.226	0.070
arcsin_pct_eptchemat	**0.921**	0.217	-0.044	-0.040
arcsin_pct_eptchematbaet	**0.890**	0.255	-0.025	-0.200
arcsin_pct_hydrosych	0.062	-0.106	**0.929**	0.114
arcsin_pct_orthoclad	-0.584	-0.025	-0.263	-0.191
arcsin_pct_plecopt	0.536	0.338	-0.334	-0.524
arcsin_pct_tricopt	0.140	-0.027	**0.949**	0.093
arcsin_pct_simul	-0.139	-0.019	0.054	0.399
arcsin_pct_dom5	0.026	**-0.822**	-0.030	0.021
arcsin_pct_ept	**0.906**	0.145	0.226	0.070

Bolded values denote which variables are considered the most heavily weighted part of the component.

The percent EPT was strongly correlated with percent Chironomidae (Spearman, -0.686) and the EPT derivatives, therefore the percent EPT variable was selected from the first component. Similarly, the genera-based total taxa was strongly correlated with the percent dominant 5 taxa (Spearman, -0.789), clinger taxa (Spearman, 0.763), Ephemeroptera taxa (Spearman, 0.625), EPT taxa (Spearman, 0.724), and Intolerant taxa (Spearman, 0.662); therefore the total taxa metric was selected from the second component. The third component was comprised of caddisfly metrics, thus the most heavily weighted variable of percent Trichoptera was selected from this component. From our initial macroinvertebrate PCA, we selected the following variables to be included in a subsequent PCA analysis:

1. percent EPT
2. genera-based total taxa
3. percent Trichoptera, and
4. density

The second macroinvertebrate PCA extracted two components with the arcsine-square root transformed variables of percent Trichoptera (0.801) and percent EPT (0.785) loading in the first component and the genera based total taxa (0.940) being heavily weighted in the second component. These two components explained approximately 64% of the variation observed among sample sites with respect to macroinvertebrate metrics. The percent EPT variable was selected from the first component due to its inclusion of both mayflies and stoneflies, and total taxa was also selected for inclusion in the overall PCA model evaluating relationships between water quality, habitat and macroinvertebrate variables.

4.4 Principal Component Analysis—Overall

As a result of the individual PCA's above, a total of 10 variables were selected for inclusion in the overall PCA to evaluate the relative importance of key water quality (5 variables), physical habitat (3 variables) and macroinvertebrates (2 variables) in characterizing sample sites with respect to the available data. The PCA extracted four components (Table 9) with the first component weighting the log transformed total magnesium with total taxa, and the second component weighted sediment deposition and arcsine-square root transformed percent fines. The log transformed total suspended solids and total iron were strongly weighted in the third component. Channel flow and log transformed dissolved oxygen were weighted heavily in the fourth component. These four components explained approximately 55% of the variation observed among sampling sites with respect to the available dataset.

Table 9: Rotated component matrix for the overall PCA including water quality, physical habitat and macroinvertebrate variables.

Variable	Component			
	1	2	3	4
log_mg_tot	**0.797**	-0.123	-0.063	-0.048
log_fecal	0.497	-0.110	0.264	-0.057
log_fe_tot	0.183	-0.267	**0.760**	0.113
log_do	-0.077	0.090	-0.181	**0.752**
log_tss	-0.059	0.096	**0.820**	-0.077
Sed_Dep	-0.074	**0.781**	-0.053	0.072
arcsin_pct_fine	0.057	**-0.817**	0.037	-0.040
Chan_Flow	0.049	0.038	0.188	**0.803**
arcsin_pct_ept	-0.492	0.419	-0.080	0.083
G_Tot_Taxa	**-0.726**	-0.078	0.025	-0.094

Bolded values denote which variables are considered the most heavily weighted part of the component.

The first component in the overall PCA indicates that total macroinvertebrate taxa is moving in the opposite direction of major ions such as magnesium, indicating a strong relationship between the response of the macroinvertebrate community and ionic chemistry. In the initial water quality PCA, total magnesium was selected as a surrogate for sulfate and calcium which may also be important factors to consider regarding biological response. The second component indicates that substrate characteristics also are an important factor when trying to explain the variation observed among sample sites in Central Appalachian streams. Lastly, total suspended solids, total iron, channel flow, and dissolved oxygen also appear to be important factors to consider when evaluating these stream site conditions. Notably, the percent EPT metric did not weight heavily in any of the components, although its coefficients for both the first and second component indicate this metric may be weakly related ionic chemistry and substrate conditions.

The key variables identified in the PCA analyses were retained and placed into a matrix for further evaluation with results from the All Possible Regressions and Chi-square Automatic Interaction Detection. This matrix will be used to refine the key variables for inclusion in a three dimensional model to evaluate the nonlinear relationships between water quality, physical habitat, and macroinvertebrate metrics.

4.5 All Possible Regressions

All Possible Regressions is another iterative method that combines one dependent response variable with many independent stressor variables, using all possible combinations of the stressor variables, to maximize the variance explained in the response variable. This data mining approach identifies the best single or subset of variables that explains the most variation observed in the biological response variable. For our preliminary analyses, we selected the genera based total taxa and percent EPT variables as our biological response variables as identified in the PCA analysis. All of the independent stressor variables identified in Table 5 were initially included in each of the water quality and physical habitat APR models. Similar to the PCA approach, the water quality variables and physical habitat variables were first analyzed independently then combined in an overall APR model for each biological response variable. The R-squared and root mean square error for each APR model was reviewed to identify a model with the largest R^2 and smallest error term, while minimizing the variable count (Figure 4). The goal of APR analysis is to identify the smallest subset of variables that explains most of the variation, rather than to provide a predictive equation for the subset of variables.

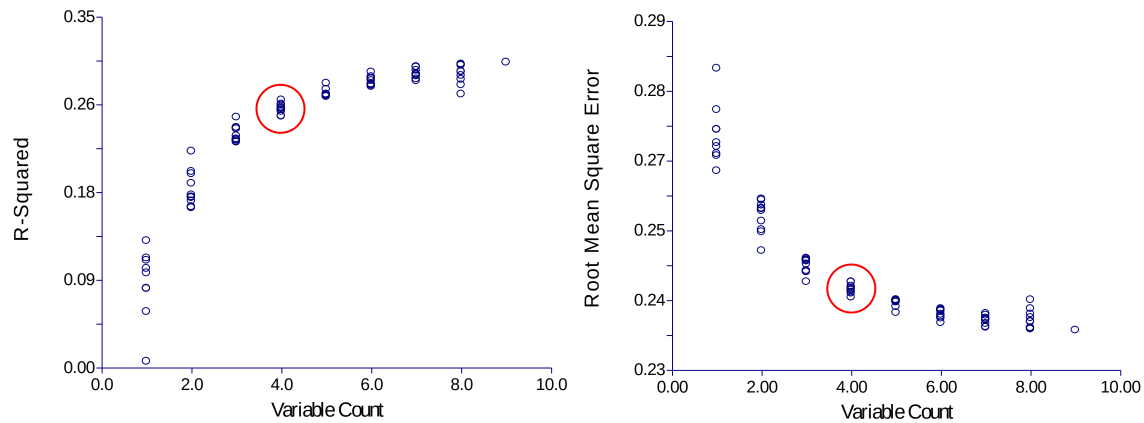

Figure 4: Example of an APR model that maximizes the R-squared and minimizes the root mean square term when four independent stressor variables are selected.

When the total taxa metric was regressed with the water quality variables, the best fit APR model was based three variables that included log transformed alkalinity, sulfate, and total aluminum. However, these three variables only explained approximately 17% of the total variation observed in total taxa. The maximum amount of variation explained by any of the models was only 19%. The best fit physical habitat based total taxa APR loaded four variables: bank stabilization, undisturbed vegetation, channel alteration and embeddedness, although the maximized R^2 was even lower at 9%.

The six variables identified as contributing to the best fit APR models for macroinvertebrate total taxa were combined for an overall APR analysis. The best fit model using both water quality and physical habitat variables loaded three variables: undisturbed vegetation, channel alteration, and log transformed sulfate, and accounted for approximately 21% of the variation observed in total taxa.

The APR analysis of the transformed percent EPT with water quality variables resulted in a best fit model containing five variables: fecal coliform, total aluminum, total calcium, chloride, and total manganese and accounted for approximately 24% of the variation observed in the percent EPT. The maximum amount of variation that could be explained using all water quality variables was 27%. The physical habitat APR resulted in a best fit model that included undisturbed vegetation, embeddedness, epifaunal substrate, and percent fines which explained 16% of the variation in the percent EPT metric. When these water quality and habitat variables were combined in an overall APR analysis, the best fit model included five variables: epifaunal substrate and the log transformed fecal coliforms, total aluminum, chloride, and total manganese. This model accounted for 28% of the variation observed in the percent EPT variable.

4.6 Summary of PCA and APR Analyses

Our preliminary data analyses indicate that conductivity alone is not the most appropriate parameter when trying to explain the variation observed among the Central Appalachian macroinvertebrate communities with respect to water quality and physical habitat. Rather, some combination of ionic composition, substrate, and channel features may be the most appropriate stressor variables to consider (Table 10). These analyses also indicate that total taxa and percent EPT abundance are the key response variables to consider when evaluating factors that shape the macroinvertebrate community, as opposed to a singular focus on Ephemeroptera alone. Additionally, total suspended solids, dissolved oxygen, and fecal coliforms appear to be key variables to consider when evaluating these stream sites, as they are strong indicators of other anthropogenic disturbances in the watersheds. Despite EPA’s underlying assumption that conductivity is the key driver in structuring macroinvertebrate community composition in the Central Appalachian streams, our preliminary analyses suggest that it is more appropriate to evaluate multiple possible stressors, including the specific ions that comprise the measure of specific conductance. Furthermore, it is also important to consider sediment characteristics and habitat disturbance when evaluating macroinvertebrate responses—conductivity, alone, is not a sufficient predictor of biological impairment.

Table 10: Preliminary list of independent stressor variables considered important in the data reduction approach when evaluating stream sites and the two dependent response variables.

Principal Component Analysis	All Possible Regressions	CHAID
Water Quality Variables		
Total magnesium	Chloride	TBD
Total iron	Sulfate	
Total suspended solids	Total aluminum	
Dissolved oxygen	Total manganese	
	Fecal coliforms	
Physical Habitat Variables		
Sediment deposition	Undisturbed vegetation	TBD
Percent fines	Channel alteration	
Channel flow	Epifaunal substrate	

5.0 Summary and Conclusions

The EPA (2010) conductivity benchmark represents a fundamentally different application of an SSD approach for derivation of regulatory benchmarks. In particular, the proposed conductivity benchmark is based on field surveys and correlations between the stressor and biological response in uncontrolled field environments, with multiple species present and all possible biotic (predation/competition/etc.) and abiotic (temperature/flow/season/etc.) interactions occurring. There are several aspects to using a field-based approach that may greatly limit the scientific reliability of using this approach to set specific regulatory thresholds for a composite water quality measurement such as conductivity. The primary disadvantages of using field data result from the fact that exposures are not controlled, and so the causal nature of the relationship between conductivity and the associated biological responses are very difficult to evaluate. As we describe in this report, EPA's arguments supporting the mechanistic plausibility of conductivity as the (virtually only) cause of "impairment" are not convincing, and so cast considerable doubt on the overall reliability of the conductivity benchmark.

Furthermore, any chemical or biological variables that are correlated with conductivity or the biotic response may confound the presumed relationship between conductivity and biological impairment. To address this, EPA (2010) conducts a relatively formal, yet incomplete, analysis of causal mechanisms and confounding factors. EPA concludes that although plausible confounding factors likely exist, their influence is not *strong enough* to prevent use of the conductivity benchmark as presented in this document. EPA needs to conduct a more formal and rigorous analysis to evaluate whether or not conductivity—as opposed to other potentially explanatory factors—is in fact the best and most reliable indicator of adverse changes in biological communities in this region.

The following discussions summarize the major conclusions from each element of our review.

5.1 Diversity of Stressor Response Profiles and Importance to Benchmark Derivation

Multiple stressor-response profiles are exhibited by the genera used in EPA (2010) to derive the conductivity benchmark and, thus, do not represent an internally consistent dataset from which to derive a regulatory benchmark. This suggests that either different invertebrate genera exhibit fundamentally different responses to elevated salinity, or that factors other than conductivity are more closely and functionally related to the capture probability of individual genera across the study region. Therefore, the inclusion of all taxa from the dataset, regardless of their stressor-response profile to conductivity, may be inappropriate for the derivation of a benchmark based on an SSD approach. Indeed, for taxa that exhibit increasing capture probabilities with increasing conductivity, it is possible that extirpation for these species would occur at low conductivities, so the benchmark would clearly not be "protective" with respect to their presence in stream sites in the region.

Therefore, we contend that the use of a SSD of XC_{95} values based on mixed stressor-respond profiles from assumed field distributions is a fundamentally flawed method for derivation of a regulatory benchmark.

5.2 Physiological Mechanisms and Causation

We suggest that there are insufficient data from the physiological and toxicology literature to rigorously support EPA's conclusion that "conductivities in the region of concern reach levels that are sufficient to cause effects on stream communities" (EPA 2010, p. 52). First, although elevated salinity can clearly induce adverse effects on aquatic invertebrates, the taxonomic patterns of sensitivity are not yet clearly defined. Although laboratory toxicity data exposing mayflies to actual or simulated mining effluents suggest they may be somewhat more sensitive than the most sensitive surrogate test species, *C. dubia*, effect concentrations are highly variable and, in some studies, overlap between species. Toxicity to ions associated with salinity also varies strongly as a function of specific ion composition and can be mitigated under conditions of elevated hardness. In fact, criteria based on individual ions—rather than those based composite variables such as conductivity—have already been considered in other states as a preferable regulatory approach that best fits the available scientific information. Additional study is needed to confirm the relative sensitivity of specific macroinvertebrate genera to elevated salinity and the extent to which other water quality variables and major ion composition will influence the consistency of these results.

5.3 Confounding Factors Analysis

The confounding factors analysis in Appendix B of EPA (2010) uses a weight of evidence approach to evaluate whether environmental factors other than conductivity could substantially interfere with or otherwise bias the presumed relationships between conductivity and biological impairment in West Virginia streams. EPA's goal was clearly not to eliminate confounding variables, nor was it an attempt to independently test the hypothesis that conductivity was the best predictor of biological impairment. We do not agree that a confounding factor analysis should take it *as a given* that these are the only or primary relationships that require evaluation. Rather, we contend that a confounding factors analysis should also include rigorous and independent tests of the primary hypothesis and first determine whether conductivity is indeed the best predictor of biological impairment that is causally related in such a way as to justify the proposed benchmark value.

We suggest that elements of EPA's confounding factor analysis would benefit from a closer evaluation to determine whether any of the following factors could provide alternative explanations for patterns in macroinvertebrate community structure relative to MTM/VF activities:

- *Habitat*: There are three clear problems with EPA's assertion that habitat presented little potential for confounding in their derivation of the conductivity benchmark needs additional scrutiny. First, RBP habitat scores may not be the most rigorous measure of habitat quality. Second, the RBP habitat scores were correlated both with conductivity and with the biological response (i.e., the HC_{05} value). Third, the analysis of the potential confounding factors in EPA (2010) focused almost exclusively on the relationship of Ephemeroptera to habitat metrics, to the exclusion of the rest of the benthic macroinvertebrate community.

- *Confounding factors analysis conducted exclusively with Ephemeroptera*: Relationships between all potential stressors (in addition to habitat) and Ephemeroptera were generally cited as reasons to reject the stressors as potential confounders in the analysis that ultimately relates to the entire aquatic benthic community. There is a clear need to include similar analyses from other members of the entire invertebrate community to conclusively reject additional environmental factors as potential confounding stressors.

- *Influence of rare taxa*: EPA (2010) attempted to control for the effect of rare taxa by including only those taxa that had been collected in at least one reference site and at least 30 general sites. Instead it may have been more appropriate for EPA to have controlled for the effects of rare taxa by including in their SSD only those genera that had a high capture probability in the reference sites. A plausible argument against excluding rare taxa from the SSD would be that the taxon is rare because of the stressor. However, this argument would not be valid if the taxon is naturally rare, a phenomenon that could be analyzed using its capture probability in reference sites. EPA (2010) did not sufficiently demonstrate that the rare taxa were rare due to conductivity or any other water quality effect, and not from general rarity itself.

5.4 Ecological Relevance of Presumed Impairment

The ultimate goal of EPA's proposed conductivity benchmark is to determine a conductivity level that, if not exceeded, would prevent extirpation of 95% of the aquatic macroinvertebrate genera. However, it is worth evaluating the relevance of this protection goal when the criterion or benchmark is derived from a very large number of genera, as is the case with the proposed conductivity benchmark. Even if one accepts the role of conductivity in being directly correlated with and responsible for loss of taxa at elevated conductivity, is this 95% protection level ecologically relevant, i.e., do communities in the presence of elevated conductivity lose important ecological functionality?

We evaluated trends in macroinvertebrate community structure and function relative to conductivity from the data presented in EPA (2010). There are few observed changes in the proportional abundance of functional feeding groups within the regional pool of taxa at conductivity levels below approximately 2,500 µS/cm to 5,000 µS/cm. This indicates that the functional aspect of the stream community may not change due to conductivity levels above 300 µS/cm, since genera from all functional feeding groups remain available in the regional taxa pool.

5.5 Independent Statistical Analysis

Our preliminary data analyses from the WABbase dataset indicate that conductivity alone is not the most appropriate parameter when trying to explain the variation observed among the Central Appalachian macroinvertebrate communities with respect to water quality and physical habitat. Rather, ionic composition, substrate, and channel features may be the most appropriate stressor variables to consider. These analyses also indicate that total taxa and percent EPT abundance are the key response variables to consider when evaluating factors that shape the macroinvertebrate community, as opposed to a singular focus on Ephemeroptera alone. Additionally, total suspended solids, dissolved oxygen, and fecal coliforms appear to be key variables to consider when evaluating these stream sites, as they are strong indicators of other anthropogenic disturbances in the watersheds.

5.6 Conclusions

We conclude that the relationship between conductivity and changes in benthic macroinvertebrate community structure is neither strong nor reliable enough to warrant derivation of a regulatory benchmark at this time. While correlations may exist between elevated conductivity and the capture probability of select invertebrate genera, there is insufficient evidence to conclude that elevated concentrations of ions related to salinity are responsible for losses of presumed sensitive taxa. For the most part, this is because EPA (2010) did not rigorously or independently test the primary hypothesis that elevated salinity (as measured by conductivity) was the best predictor of changes in macroinvertebrate community structure in West Virginia streams associated with MTM/VF activities. Rather, most of the analysis conducted in EPA (2010) takes it as a given that conductivity is the best predictor. Furthermore, insufficient laboratory studies are available to confirm either the causal mechanisms or conductivity thresholds that would confirm the proposed benchmark of 300 µS/cm under the specific ion composition of streams in this region. For similar reasons, Illinois, Indiana, and Iowa have rejected the use of TDS or conductivity-based criteria in lieu of criteria for individual ions such as sulfate or chloride.

We also conclude that the use of a SSD of XC_{95} values based on mixed stressor-response profiles from assumed field distributions is a fundamentally flawed method for derivation of a regulatory benchmark. EPA (2010) contends that this approach is appropriate “because sufficient and appropriate laboratory data were not available and because high quality field data were available to relate conductivity to effects on aquatic life in streams and rivers.” We agree that sufficient and appropriate laboratory are not available, but a preferred approach would be to generate such data as could be used for derivation of aquatic life criteria using standard methods

(Stephan et al. 1985). This would avoid the use of mixed stressor-response profiles in deriving the HC_{05} benchmark value, and would also help to confirm or refute the causal linkages between elevated concentrations of salinity and biological impairment to invertebrates. We disagree, however, that the field data rigorously support the hypothesis conductivity is the best predictor of changes in benthic macroinvertebrate community structure in West Virginia streams. Additional study is needed to confirm or refute this hypothesis, both through use of additional statistical hypothesis testing with the existing dataset, and through additional study of West Virginia streams associated with MTM/VF activities.

Therefore, we strongly suggest that it is inappropriate and inadvisable to adopt a conductivity benchmark until or unless such additional study is conducted. To adopt this benchmark without the additional study runs a significant risk of forcing mining operations to expend significant financial resources to reduce conductivity from MTM/VF outfalls, with little confidence that this would achieve the desired goal of preventing extirpation of sensitive genera.

6.0 References

Aladin, N. V., and W. T. W. Potts. 1995. Osmoregulatory capacity of Cladocera. *Journal of Comparative Physiology B* 164:671-683.

Avenet, P., and J. M. Lignon. 1985. Ionic permeabilities of the gill lamina cuticle of the crayfish, *Astacus leptodactylus* (E). *Journal of Physiology* 363:377-401.

Chapman, P. M., H. Bailey, and E. Canaria. 2000. Toxicity of total dissolved solids associated with two mine effluents to chironomid larvae and early life stages of rainbow trout. *Environmental Toxicology and Chemistry* 19:210-214.

Clark, T. M., B. J. Flis, and S. K. Remold. 2004. Differences in the effects of salinity on larval growth and developmental programs of a freshwater and a euryhaline mosquito species (Insecta: Diptera, Culicidae). *Journal of Experimental Biology* 207:2289-2295.

Echols, B. S., R. J. Currie, and D. S. Cherry. 2009. Preliminary results of laboratory toxicity tests with the mayfly, *Isonychia bicolor* (Ephemeroptera: Isonychiidae) for development as a standard test organisms for evaluating streams in the Appalachian coalfields of Virginia and West Virginia. *Environmental Monitoring and Assessment* published online: DOI 10.1007/s10661-009-1191-3.

GEI Consultants, Inc. (GEI). 2009a. Identification of Issues in Regard to the "Pond et al. Study" on Effects of Mountaintop Mining and Valley Fill on Benthic Invertebrate Communities. Technical Memorandum, GEI Consultants, Inc., Denver, CO.

GEI Consultants, Inc. (GEI). 2009b. Update on Issues for Further Investigation in Regard to the "Pond et al. Study" on Effects of Mountaintop Mining and Valley Fill on Benthic Invertebrate Communities. Technical Memorandum, GEI Consultants, Inc. Denver, CO.

Goodfellow, W. L., L. W. Ausley, D. T. Burton, D. L. Denton, P. B. Dorn, D. R. Grothe, M. A. Heber, T. J. Norberg-King, and J. H. Rodgers, Jr. 2000. Major ion toxicity in effluents: a review with permitting recommendations. *Environmental Toxicology and Chemistry* 19: 175-182.

Hart, B. T., P. Bailey, R. Edwards, K. Hortle, K. James, A. McMahon, C. Meredith, and K. Swadling. 1991. A review of the salt sensitivity of the Australian freshwater biota. *Hydrobiologia* 210:105-144.

Hassell, K. L., B. J. Kefford, and D. Nugegoda. 2006. Sub-lethal and chronic lethal salinity tolerances of three freshwater insects: *Cloeon* sp. and *Centroptilum* sp. (Ephemeroptera:

Baetidae) and *Chironomus* sp. (Diptera: Chironomidae). *Journal of Experimental Biology* 209:4024−4032.

Hill, A. B. 1965. The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine* 58:295-300.

Iowa Department of Natural Resources (IDNR). 2009. *Fact Sheet: Revising Criteria for Chloride, Sulfate, and Total Dissolved Solids.* http://www.iowadnr.gov/water/standards/files/ws_fact.pdf, accessed July 7, 2010. Iowa Department of Natural Resources, Des Moines, IA.

Kapoor, N. N. 1979. Osmotic regulation and salinity tolerance of the stonefly nymph, *Paragnetina media*. *Journal of Insect Physiology* 25:17-20.

Kefford, B. J., P. J. Papas, and D. Nugegoda. 2003. Relative salinity tolerance of macroinvertebrates from Barwon River, Victoria, Australia. *Marine and Freshwater Research* 54:755-765.

Kefford, B. J., A. Dalton, C. G. Palmer, and D. Nugegoda. 2004. The salinity tolerance of eggs and hatchlings of selected aquatic macroinvertebrates in south-east Australia and South Africa. *Hydrobiologia* 517:179−192.

Kefford, B. J., and D. Nugegoda. 2005. No evidence for a critical salinity threshold for growth and reproduction in the freshwater snail *Physa acuta*. *Environmental Pollution* 134:377-383.

Kefford, B. J., L. Zalizniak, and D. Nugegoda. 2006. Growth of the damselfly *Ischnura heterosticta* is better in saline water than freshwater. *Environmental Pollution* 141:409-419.

Kefford, B. J., D. Nugegoda, L. Zalizniak, E. J. Fields, and K. L. Hassell. 2007. The salinity tolerance of freshwater macroinvertebrate eggs and hatchlings in comparison to their older life-stages: a diversity of responses. *Aquatic Ecology* 41:335-348.

Kemble, N. 2010. Columbia Environmental Research Center (CERC) quarterly project summary for the project entitled: "Toxicity of Total Dissolved Solids to Appalachian Aquatic Invertebrates". Memorandum to T. Norberg-King and G. Pond. Columbia Environmental Research Center, U.S. Geological Survey, Columbia, MO.

Kennedy, A. J., D. S. Cherry, and R. J. Currie. 2003. Field and laboratory assessment of a coal processing effluent in the Leading Creek watershed, Meigs County, Ohio. *Archives of Environmental Contamination and Toxicology* 44:324-331.

Kennedy, A. J., D. S. Cherry, and R. J. Currie. 2004. Evaluation of ecological relevant bioassays for a lotic system impacted by a coal-mine effluent, using *Isonychia*. *Environmental Monitoring and Assessment* 95:37-55.

Kennedy, A.G., D.S. Cherry, and C.E. Zipper. 2005. Evaluation of ionic contribution to the toxicity of a coal-mine effluent using *Ceriodaphnia dubia*. *Archices of Environmental Contamination and Toxicology* 49:155-162.

Komnick, H. 1977. Chloride cells and chloride epithelia of aquatic insects. *International Review of Cytology* 49:285−328.

Merricks, T. C., D. S. Cherry, C. E. Zipper, R. J. Currie, and T. W. Valenti. 2007. Coal-mine hollow fill and settling pond influences on headwater streams in southern West Virginia, USA. *Environmental Monitoring and Assessment* 129: 359-378.

Mount, D. R., D. D. Gulley, J. R. Hockett, T. D. Garrison, and J. M. Evans. 1997. Statistical models to predict toxicity of major ions to *C. dubia*, *D. magna* and *P. promelas*. *Environmental Toxicology and Chemistry* 10:2009-2019.

Natochin, Y. V., and R. G. Parnova. 1987. Osmolality and electrolyte concentration of hemolymph and the problem of ion and volume regulation of cells in higher insects. *Comparative Biochemistry and Physiology A* 88:563-570.

Norwest Co. 2010. *Comments on "A Field Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams" EPA/600/R-10/023A*. White paper prepared for National Mining Association, Washington DC.

Paulson, C. L., S. Reeves, S. P. Canton, and T. F. Moore. 2001. 2001. Distinguishing the relative influence of habitat and water quality on aquatic biota. *Water Environment Research Foundation, 98-WSM-1*.Pillard, D. A., J. R. Hocket, and D. R. DiBona. 1999. *The Toxicity of Common Ions to Freshwater and Marine Organisms*. Publication number 4666, American Petroleum Institute, Washington, DC.

Piscart, C., P. Usseglio-Polatera, J.-C. Moreteau, and J.-N. Beisel. 2006. The role of salinity in the selection of biological traits of freshwater invertebrates. *Archiv für Hydrobiologie* 166:185-198.

Pond, G. J. 2010. Patterns of Ephemeroptera taxa loss in Appalachian headwater streams (Kentucky, USA). *Hydrobiologia* 641:185−201.

Pond G. J., M. E. Passmore, F. A. Borsuk, L. Reynolds, and 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.

Posthuma, L., G. W. Suter II, and T. P. Traas (eds.). 2002. *Species Sensitivity Distributions in Ecotoxicology*. Lewis Publishers, Boca Raton, FL.

Smith, D. G. 2001. *Pennak's Freshwater Invertebrates of the United States: Porifera to Crustacea*, 4th edition. John Wiley & Sons Inc., New York, NY.

Soucek, D. J. 2007. Comparison of hardness- and chloride-regulated acute effects of sodium sulfate on two freshwater crustaceans. *Environmental Toxicology and Chemistry* 26:773-779.

Soucek, D. J., and A. J. Kennedy. 2005. Effects of hardness, chloride, and acclimation on the acute toxicity of sulfate to freshwater invertebrates. *Environmental Toxicology and Chemistry* 24:1204-1210.

Stephan, C. E., D. I. Mount, D. J. Hansen, J. H. Gentile, G. A. Chapman, and W. A. Brungs. 1985. *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*. PB-85-227049. U.S. Environmental Protection Agency, Office of Research and Development, Duluth, MN.

Stewart, K. W., and B. P. Stark. 2002. *Nymphs of North American Stonefly Genera (Plecoptera)*, 2nd edition. Caddis Press, Columbus, OH.

Thorp, J., and A. P. Covich. 2001. *Ecology and Classification of North American Freshwater Invertebrates.* Academic Press, San Diego, CA.

U.S. Environmental Protection Agency (EPA). 2000. Stressor identification guidance document. EPA/822/B-00/025, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (EPA). 2008. *Supporting Rationale for Decision, Revisions to Indiana's Sulfate Water Quality Standards. WQSTS IN2008-257.* U.S. Environmental Protection Agency, Region 5, Chicago, IL.

U.S. Environmental Protection Agency (EPA). 2009. *The Effects of Mountaintop Mines and Valley Fills on Aquatic Ecosystems of the Central Appalachian Coalfields*. Draft EPA/600/R-09/138A, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (EPA). 2010. *A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams.* Draft EPA/600/R-10/023A, U.S. Environmental Protection Agency, Washington, DC.

Woodward, D. F., R. G. Riley, M. G. Henry, J. S. Meyer, and T. R. Garland. 1985. Leaching of retorted oil shale: assessing the toxicity to Colorado squawfish, fathead minnow, and two food-chain organisms. *Transactions of the American Fisheries Society* 114:887-894.

Zalizniak, L. B., B. J. Kefford, and D. Nugegoda. 2009. Effects of pH on salinity tolerance of selected freshwater invertebrates. *Aquatic Ecology* 43:135-144.