

Technical Memo

Date: April 19, 2017

Re: **Evaluation of data and methodology associated with USEPA's (Susan M. Cormier, Ph.D.) review of the Johnson and Johnson report titled "An Evaluation of a Field-Based Aquatic Benchmark for Specific Conductance in Northeast Minnesota" (November 2015)**

Executive Summary

GEI Consultants, Inc. (GEI) reviewed the U.S. Environmental Protection Agency's white paper and corresponding data regarding the development of a conductivity benchmark for Ecoregion 50 (Northern Lakes and Forest) in Minnesota. It is GEI's professional opinion that the methodology used to develop the benchmark contains inconsistencies and flaws, as does the supporting data, which precludes their use in establishing a conductivity benchmark for Ecoregion 50.

GEI found multiple stressor-response profiles in the most sensitive taxa that provided conflicting evidence for the genera's presumed physiological limits to conductivity which highlights flaws in the conductivity benchmark approach. For example, the most sensitive benthic macroinvertebrate genus found in Ecoregions 69 & 70, *Lepidostoma* (XC95 = 121 $\mu\text{S}/\text{cm}$), has a positive 1,162 percent change in the extirpation coefficient compared to Ecoregion 50 (XC95 = 1,527 $\mu\text{S}/\text{cm}$). The third most sensitive genus in Ecoregion 50, *Rhyacophila* (XC95 = 254 $\mu\text{S}/\text{cm}$), has a negative 87 percent change in Ecoregions 69 & 70 (XC95 >1,890 $\mu\text{S}/\text{cm}$). These substantially conflicting extirpation coefficients for the same genera highlight the inconsistencies in the supposed physiological responses at the genus level and represent a flaw in the EPA's conductivity benchmark approach. The finding that virtually all of the common genera found in Ecoregion 50 had substantially different extirpation coefficients when compared to the same genera, and in some cases the same species, found in Ecoregion 69 & 70 undermines the premise that measurements of conductivity are the dominant stressor that affect the distribution of benthic invertebrate taxa.

GEI also observed multiple data-related inconsistencies and questionable water chemistry values during evaluation of benthic invertebrate and water chemistry data sets utilized to develop the Ecoregion 50 conductivity criteria. For example, approximately 60 percent of the MPCA macroinvertebrate data was missing paired water quality data, including conductivity measurements. GEI also identified significant issues in the reproducibility and traceability of water quality data. It is unclear the extent to which these factors may affect the analysis but does suggest the possibility of underlying bias. The key findings outlined above are elaborated upon in the sections that follow and demonstrate that the conductivity benchmark approach is not appropriate for Ecoregion 50.



1. Background and Introduction

GEI Consultants, Inc. (GEI) was asked to continue assisting Cliffs Natural Resources, U.S. Steel, and ArcelorMittal (“the clients”) in responding to the issue of the possible development of a conductivity benchmark in Minnesota. On February 4, 2016, the U.S. Environmental Protection Agency (USEPA) published a memo (authored by Dr. Susan Cormier) that reviewed the November 2015 Johnson and Johnson report titled “An Evaluation of a Field-Based Aquatic Life Benchmark for Specific Conductance in Northeast Minnesota” (Johnson and Johnson 2015). USEPA’s memo evaluated benthic invertebrate and water quality data sets generated by the Minnesota Pollution Control Agency (MPCA) and supported the conclusions of Johnson and Johnson (2015) concerning the effects of conductivity on benthic invertebrates, citing that “...that more than 5% of genera would be extirpated in streams greater than 320 $\mu\text{S}/\text{cm}$.” GEI has obtained the MPCA data utilized by Dr. Cormier in USEPA’s review of the Johnson and Johnson report. This memo provides a summary of GEI’s in-depth examination of the data and underlying methodology utilized in USEPA’s memo. It should be noted that these comments will also have direct relevance to USEPA’s more recent national December 2016 draft “Field-Based Methods for Developing Aquatic Life Criteria for Specific Conductivity” (USEPA 2016).

GEI conducted a brief evaluation of the data provided by Dr. Cormier and those retrieved from the MPCA FTP site and provided initial findings in September 2016. The data were then further examined in detail with respect to 1) the extent to which the Dr. Cormier dataset could be reproduced from MPCA Macroinvertebrate and MPCA Water Quality files, and 2) the completeness of the available data and its integration into the final data set utilized by Dr. Cormier. GEI’s findings are as follows.

1.1 Conductivity Benchmark Approach for Ecoregion 50

- *GEI found multiple species response profiles in the most sensitive taxa that provided conflicting evidence for the genera’s physiological limits to conductivity which raises significant uncertainty in the benchmark approach.*

One of the major conceptual issues with the USEPA conductivity benchmark is the assumption that the absence of any benthic invertebrate genera is solely due to a conductivity level that exceeds the physiological limits of that genus. However, as we have noted in our prior comments to the underlying benchmark document (GEI 2010, Roark et al. 2013), there are many factors that contribute to the absence of benthic invertebrates from a stream sample, such as interspecific competition, habitat suitability, other stressors (i.e., metals and sedimentation), or simply sampling and sample processing bias, which were not addressed in the document.

However, if the premise is true that absence is due to conductivity’s effects on a genus’ physiological limits, then any particular genus’ extirpation coefficient (XC_{95}) should be consistent across ecoregions. A comparison of the extirpation coefficients (XC_{95}) for

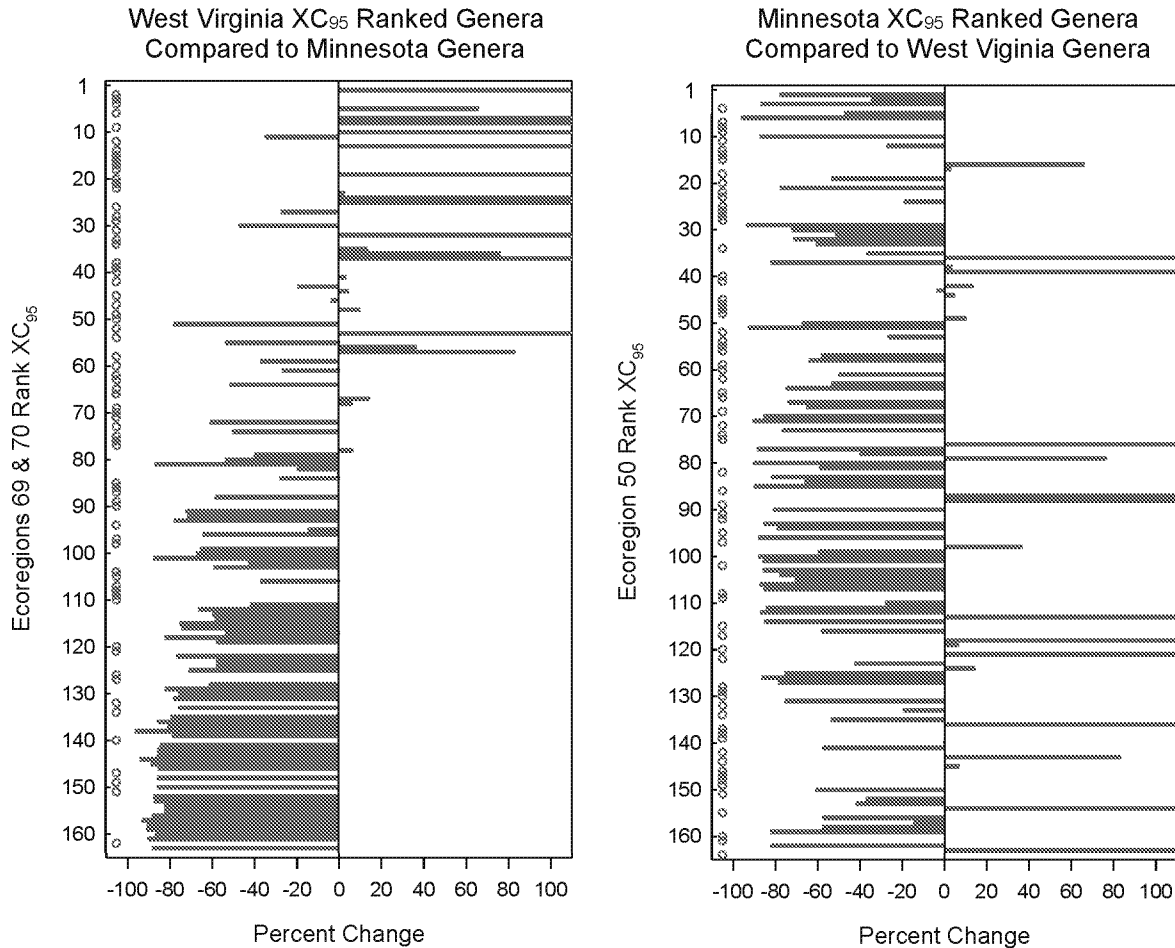


Ecoregion 50 with Ecoregions 69 & 70 provides some insight to the physiological limits and/or conflicting limits previously identified for some taxa. The Minnesota taxa list was compared to the West Virginia taxa (from USEPA 2011) list to identify common genera among the ecoregions as well as to identify unique genera in Ecoregion 50. Of the total 164 genera found in Minnesota, 95 genera were common to Ecoregions 69 & 70 of the Appalachians, while 69 genera were unique to Ecoregion 50 in northeastern Minnesota.

When considering the entire taxa list, most genera rankings greatly changed between ecoregions, putting the concept of conductivity as the prime reason for presence/absence into significant doubt. In fact, when considering the 20 most sensitive taxa based on their XC₉₅ ranking for each ecoregion (i.e., 50 and 69 & 70), there are only two genera – *Leptophlebia* and *Epeorus* – that are common to both lists strongly indicating there is not a universally expressed relationship between presence of genera and conductivity. Stated plainly, differences in extirpation coefficients (Figure 1 – next page) for common genera suggests that any purported relationship to “conductivity” is simply an artifact of the benchmark methodology, or other factors affect the frequency of occurrence (or absence) of invertebrates.



Figure 1: The percent change in extirpation coefficients for common genera from Ecoregions 69 & 70 compared to Ecoregion 50. Rank 1 is the smallest XC_{95} value for each ecoregion. Open circles denote unique genera to each ecoregion. Positive percent change values are truncated at 110%, because the maximum percent changes was 1,162 percent. Percent Change = $[(MN\ XC_{95} - WV\ XC_{95}) / WV\ XC_{95}] \times 100$.



Notably, the most sensitive genus found in Ecoregions 69 & 70 was *Lepidostoma* ($XC_{95} = 121\ \mu\text{S}/\text{cm}$). However, in Ecoregion 50, *Lepidostoma* was actually one of the least sensitive genera ($XC_{95} = 1,527\ \mu\text{S}/\text{cm}$). The difference between these two extirpation coefficients represents a *positive 1,162 percent change* for this genus (Figure 1, left panel Rank 1, right panel Rank 121) in Ecoregion 50.

Similarly, the third most sensitive genus in Ecoregion 50 – *Rhyacophila* ($XC_{95} = 254\ \mu\text{S}/\text{cm}$) – was shown to be fairly tolerant of conductivity ($XC_{95} > 1,890\ \mu\text{S}/\text{cm}$) in Ecoregions 69 & 70; a *negative 87 percent change* (Figure 1, right panel Rank 3). These differences highlight our concern that genera characterized as being sensitive to conductivity in one ecoregion may in fact not be sensitive to conductivity in another ecoregion. Such large variability in the purported physiological limits of “sensitive” genera raises considerable uncertainty regarding



the applicability of a surrogate measurement such as conductivity to determine the frequency of occurrence (or absence) of a given taxa. This is an extremely important issue that puts into question the entire approach used in EPA’s conductivity benchmark analysis and needs to be more fully investigated by EPA before any conductivity “criterion” is adopted.

To further explore these differences, we examined the 20 most sensitive genera in Ecoregion 50 (Table 1) in the context of previously identified species sensitivity distributions for conductivity (USEPA 2011). The Dr. Cormier data file was filtered accordingly (i.e., pH > 6, genera present \geq 25 samples, invertebrates identified to at least genera level) which resulted in 743 sampling events representing 600 sampling locations in Ecoregion 50. The dataset was not censored for conductivity values on the extreme left or right tails of the distribution. The list includes 7 genera of Ephemeroptera (mayflies), 5 genera of Diptera (midges), 3 genera of Trichoptera (caddisflies), 2 genera of Odonata (dragonflies), and 1 genera each of Plecoptera (stoneflies), Lepidoptera (moth), and Basommatophora (snail). For these genera, the weighted cumulative distributions were generated to determine if the XC₉₅ values could be replicated and species sensitivity distributions were generated to evaluate the relationship with conductivity. Of the 20 genera identified in Ecoregion 50, only 10 genera were common to Ecoregions 69 & 70 in West Virginia. GEI’s calculated metrics (e.g., number samples, XC₉₅) were generally within \pm 2% (RPD) with the differences likely due to the binning approach (Roark et al. 2013).

Table 1: The twenty most sensitive genera in Ecoregion 50 with summary information and comparison to extirpation coefficients in Ecoregions 69 & 70. Shaded cells denote common genera. Percent Single Individual per Sample is the percentage of the total number of occurrences represented by only 1 individual.

Rank	EcoR 50 XC ₉₅	EcoR 69 & 70 XC ₉₅	Order	Family	Genus	Present in Number of Samples	% Single Individual per Sample
1	191	863	Trichoptera	Philopotamidae	<i>Dolophilodes</i>	86	7%
2	201	307	Ephemeroptera	Heptageniidae	<i>Epeorus</i>	96	24%
3	254	1,890	Trichoptera	Rhyacophilidae	<i>Rhyacophila</i>	39	62%
4	272		Odonata	Gomphidae	<i>Ophiogomphus</i>	75	63%
5	283	535	Ephemeroptera	Ephemerellidae	<i>Serratella</i>	43	23%
6	298	7,340	Odonata	Aeshnidae	<i>Boyeria</i>	126	60%
7	302		Plecoptera	Perlidae	<i>Agnatina</i>	25	52%
8	327		Diptera	Chironomidae	<i>Trissopelopia</i>	26	35%
9	335		Diptera	Chironomidae	<i>Xenochironomus</i>	36	58%
10	338	2,630	Diptera	Chironomidae	<i>Larsia</i>	25	36%
11	338		Lepidoptera	Crambidae	<i>Paraponyx</i>	33	58%
12	357	490	Ephemeroptera	Ephemerellidae	<i>Eurylophella</i>	157	34%
13	361		Diptera	Chironomidae	<i>Stictochironomus</i>	47	53%
14	374		Basommatophora	Planorbidae	<i>Helisoma</i>	96	38%
15	390		Diptera	Chironomidae	<i>Lopescladius</i>	60	47%
16	416	251	Ephemeroptera	Leptophlebiidae	<i>Leptophlebia</i>	43	23%
17	435	424	Ephemeroptera	Heptageniidae	<i>Leucrocota</i>	129	29%



Rank	EcoR 50 XC ₉₅	EcoR 69 & 70 XC ₉₅	Order	Family	Genus	Present in Number of Samples	% Single Individual per Sample
18	456		Ephemeroptera	Baetidae	<i>Labiobaetis</i>	60	30%
19	464	996	Ephemeroptera	Baetidae	<i>Plauditus</i>	38	24%
20	502		Trichoptera	Leptoceridae	<i>Triaenodes</i>	58	47%

In addition to the notable differences in XC₉₅ values outlined above, Table 1 also points out another concern with developing extirpation coefficient and species sensitivity distributions using field based count data. Specifically, the relative abundance of any one genera is not factored into the presence/absence benchmark approach. Thus, single individuals are afforded the same weighting as multiple individuals in a sample. This is tenuous when extirpation of a genus is largely pinned on the presence or absence of a single individual. The benthic invertebrate processing approach that often utilizes subsampling of the entire sample that can greatly affect the outcome of distributions. In the case of the Ecoregion 50 data, for the 20 most sensitive genera, single individuals represented from 7% to 63% of their respective occurrences (see far right column in Table 1) with an average of 40% of the data being used to develop extirpation coefficients and species sensitivity distributions based on genera represented by a single organism in a sample. With such a significant portion of the data comprised of single individuals, sampling bias could be significantly influencing the outcome of the use of presence/absence data.

1.2 Diversity of Conflicting Stressor-Response Profiles

- *GEI reviewed the twenty most sensitive genera in Ecoregion 50, and found multiple stressor-response profiles that provide conflicting results within the most sensitive genera as well as compared to response-profiles from other ecoregions, a key fundamental flaw in the approach.*

One of the underlying principles governing the use of a species sensitivity distribution (SSD) to derive biological thresholds is that all of the organisms represented in the distribution exhibit the same type of response to the stressor in question (Posthuma et al. 2002). However, three types of stressor-responses are recognized by Dr. Cormier (2016, EPA 2010), as exemplified in Figure 2, and a fourth type (GEI 2010) not recognized by EPA—but was observed in the Ecoregion 50 dataset. The fourth type of profile is basically characterized by no response or a bimodal (i.e., inverse optimal) response to conductivity (Figure 2). The no response profile results in conflicting stressor response concentrations when the tails of the distribution are used to establish thresholds. The four response profiles are:

- Decreasing probability of observation with increasing conductivity,
- Increasing probability of observation with increasing conductivity, and
- Optimal or “bell-curve” probability of observation with increasing conductivity.



- No response or bimodal, where probability of observation is not related to increasing conductivity



Figure 2: Biological response profiles with respect to conductivity and probability of capture.

Based on our review of the twenty most sensitive taxa in Ecoregion 50, the most common response profile was the no response/bimodal profile (9 of 20), followed by the decreasing profile (7 of 20), optimum profile (2 of 20), and increasing profile (2 of 20)

In Ecoregion 50, the most sensitive genus - *Dolophilodes* - was comprised of three species with *Dolophilodes distinctus* being the most common taxon observed. These three taxa generated a decreasing SSD (Figure A-1, Attachment) that was very similar in shape to previously identified SSD for *Dolophilodes* in Ecoregions 69 & 70, albeit the extirpation coefficient was considerably greater at 863 $\mu\text{S}/\text{cm}$. *Dolophilodes distinctus* was present in West Virginia but an unidentified species was more abundant.

The genera *Epeorus* and *Rhyacophila* exhibited similar decreasing patterns in their SSD although their response range was considerably more narrow than the West Virginia data relationships (Figure 3). As discussed above, *Rhyacophila* exhibited vastly different extirpation coefficients between the two ecoregions (MN = 254 $\mu\text{S}/\text{cm}$, WV >1,890 $\mu\text{S}/\text{cm}$) despite sharing a common species *R. fuscata*, although the most abundant taxon remained unidentified in both ecoregions.

The 4th ranked genera (*Ophiogomphus*) is a dragonfly that exhibited relatively low capture probabilities (i.e., <30%) with a decreasing pattern through the range of conductivity values from 45 to 200 $\mu\text{S}/\text{cm}$, but this genus was not observed in WV which limited comparability of the SSD (Figure A-1, Attachment).

The genus *Seratella* (mayfly) was comprised of the species *S. serrata* and an unidentified sp. that exhibited low capture probabilities (i.e., <15%) from 45 to 420 $\mu\text{S}/\text{cm}$. Similarly, in West Virginia, this genus exhibited low capture probabilities (i.e., <15%) through the range of conductivity from approximately 40 to 500 $\mu\text{S}/\text{cm}$ (Figure A-1, Attachment). The extirpation coefficient in Ecoregion 50 was calculated as 283 $\mu\text{S}/\text{cm}$ and was 535 $\mu\text{S}/\text{cm}$ in Ecoregions 69 & 70, representing nearly a two-fold difference in the sensitivity of this genera.

The dragonfly *Boyeria* was the third most commonly observed genera of the MN top 20 genera with *B. grafiana*, *B. vinosa*, and an unidentified sp. comprising the observed individuals. *B. grafiana* and *B. vinosa* were also present in the WV database as well as an unidentified sp.,



thus given the similar taxonomic composition it would be expected that the taxonomic response to conductivity would be similar for each ecoregion. However, the SSDs are extremely different with the Ecoregion 50 SSD exhibiting a decreasing relationship with conductivity ($XC_{95} = 298 \mu\text{S}/\text{cm}$, and the Ecoregions 69 & 70 SSD exhibiting a unimodal relationship with an estimated extirpation coefficient of $>7,340 \mu\text{S}/\text{cm}$ (Figure A-1, Attachment). The differences in physiological responses and extirpation levels for this genus raises concern for confounding factors such a seasonal life-stages, habitat preferences, and timing of sampling. Seasonal patterns and habitat availability such as debris dams, emergent vegetation, and coarse substrates are likely important factors in regulating the distribution of *Boyeria* in Northeastern Minnesota (Haarstad 1997). In fact, *B. grafiana* is a species of special concern in Minnesota (MN Administrative Rule 6134.0200) indicating the species is extremely uncommon in Minnesota, and has a unique or highly specific habitat requirements.

The 7th through 9th ranked genera consisted of a stonefly (*Agnatina*) and two midges (*Trissopelopia* and *Xenochironomus*) that bracket the proposed conductivity benchmark of $320 \mu\text{S}/\text{cm}$ for Ecoregion 50. These three taxa were not present in the Ecoregion 69 & 70 dataset so there are no comparative SSD. These three taxa exhibit poor capture probabilities, generally less than 15%, through the conductivity range of approximately 50 to $300 \mu\text{S}/\text{cm}$ and do not show a clear response to conductivity (Figure A-1, Attachment). Furthermore, these taxa were observed in a small number of samples just above the ≥ 25 sample cut-off ranging from 25 to 36 samples out of the 734 samples comprising the Ecoregion 50 dataset used by Dr. Cormier, with nearly 50% of the count data represented by a single individual found in the sample. The absence of a well-defined response to conductivity raises concern regarding the appropriateness of calculating “extirpation coefficients” when the rarity of a taxon is not fully vetted with respect to habitat preferences, substrate conditions, or flow characteristics, much less a lack of significant response to conductivity.

The 10th ranked genus *Larsia*, also a midge, was observed in the Ecoregion 69 & 70 data set and exhibited an extirpation coefficient of $2,360 \mu\text{S}/\text{cm}$ as compared to extirpation coefficient of $338 \mu\text{S}/\text{cm}$ for Ecoregion 50 (Figure A-1, Attachment). This genus also exhibited poor capture probabilities of $<10\%$ through the range of conductivity that dominated conditions observed in Ecoregion 50. Similarly, the absence of a well-defined response to conductivity raises concerns.

Comparisons of the 11th through 20th ranked genera identified only four common genera between Ecoregions 50, 69, and 70 which were all mayflies – *Eurylophella*, *Leptophlebia*, *Leucrocuta*, *Plauditus*. Both *Eurylophella* and *Leucrocuta* exhibited decreasing SSD that was similar in shape among the ecoregions and with respect to extirpation coefficient (Figure A-1, Attachment). *Leptophlebia* exhibited a no response/bimodal response in Ecoregion 50 while the SSD for Ecoregions 69 & 70 was decreasing. *Plauditus* exhibited an optimum response with observations noted on the lower and upper tail of the distribution, which is very similar to the response in Ecoregions 69 & 70, but the extirpation coefficient was very different in



Ecoregion 50. *Leptophlebia* and *Leucrocuta* exhibited a greater extirpation coefficient in Ecoregion 50 when compared to the Ecoregions 69 & 70. The remaining six genera ranked 11th through 20th were unique to Ecoregion 50 with respect to their XC₉₅ data – *Paraponyx*, *Stictochironomus*, *Helisoma*, *Lopescladius*, *Labiobaetis*, and *Triaenodes*. These genera exhibited a mix of no response/bimodal and increasing SSD profiles as a function of conductivity, indicating low sensitivity to increasing specific conductance.

Based on our review of the twenty most sensitive genera in Ecoregion 50, there are multiple stressor-response profiles that provide conflicting results within the most sensitive genera as well as compared to response-profiles from other ecoregions. These conflicting stressor-response profiles do not represent an internally consistent dataset especially when considering the genera that bracket the 5th centile hazardous concentration (320 µg/L). These genera do not exhibit an increasing sensitivity to increasing conductivity. This is a key fundamental flaw in the approach, as it suggests that either invertebrate genera are exhibiting fundamentally different physiological responses to elevated conductivity levels or, more likely, that factors other than conductivity are much more closely and functionally related to the probability of finding a genus in an ecoregion.

2. Data Validation

- *GEI's review of the data sets revealed that approximately 60 percent of the MPCA macroinvertebrate data was missing paired water quality data, including conductivity measurements. Significant issues in the reproducibility and traceability of water quality data were also documented.*

Dr. Cormier provided GEI a zip file containing a single file (in Microsoft Excel® csv format) with benthic invertebrate and water quality data that were apparently used to develop the XC₉₅ values published in USEPA's February 4, 2016 memo. GEI also obtained 23 Excel files from an MPCA FTP site, one of which was the file that Dr. Cormier provided to GEI. The MPCA files contain a substantially larger amount of data than the file passed along by Dr. Cormier (**Error! Reference source not found.2**). Twenty of 23 files contained a mix of water quality and physicochemical measurements with a minimum of habitat quality site descriptors. Of these 20 files, eight files did not contain data for Ecoregion 50 – Northern Lakes and Forest – the region of interest in Johnson and Johnson (2015) and USEPA (2016). The remaining three files contained benthic invertebrate (2 files) and fishery (1 file) data along with limited water quality data and Minnesota Stream Habitat Assessment (MSHA) data. In summary, there are water quality files and aquatic life use files with limited water quality (e.g. specific conductance and nutrients) collected at the time of the sampling as well as descriptive habitat data (e.g., percent riffles and substrate etc.). As such, information is lacking for many sampling efforts, precluding a full understanding of water quality and physical characteristics during each sampling effort.

**Table 2: Data files obtained from Dr. Cormier and the MPCA FTP site.**

Data File Names
Dr. Cormier data file
InvertsWFieldChem_Rev20150123.csv
MPCA Macroinvertebrate data file
SendMeAllTheData_InvertsChemMSHAHabitat_2014Oct21.xlsx
MPCA Water Quality data files
04010201_07010203_07020005_2007_2013.xlsx
04010301_07010107_07010108_2007_2013.xlsx
07010101_07020007_07020010_2007_2013.xlsx
Data File Names
07010205_07040004_09020303_2007_2013.xlsx
07020002_07010204_07010202_2007_2013.xlsx
07020004_07040003_09020101_2007_2013.xlsx
07020011_07040001_07040008_2007_2013.xlsx
07030005_07080201_07080202_2007_2013.xlsx
07040002_09020301_09020304_2007_2013.xlsx
09020102_09030006_04010102_2007_2013.xlsx
09020104_09020311_09030005_2007_2013.xlsx
09020106_07010106_07010206_2007_2013.xlsx
09020306_09030009_09020312_2007_2013.xlsx
09020309_04010101_07010207_2007_2013.xlsx
10170202_10170203_10170204_2007_2013.xlsx
10230003_07010102_07010105_2007_2013.xlsx
Samples_1996_2000.xlsx
Streams_1996_2000.xlsx
Streams_2001_2003.xlsx
Streams_2004_2006.xlsx
MPCA Fish data file
SendMeAllTheData_FishChemMSHAHabitat_2014Oct21.xlsx

2.1 Data Completeness and Consistency

- *GEI found multiple data-related inconsistencies and questionable water chemistry values during evaluation of how benthic invertebrate and water chemistry data were paired up in Dr. Cormier's final data set.*

For Dr. Cormier's review of "An Evaluation of A Field-Based Aquatic Life Benchmark For Specific Conductance In Northeast Minnesota" (Johnson and Johnson 2015), a dataset was developed from multiple water quality and macroinvertebrate datasets provided by the Minnesota Pollution Control Agency (MPCA). The MPCA macroinvertebrate dataset is a compilation of data from a variety of sampling events of which roughly 40 percent included



paired water quality data (i.e., temperature, conductivity, pH, dissolved oxygen and limited nutrients) while the remaining 60 percent of the sample events did not include any water quality data. The MPCA macroinvertebrate dataset did not include any ion chemistry such as sulfate, carbonates, or chloride.

Given these findings, it's not surprising the file provided by Dr. Cormier corroborated that *approximately 60 percent of the MPCA macroinvertebrate data was missing paired water quality data*, including conductivity measurements. Therefore, the annual geometric mean for selected water quality measures – temperature, pH, dissolved oxygen and others – was calculated for each sample location in the MPCA water quality data files and used by MPCA to provide paired water quality with macroinvertebrate data. These calculated geometric mean values are essentially “estimated” values that cannot be verified and thus, their impact on the overall data set cannot be quantified. The Dr. Cormier dataset included a data qualifier indicating an “Exact Match” or “Same Year Match” which presumably identified the paired water quality data collected at the time of the macroinvertebrate sample, or the calculated geometric mean water quality data used to provide paired results, respectively.

GEI investigated the continuity and completeness of the dataset used by Dr. Cormier to establish the conductivity benchmark value of 320 $\mu\text{S}/\text{cm}$ by retracing the origin of the water quality data. Within the “Exact Match” and “Same Year Match” groups we selected the sampling events with the highest and lowest conductivity measurements and three other sampling events that exhibited a high, mid, and low conductivity measurement for a total of 10 samples to trace the origins of the water quality data (Table A-1 and Table A-2, Attachment). Notably, the file provided by Dr. Cormier contained macroinvertebrate and water quality data for Ecoregions 46, 47, 48, 49, 50, 51, and 52, but we filtered the data to only examine Ecoregion 50 data as well as for other MPCA water quality data files.

For our comparison, the data sources – Dr. Cormier file, MPCA Macroinvertebrate file, and 20 MPCA Water Quality files were cross referenced using a combination of Field Number, Location Code, Location Description, Basin Code, Sample Date, and Latitude/Longitude to trace the origin/availability of the water quality data. Each sampling location was spatially referenced in Google Earth to evaluate data availability for nearby sampling locations. The geometric mean of selected water quality parameters, from the MPCA Water Quality datasets, was calculated for cases when multiple samples were collected in the same year at the same site, which was consistent with the data estimation procedure used by Dr. Cormier.

The macroinvertebrate total taxa, total count and mean conductivity for each of the “Exact Match” qualified sampling events were the same between the Dr. Cormier and MPCA Macroinvertebrate data files. This was reassuring because these two parameters provided the basis of the benchmark approach.



In contrast, however, GEI found that the “Exact Match” water quality data (e.g. water temperature, pH, dissolved oxygen) was not always the same between the Dr. Cormier data file and the presumed parent file – MPCA Macroinvertebrate data file that provided the data (Table A-1 and Table A-2, Attachment). Three out of the five sampling events showed data inconsistencies. For two of the sampling events that revealed data inconsistencies, there was no data present in the MPCA Water Quality data files that could supplement these data. Based on the MPCA Macroinvertebrate data file, it is reasonable to assume that if water quality data was collected at the time of the macroinvertebrate sampling the data would remain paired to ensure data integrity. This is important from the standpoint of evaluating confounding effects other than conductivity on the macroinvertebrate assemblages.

We also found data inconsistencies for the selected sampling events qualified as “Same Year Match” in the Dr. Cormier data file. All sampling events selected in the Dr. Cormier data file were located in the MPCA Macroinvertebrate dataset, along with the macroinvertebrate total taxa and count being the same. Two of the five sampling events, however, included water quality data in the MPCA Macroinvertebrate data file which was not expected. These two sampling events should have been qualified as “Exact Match” but weren’t. Three of the five samples selected from the Dr. Cormier dataset were located in the MPCA Water Quality data files but two of these were not sampled in the same year as the macroinvertebrates. Only one of the five sampling events had water quality data associated with the sampling location and in the same year to estimate the annual geometric mean value to replace data gaps, and the geometric mean or mean values are close but do not match the estimated data in Dr. Cormier’s file. The data file should have included comments and/or notes enabling users to understand what values were used when paired water quality values were not available.

During this investigation of continuity and completeness of all datasets, GEI noticed a repetitive pattern in the Dr. Cormier and MPCA Macroinvertebrate data files. In both data files, specific numbers were often repeated across dates, sample sites, waterbodies, ecoregions, and analytes (Table 3 and Table 4). This repetitiveness may not appear odd if the values were whole numbers, but instead, some were to the 14th decimal place. The chances of, for example, both DO and pH having a calculated geometric mean value of 8.39000034332275 is very unlikely. At some stage in the data management process it appears that data columns and values were erroneously sorted or pasted into cells causing the repetition throughout the data file. From a data management perspective, the repetitive values raise a substantial concern on the validity of evaluating the effects of water quality on macroinvertebrates in Ecoregion 50.

**Table 3: Examples of commonly repeated values (shaded cells) for estimated data used to fill data gaps in the Dr. Cormier data file. Shaded cells denote consistent values repeated in the data.**

Field Number	Ecoregion	Sample Date	DO (mg/L)	pH	Water Temperature (C°)	Data Qualifier
05RN079	50	8/8/2005	10.30000019000000	7.89999985700000	20.32499981000000	Same Year Match
97LS047	50	9/11/1997	10.30000019000000	7.69000005700000	15.50000000000000	Same Year Match
05RD009	51	9/22/2005	10.30000019000000	8.00000000000000	22.31999970000000	Same Year Match
04LM095	52	9/1/2004	10.30000019000000	7.76999998100000	17.05999947000000	Same Year Match
97LS071	50	9/10/2013	9.89999961900000	6.26000022900000	17.70000076000000	Same Year Match
04LM127	52	8/17/2004	9.89999961900000	7.69999980900000	16.29999924000000	Same Year Match
09CD058	47	8/12/2009	9.89999961900000	8.18000030500000	17.79999924000000	Same Year Match
09CD029	47	8/6/2009	6.26000022900000	6.65999984700000	19.29999924000000	Same Year Match
04CD017	47	8/31/2004	17.70000076000000	9.39999961900000	22.22999954000000	Same Year Match
00UM101	51	10/9/2000	10.80000019000000	8.90999984700000	17.70000076000000	Same Year Match
04LM083	51	8/18/2004	7.59999990500000	8.00000000000000	17.70000076000000	Same Year Match

Table 4: Examples of commonly repeated values (shaded cells) found in the MPCA Macroinvertebrate data file for sampling events where paired water quality data was collected with the macroinvertebrate sample.

Field Number	Water Body Name	Ecoregion	Sample Date	DO (mg/L)	pH
10UM055	Shell River	51	8/30/2010	8.39000034332275	8.39000034332275
11MS114	Rock River	47	8/8/2011	8.39000034332275	8.36999988555908
13RD007	Snake River	48	8/6/2013	8.39000034332275	7.98999977111816
00UM010	Mississippi River	50	8/26/2013	7.96999979019165	7.53000020980834
11LS017	Captain Jacobson Creek	50	8/16/2011	8.60999965667725	7.53000020980834
13UM021	Mississippi River	50	9/5/2013	7.53000020980834	7.88999986648560
98LS026	Crow Creek	50	8/8/2011	9.25000000000000	7.53000020980834
12UM140	Dabill Creek	50	8/2/2012	0.20000000298023	7.40999984741211
12UM112	Leech Lake River	50	8/29/2012	7.40999984741211	7.48999977111816
10MN108	Smith Creek	47	8/9/2010	7.40999984741211	7.80000019073486
10RN040	Big Fork River	49	9/2/2010	7.40999984741211	7.88000011444092
11LM054	Trib. to Prairie Creek	47	8/3/2011	7.40999984741211	8.09000015258789

Our review of the data sets revealed that approximately 60 percent of the MPCA macroinvertebrate data was missing paired water quality data, including conductivity measurements, and identified significant issues in the reproducibility and traceability of water quality data. It is unclear the extent to which these factors may affect the analysis but does suggest the possibility of underlying bias.



3. Summary and Conclusions

The lack of internal consistency for species response profiles and extirpation coefficients demonstrates significant flaws with the EPA's conductivity benchmark approach. The finding that virtually all of the common genera found in Ecoregion 50 had substantially different calculated XC_{95} values when compared to the same taxa found in Ecoregion 69 & 70 invalidates the entire underlying premise of the conductivity benchmark that conductivity levels affect the distribution of benthic invertebrate taxa.

While arguments have been presented elsewhere (EPA 2016) that differing extirpation coefficients for the same genera would be expected because the physiological limits of species within a genus may be different is a concern with how a national level criterion approach is applied to site-specific water quality conditions. As noted above, the genus *Boyeria* exemplifies this issue in the Ecoregion 50 and 69 & 70 data sets. All three ecoregions are represented by two common species *Boyeria grafiiana* and *Boyeria vinosa* as well as an unidentified species. Given the commonality of the taxa at the species level it would be expected that extirpation coefficients would be similar for the genus – ***yet there was a difference of over 7,000 $\mu\text{S}/\text{cm}$ in their XC_{95} values***, again invalidating the premise that conductivity is the primary factor that affects the distribution of this genus.

The multiple response profiles found in the twenty most sensitive genera also raises concern from a criterion development standpoint. The most "sensitive" genera should all exhibit a decreasing SSD profile such that the effect of a stressor (i.e., conductivity) is well defined and does not provide conflicting responses. However, in Ecoregion 50, the genera that bracket the 5th centile hazard concentration (320 $\mu\text{S}/\text{cm}$) – and, thus, are largely responsible for the calculated benchmark – exhibit very poor capture probabilities (<10%) over the range of conductivity conditions and generally show no response/bimodal or an optimum response which results in conflicting stressor thresholds indicating their distribution is not related to conductivity.

While a premise of this approach is that different species within a genus can have vastly different physiological limits, it is the intent of the 1985 criteria development guidelines to document these physiological limits for each species such that genus level chronic and acute responses can be weighted accordingly. As such, the use of XC_{95} values derived from conflicting SSD response profiles at this low end of the SSD is yet another fundamental flaw in the approach used to establish a benchmark conductivity value for Ecoregion 50, and is not consistent with guidelines for developing a water quality criterion (Stephan et al. 1985).

Further, there were significant issues in the reproducibility and traceability of water quality data that raise concern in using the dataset to validate macroinvertebrate response to conductivity or other water quality conditions. There are numerous instances where water quality data was estimated to fill data gaps – almost 60% of the sites where there was not a site



or data available in that year to fill the missing information. Furthermore, water quality data that was collected at the time of the macroinvertebrate sampling event was not always the same value found in Dr. Cormier's data file. This lack of matched invertebrate/conductivity data, as well as the lack of agreement and traceability between Dr. Cormier, MPCA Macroinvertebrate, and MPCA Water Quality data files combined with the repeated parameter values for a small subset of samples examined indicates inconsistent and incomplete datasets and is a flawed foundation for attempting to establish a conductivity benchmark for Ecoregion 50.



References

- Cormier, S.M. 2016. *An Evaluation of a Field-based Aquatic Life Benchmark for Specific Conductance in Northeast Minnesota*. National Center for Environmental Assessment – Cincinnati, Ohio. Office of Research and Development, U.S. EPA. February 2016.
- GEI Consultants, Inc. 2010. *Technical Review: A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams*. Prepared for National Mining Association, Washington, DC. September 2010.
- Haarstad, J. 1997. *The Dragonflies of Selected Eastern Minnesota Rivers*. Prepared for the Minnesota Department of Natural Resources.
- Johnson, B.L. and M.K. Johnson. 2015. Review of *An Evaluation of a Field-based Aquatic Life Benchmark for Specific Conductance in Northeast Minnesota*. Prepared for Water Legacy. November 2015.
- Posthuma, L., G. W. Suter II, and T. P. Traas (eds.). 2002. *Species Sensitivity Distributions in Ecotoxicology*. Lewis Publishers, Boca Raton, FL.
- Roark, S. A, C.F. Wolf, G.D., De Jong, R.W. Gensemer, and S.P. Canton. 2013. Influences of subsampling and modeling assumptions on the US Environmental Protection Agency field-based benchmark for conductivity. *Integrated Environmental Assessment and Management*: 9(3) 2013.
- 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. EPA, Office of Research and Development, Duluth, MN.
- USEPA. 2011. *A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams*. National Center for Environmental Assessment, Office of Research and Development, U.S. EPA Washington, DC. EPA-600-R-10-023F. March 2011.
- USEPA. 2016. *DRAFT: Field-based Methods for Developing Aquatic Life Criteria for Specific Conductivity*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA-822-R-07-010. December 2016.

Attachments

Table A-1: Traceability of the “Exact Match” water quality data present in Dr. Cormier’s data file with the MPCA Macroinvertebrate data file and the MPCA Water Quality data files. Shaded cells denote inconsistencies in the data.

Conductivity	Comparison	Dr. Cormier	MPCA Macroinvertebrate	MPCA Water Quality
Lowest	Location	12UM140	12UM140/ Dabill Creek	No data. No sites located on Dabill Creek
	Lat, long	--	46.72779, -94.53639	
	Sample date	8/2/2012	8/2/12 8:43	
	Macroinvertebrate taxa	89	89	
	Macroinvertebrate count	609	609	
	Mean conductivity	3.880000114	3.880000114	
	Minimum conductivity	3.880000114	--	
	Maximum conductivity	3.880000114	--	
	DO	2.290000059	0.200000003	
	pH	7.279999971	7.409999847	
Temperature	18.09999999	21.5		
Low	Location	13LS008	13LS008/ South Brule River	5007-327/ S Brule R. at Gunflint Tr. (Cook Cr-12), 6.6 mi. SW of E Cook, MN
	Lat, long	--	47.92698, -90.31118	47.926659, -90.307443
	Sample date	8/13/2013	8/13/13 14:25	8/13/2013 7:30
	Macroinvertebrate taxa	45	45	--
	Macroinvertebrate count	322	322	--
	Mean conductivity	33.90000153	33.90000153	34.1
	Minimum conductivity	33.90000153	--	--
	Maximum conductivity	33.90000153	--	--
	DO	8.895000219	9.81000042	8.74
	pH	7.069999933	7	--
Temperature	20.5	18.70000076	16	
Mid	Location	12UM088	12UM088/ Necktie River	5006-256/ Necktie R. at county state aid Highway 45. Site is located 2.5 mi. NE of Laporte, MN
	Lat, long	--	47.24681, -94.72887	47.248083, -94.727972
	Sample date	8/29/2012	8/29/12 14:43	8/27/2012 11:12
	Macroinvertebrate taxa	21	21	--
	Macroinvertebrate count	314	314	--



Conductivity	Comparison	Dr. Cormier	MPCA Macroinvertebrate	MPCA Water Quality
	Mean conductivity	438	438	449
	Minimum conductivity	438	--	--
	Maximum conductivity	438	--	--
	DO	3.089999914	3.089999914	2.59
	pH	7.170000076	7.170000076	7.37
	Temperature	23.60000038	23.60000038	20
High	Location	11LS075	11LS075/ Trib. to McQuade Lake	5007-255/ Unn. Str. at Hayes Rd., 3 mi. SSE of Buhl, MN. T58R19WS34
	Lat, long	--	47.45847, -92.74807	47.45761, -92.748022
	Sample date	9/12/2011	9/12/11 13:41	No 2011 data. Site in database but samples did not occur in 2011. No other sites on same tributary have sample dates in 2011.
	Macroinvertebrate taxa	36	36	
	Macroinvertebrate count	316	316	
	Mean conductivity	825	825	
	Minimum conductivity	825	--	
	Maximum conductivity	825	--	
	DO	9.390000343	9.390000343	
	pH	8.039999962	8.039999962	
	Temperature	21.29999924	21.29999924	
Highest	Location	78SC001	78SC001/ Wolf Creek	
	Lat, long	--	46.1296704672535, -92.6271409649827	
	Sample date	8/17/2010	8/17/10 11:40	
	Macroinvertebrate taxa	53	53	
	Macroinvertebrate count	319	319	
	Mean conductivity	1594	1594	
	Minimum conductivity	1594	--	
	Maximum conductivity	1594	--	
	DO	7.700000048	8.460000038	
	pH	7.424999952	7.619999886	
	Temperature	19.59999943	15.89999962	



Table A-2: Traceability of the “Same Year Match” water quality data present in Dr. Cormier’s data file with the MPCA Macroinvertebrate data file and the MPCA Water Quality data files. Shaded cells denote inconsistencies in the data.

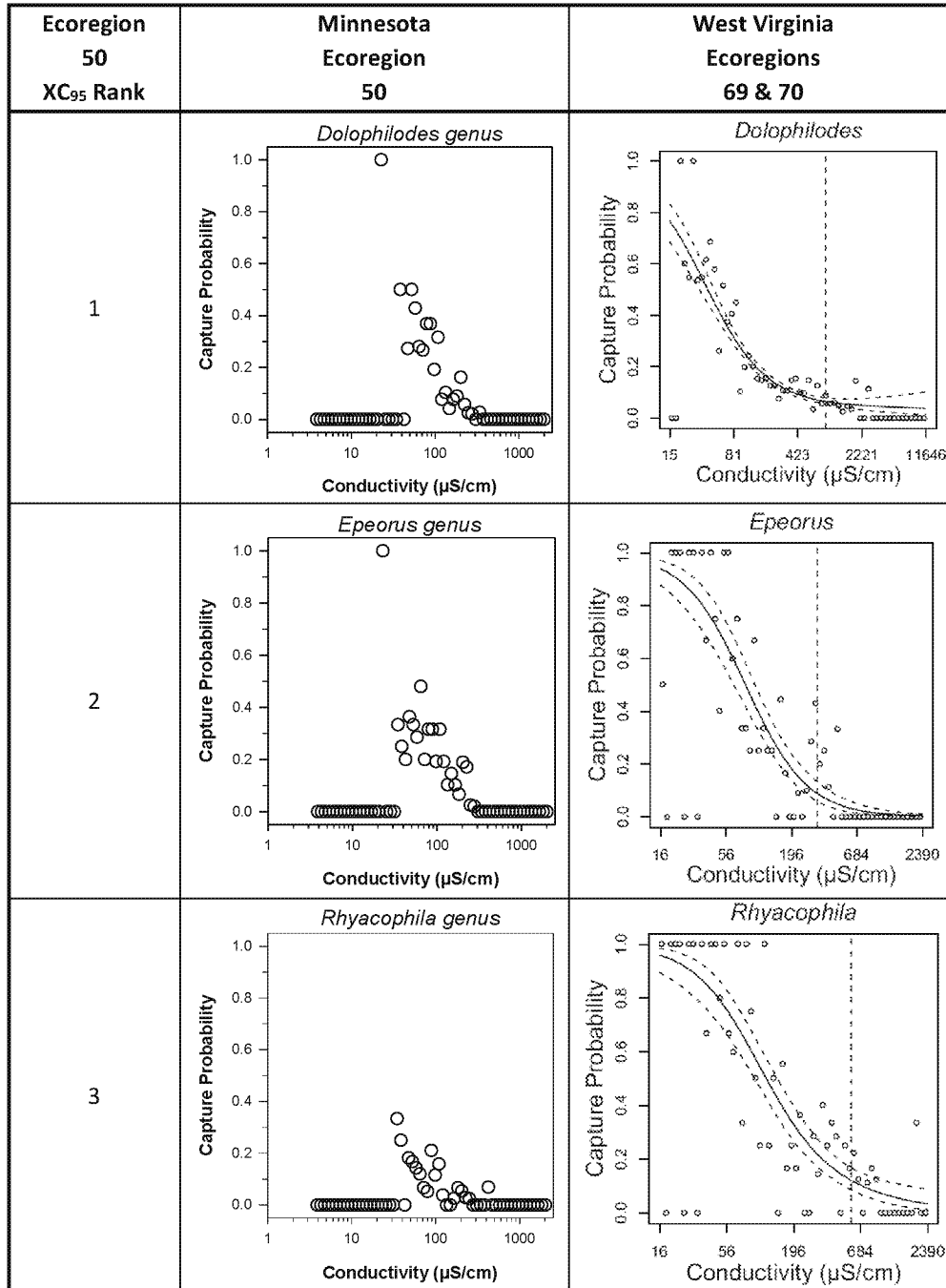
Conductivity	Comparison	Dr. Cormier	MPCA Macroinvertebrate	MPCA Water Quality
Lowest	Location	13LS001	13LS001/ Portage Brook	No data. No sites located on Portage Brook
	Lat, long	--	47.99209, -90.04431	
	Sample date	9/10/2013	9/10/13 10:09	
	Macroinvertebrate taxa	45	45	
	Macroinvertebrate count	313	313	
	Mean conductivity	22.16499996	26	
	Minimum conductivity	21.53000069	--	
	Maximum conductivity	22.79999924	--	
	DO	9.505000115	9.329999924	
	pH	6.700000048	7.429999828	
	Temperature	19.64999962	16.10000038	
Low	Location	97LS071	97LS071/ Stump River	No data. No sites located on Stump River
	Lat, long	--	48.0188748454965, -90.0340544538571	
	Sample date	9/10/2013	9/10/13 14:04	
	Macroinvertebrate taxa	132	132	
	Macroinvertebrate count	630	630	
	Mean conductivity	31.20000076	49.09999847	
	Minimum conductivity	31.20000076	--	
	Maximum conductivity	31.20000076	--	
	DO	9.899999619	8.079999924	
	pH	6.260000229	7.099999905	
	Temperature	17.70000076	17.10000038	
Mid	Location	09LS073	09LS073/ West Two River	West Two R. at CR-661, 4 mi. SW of Forbes
	Lat, long	--	47.34011667, -92.68305	47.338853, -92.683021
	Sample date	8/11/2009	8/11/2009	12 samples in 2009
	Macroinvertebrate taxa	54	54	--
	Macroinvertebrate count	314	314	--
	Mean conductivity	533	No data as expected	520.7
	Minimum conductivity	529		341



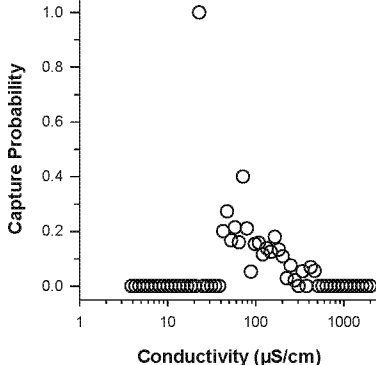
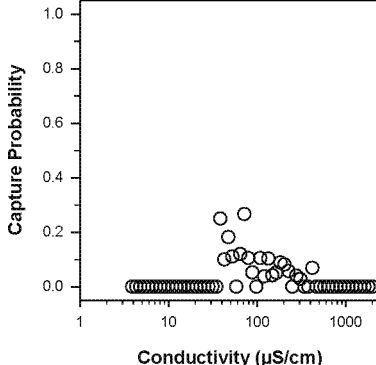
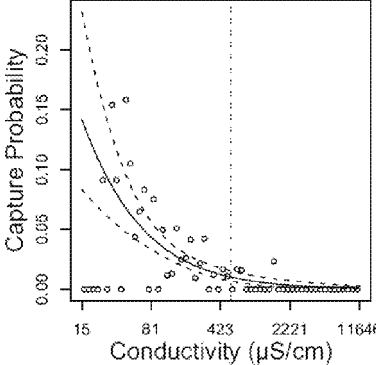
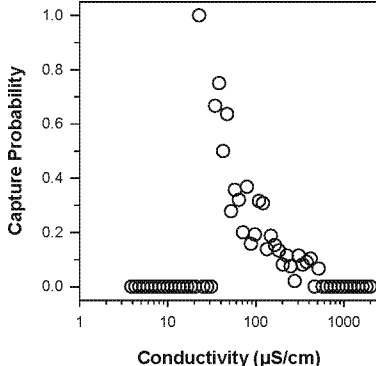
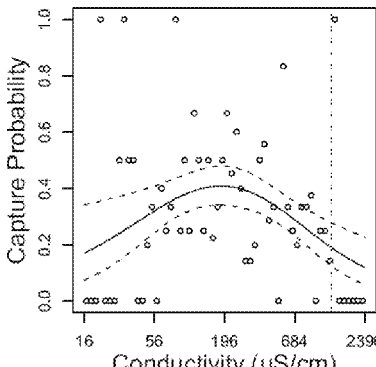
Conductivity	Comparison	Dr. Cormier	MPCA Macroinvertebrate	MPCA Water Quality
	Maximum conductivity	537		601
	DO	7.329999924		9.1
	pH	8.085000038		8.081818182
	Temperature	19.10000038		18.86363636
High	Location	05RN079	05RN079/ Dark River	Dark R. at Sherwood Anderson Rd., 7.5 mi. N of Buhl
	Lat, long	--	47.6228004276397, -92.7309723802619	47.62358, -92.73175
	Sample date	8/8/2005	8/8/2005	No samples in 2005
	Macroinvertebrate taxa	44	44	--
	Macroinvertebrate count	296	296	--
	Mean conductivity	1447	No data as expected	No data. Site in database but samples did not occur in 2005. No other sites on Dark R. have conductivity data.
	Minimum conductivity	1083		
	Maximum conductivity	1811		
	DO	10.30000019		
	pH	7.899999857		
Temperature	20.32499981			
Highest	Location	09LS005	09LS005/ Otter Creek	Otter Ck. at CSAH-1 / 3rd St. Br. in Carlton, MN
	Lat, long	--	46.66095, -92.42676667	46.660849, -92.42444
	Sample date	9/10/2009	9/10/2009	No samples in 2009
	Macroinvertebrate taxa	104	104	--
	Macroinvertebrate count	648	648	--
	Mean conductivity	1997.699951	No data as expected	No data. Site in database but samples did not occur in 2009.
	Minimum conductivity	1997.699951		
	Maximum conductivity	1997.699951		
	DO	10.14000034		
	pH	7.53000021		
Temperature	8.399999619			



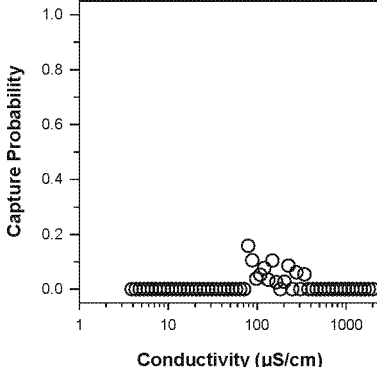
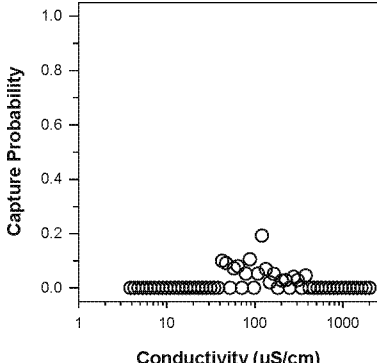
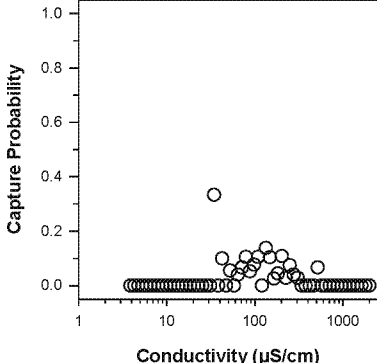
Figure A-1: The species sensitivity distributions for the 20 most sensitive genera found in Ecoregion 50 and a comparison to those genera that were common in Ecoregions 69 & 70. The Ecoregion 50 SSD were generated as part of this review and Ecoregion 69 & 70 SSD are from Appendix E (EPA 2011).



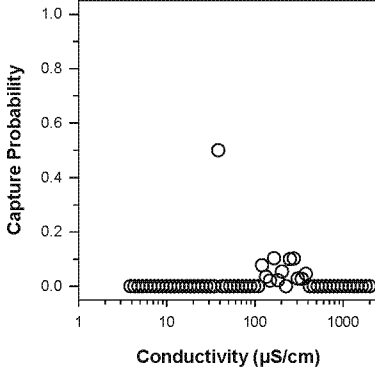
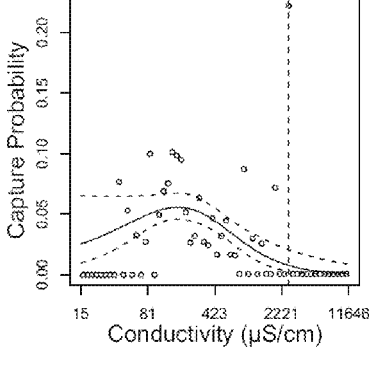
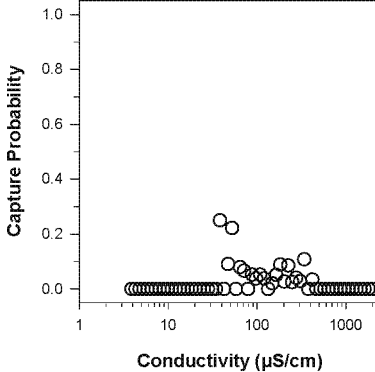
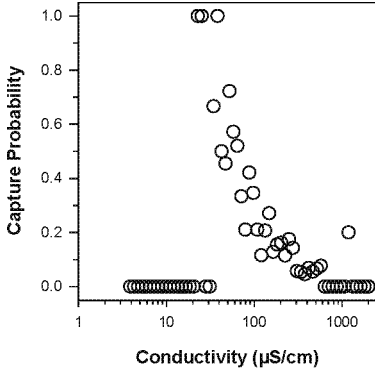
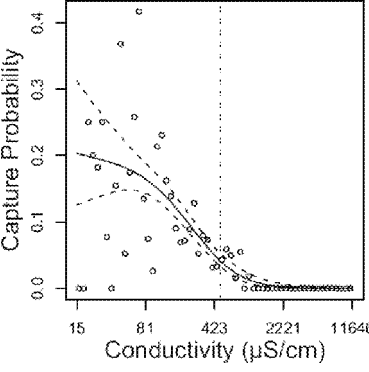


Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
4	<p data-bbox="657 362 873 392"><i>Ophiogomphus</i> genus</p> 	Not Available
5	<p data-bbox="685 776 857 806"><i>Serratella</i> genus</p> 	<p data-bbox="1107 771 1214 802"><i>Serratella</i></p> 
6	<p data-bbox="695 1190 847 1220"><i>Boyeria</i> genus</p> 	<p data-bbox="1123 1185 1205 1215"><i>Boyeria</i></p> 

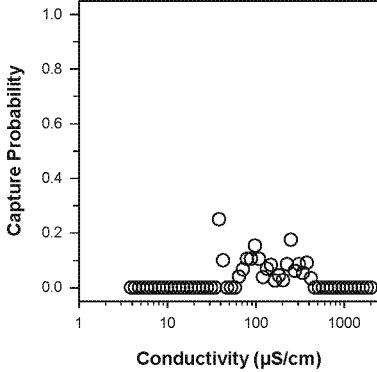
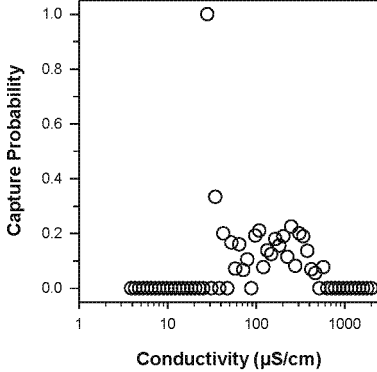
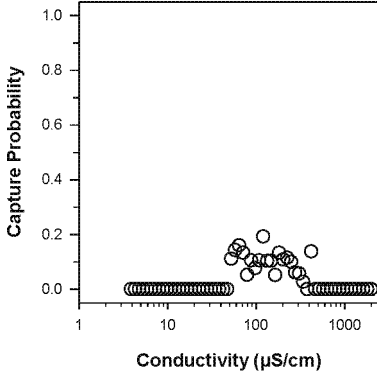


Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
7	<p data-bbox="686 362 849 390"><i>Agetina</i> genus</p> 	Not Available
8	<p data-bbox="670 776 865 804"><i>Trissopelopia</i> genus</p> 	Not Available
9	<p data-bbox="654 1190 881 1218"><i>Xenochironomus</i> genus</p> 	Not Available

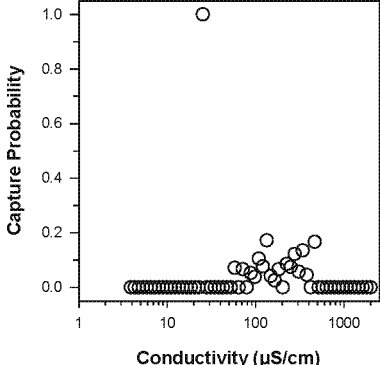
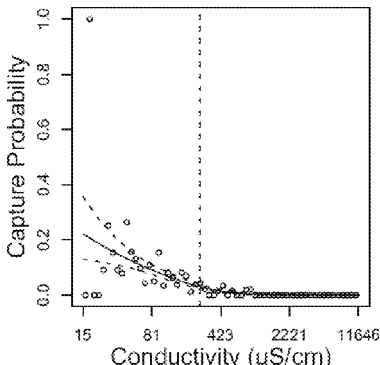
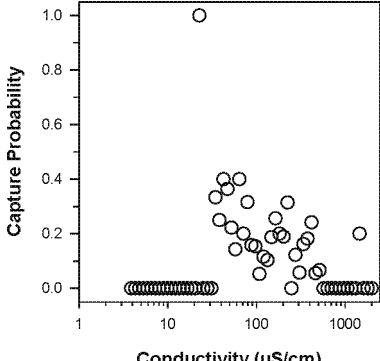
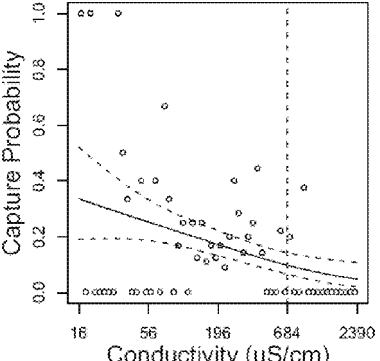
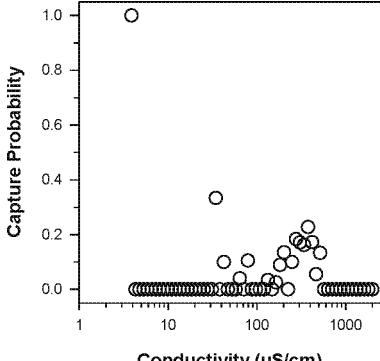


Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
10	<p data-bbox="699 362 829 390"><i>Larsia</i> genus</p> 	<p data-bbox="1133 362 1198 390"><i>Larsia</i></p> 
11	<p data-bbox="678 774 850 802"><i>Paraonyx</i> genus</p> 	<p data-bbox="1060 961 1206 989">Not Available</p>
12	<p data-bbox="667 1185 862 1213"><i>Eurylophella</i> genus</p> 	<p data-bbox="1097 1181 1227 1209"><i>Eurylophella</i></p> 

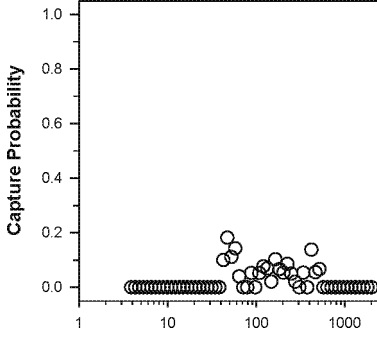
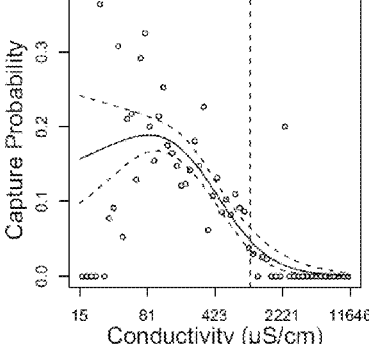
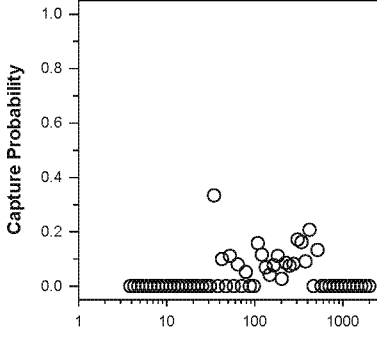


Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
13	<p data-bbox="646 362 889 390"><i>Stictochironomus</i> genus</p> 	Not Available
14	<p data-bbox="683 774 852 802"><i>Helisoma</i> genus</p> 	Not Available
15	<p data-bbox="667 1185 868 1213"><i>Lopescladius</i> genus</p> 	Not Available



Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
16	<p data-bbox="667 362 862 390"><i>Leptophlebia</i> genus</p> 	<p data-bbox="1094 362 1230 390"><i>Leptophlebia</i></p> 
17	<p data-bbox="675 774 854 802"><i>Leucrocuta</i> genus</p> 	<p data-bbox="1102 774 1219 802"><i>Leucrocuta</i></p> 
18	<p data-bbox="675 1190 854 1218"><i>Labiobaetis</i> genus</p> 	Not Available



Ecoregion 50 XC ₉₅ Rank	Minnesota Ecoregion 50	West Virginia Ecoregions 69 & 70
19	<p data-bbox="683 362 849 390"><i>Plauditus</i> genus</p>  <p data-bbox="670 739 862 761">Conductivity (µS/cm)</p>	<p data-bbox="1117 362 1214 390"><i>Plauditus</i></p>  <p data-bbox="1060 717 1279 739">Conductivity (µS/cm)</p>
20	<p data-bbox="678 778 854 806"><i>Triænodes</i> genus</p>  <p data-bbox="670 1151 862 1172">Conductivity (µS/cm)</p>	Not Available