Peer Review Workshop of EPA's Draft Report, Implications of Climate Change for Bioassessment Programs and Approaches to Account for Effects

Reviewer Post-Meeting Comments

June 16, 2011

Contents

Reviewer Biographies

Daren M. Carlisle, Ph.D.

Chief, Ecology National Synthesis
U.S. Geological Survey

After obtaining B.S. degree in Fishery Science, Dr. Carlisle studied invertebrate and fish communities in alpine lakes for his M.S. in Aquatic Ecology. He then worked for nearly a year as a water-quality specialist for the Idaho Division of Environmental Quality, as the technical lead on a team to develop a TMDL for nutrients for the Middle Snake River. He then began Ph.D. program in Ecotoxicology at Colorado State University studying the food webs of streams under the influence of legacy mining. After completing graduate school he worked as a regional ecologist for the National Park Service, where he assisted park units develop long-monitoring plans, investigate contaminant issues, and inventory aquatic species. Since 2002 Dr. Carlisle has been with the USGS National Water-Quality Assessment Program, where he led an effort to integrate and synthesize biological, chemical, and physical data collected over the life of the Program.

M. Siobhan Fennessy, Ph.D.

Professor, Biology Department
Kenyon College

Siobhan Fennessy is a professor of Biology at Kenyon College where she studies freshwater wetlands, biological assessment methods, watershed-based assessment of wetlands, restoration ecology, and the role of temperate wetlands in the global carbon cycle. She previously served on the faculty of the Geography Department of University College London and held a joint appointment at the Station Biologique du la Tour du Valat investigating human impacts to Mediterranean wetlands. During a subsequent position at Ohio EPA, she founded Ohio's wetland bioassessment program and wrote the current wetland water quality standards designed to protect Ohio's wetlands. She was a member of the U.S EPA's Biological Assessment of Wetlands Workgroup, a national technical committee working to develop biological indicators of ecosystem condition, has been lead on several grant-funded projects designed to test the use of rapid and biotic assessment methods to evaluate the ecological condition of wetlands on a watershed basis. She is currently working with USEPA to develop a national rapid assessment method for use in the national wetland condition assessment project.

Eric P. Smith, Ph.D.

Professor of Statistics
Virginia Polytechnic Institute and State University

Dr. Eric P. Smith is a Professor of Statistics at Virginia Tech. He joined the faculty in 1982 and has been the chair of the department since 2006. His research focuses on the development and application of statistical methods to help understand and solve environmental and ecological problems. Some recent problems include design of multimetric indices, development and evaluation of biological and chemical standards and design of monitoring programs. He was the director of the Statistical Consulting Center 1995-2004. In that position he was responsible for providing statistical support to students, faculty and staff and provided training to statistics students on the art of consulting. He has worked on a variety of statistical and scientific problems from areas such as engineering, education and natural resources. Dr. Smith teaches courses on multivariate analysis and linear models (regression, analysis of variance). Dr. Smith serves as Associate Editor of *Environmetrics* and is a former associate editor of the *Journal of Agricultural, Biological and Environmental Statistics* and the

Journal of the American Statistical Association. He has supervised 14 Ph.D. students and has authored numerous papers on environmental and ecological statistics.

R. Jan Stevenson, Ph.D.

Professor, Department of Zoology
Co-Director, Center for Water Sciences
Michigan State University

Dr. Stevenson received his Ph.D. from the University of Michigan on 1981. He is currently a profesosr in the Department of Zoology and Co-Driector of the Center for Water Sciences at Michigan State University. He employs his technical expertise in algal taxonomy and ecology to test ecological theory and to develop approaches for solving environmental problems. Dr. Stevenson is particularly interested in how ecological systems respond to environmental change. He also works with federal and state officials to develop protocols for ecological assessment. Working with resource managers and policy makers often stimulates new directions for his research.

Most of Dr. Stevenson's projects use field observations, experiments, and modeling to better understand the effects of natural and human factors on algae and the role of algae in aquatic ecosystems. Field studies are used to identify interesting patterns in nature that may indicate an environmental problem or an interesting natural phenomenon. Manipulative experiments in artificial streams and mesocosms are used to confirm cause-effect relationships. He uses models to scale their observations up to better understand cause-effect relations among natural and anthropogenic factors and ecological conditions.

Major Research Projects

- developing national protocols for assessment of streams and wetlands with the US EPA;
- a survey of 100 Kentucky and Michigan wetlands to determine abiotic effects on algal biomass and species composition;
- a survey of 500 streams in Western US to develop predictive models of that algal species that should be found given natural climatic, geologic, and hydrological features of the stream;
- assessing ecological condition of streams in western US with periphyton assemblages;
- relating ecological response of streams, lakes and wetlands to human activities in the Muskegon River Watershed;
- evaluating the value of grouping similar types of streams for improving predictions of relations between ecological condition, stressors, and human activities in the Great Lakes region;
- determining the causes of nuisance algal growths in Florida Springs; and
- developing a better understanding of relations between nitrogen and phosphorus contamination, algae, bacteria, and low dissolved oxygen problems in streams.

N. Scott Urquhart, Ph.D.

Statistical Consultant
Windsor, CO

N. Scott Urquhart was trained as a statistician at Colorado State University (CSU) – PhD, 1965. He held tenured faculty positions in the Ag schools at Cornell University (1965-1970), and New Mexico State University (1970- 1991), and untenured research positions at Oregon State University (OSU, 1991 – 2000), and CSU (2001 – 2006). He was elected a fellow of the American Statistical Association (a status limited to no more than 5% of ASA's membership) in 1988 in recognition for his contributions to consulting, teaching, and professional service.

Throughout his career Urquhart engaged in collaborated research with a wide range of environmental, ecological, agricultural and engineering scientists. These collaborations led to 1 book, about 60 journal articles, 19 proceedings or book chapters, 9 agency publications, and about 100 other pieces of professional output such as abstracts, and technical reports. Journals relevant to this review include the Journal of Agricultural, Biological and Environmental Statistics, Environmetrics, Biometrics, Ecological Applications, Environmental Monitoring and Assessment, Freshwater Biology, and Environmental Management.

During his time at OSU, Urquhart cooperated closely with EPA's Environmental Monitoring and Assessment Program (EMAP), developing and implementing methods for comparing different proposed revisit patterns relative to their power to detect trend, and evaluate status (= response size in a population); he became familiar with the sampling and evaluation of macro invertebrates. During his time at CSU he managed the EPA-funded Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP), a program also affiliated with EMAP.

Urquhart has participated in at least four other EPA review panels, chaired one for the Grand Canyon Monitoring and Research Center, and reviewed publications for several federal agencies.

Responses to Charge Questions

General Comments

Carlisle	This report tackles an issue that has far-reaching implications for how aquatic ecosystems are assessed and managed. The primary focus of this study is the potential consequences of climatic-induced shifts in the baselines and climatic-related confounding of the indicators used to assess ecological health. Although the report contains much useful information and findings relevant to this important topic, the current organizational structure renders this information extremely difficult to locate, interpret, and judge. In addition, there are several important lines of evidence on which the report is inexplicably silent—most notably hydrological variability and characterization of the longer-term climatic context of the period examined (i.e., the last 30 years). Finally, in an effort to fully understand the depth of the analyses in the Report, I focused on a site (Weber, Utah) about which I have first-hand knowledge. I have provided several Attachments from data sets that I was able to readily obtain and that I believe would enhance the study. In doing so, I realize that scope is a limitation of every study. However, I think that examination of some additional data sets would be beneficial and strengthen the evidence for the Study's primary conclusions.
Smith	The purpose of this report is to evaluate the potential effects of climate change on benthic monitoring and its use for evaluation of water quality. The researchers used four sampling programs to try to find consistent changes in macroinvertebrate indices associated with changes in climate. This type of data is what I call "found" data in that it was not collected to answer the question of interest. Therefore my expectations are low as there are confounding factors that cannot be easily controlled. While the researchers have done an admirable job of trying to address a difficult problem, there are several issues that severely limit the utility of the work and its conclusions: 1. There is an excessive focus on the use of p-values and significance as the measure of the importance of a relationship between a benthic metric and climate change. I suggest the authors consider using an approach that is more connected to the science and expected changes or look at other statistical measures such as slope and size of correlation. At the least, there should be an adjustment of the p-values for the number of tests that are run. 2. Climate change has many components and expected changes. The authors seem to focus on defining climate change as temperature change. Aren't there other components that merit evaluation such as precipitation in the east and fire in the west? Correlation analysis also is susceptible to spurious relationships. I suspect correlations that are based on temperature are confounded with urbanization. It will be hard to argue that the changes are due just to temperature. 3. A further difficulty is the temperature metric is not well connected to the site since it is not directly measured at the site. Use of average temperature seems somewhat arbitrary. Even if temperature data are available it is not clear how to combine the information in temperature into a metric that is connected to benthic metrics.

	Temperature data tends to be measured over a short time interval while benthic samples are taken in either spring or fall. What is the appropriate temperature metric? If PRISM data is used, uncertainty associated with the resulting temperature values need to be discussed and possibly estimated. How good are the estimate temperatures from Prism? 4. In my opinion, the authors overstate the conclusions associated with the data analysis. With found data and observational data one must be careful to avoid the trap of correlation versus causation. Many of the conclusions are obvious and not a result of the analysis but seems to be prior views on climate change. I don't see where the support is for the statement that benthic metrics are vulnerable to climate change. Is this based on 1 significant correlation?
Urquhart	A. The report under review represents an important and substantial effort. The effort should be applauded. It demonstrates how difficult it is to extract meaningful findings from data gathered for another purpose. The report ignores its potential reader. More detail on that below. B. Virtually every comment made by the other reviewers is valid and supported by this reviewer. In particular Eric Smith's comments about the distortion of significance levels by the application of thousands of statistical tests identifies an important limitation in the results. C. The main conclusion from this study/report should be that data relevant to the impact of global change on bioassessment is practically unavailable, especially on the spatial scale of the entire US. If the relevant personnel at EPA agree with this evaluation, and the need for such data is judged to be important, then EPA needs to initiate a program to gather the relevant data on a national scale, and with consistently applied collection and evaluation protocols. By starting soon, the needed calibration data should start to become available within a decade. D. PRESENTATION: Look at how EPA has presented its REPORT ON THE ENVIRONMENT (ROE). (http://www.epa.gov/roe/). It has a solid report as a large pdf file, but with web access to extensive supporting data, as well as briefer summaries for nontechnical audiences. This reviewer suggests the ROE material and presentation should serve as a model for how this report and its supporting information could be presented. The important point here is that the rewrite should consider the potential reader of the document and supporting information. Organize to communicate. Explain to the reader how the material, especially what was in the appendices in the draft, is organized and to which points it is related. One member of the target audience made observations about the sort of information which would be of interest to people in positions such as hers.

1. Based on your knowledge of bioassessment/biomonitoring programs, biological indicators, and climate change effects, please comment on the report with respect to:

a. Providing sufficient technical evidence to support programmatic modifications to address climate change effects; what additional steps, data, or analyses would improve the evidence?

Reviewer	Comments
Carlisle	The conceptual model depicted in Figure A-1 serves as a useful framework for organizing the evidence provided in the Study. The stated objective of the Study is to evaluate whether "components" of bioassessment programs are influenced by climate change, although most of the effort appears to be focused on biological indicators. As illustrated in Figure A-1, the effects of climate change on stream ecosystems begin with two driving factors: air temperature and precipitation. These factors in turn influence the thermal and hydrological regime of streams, which influence biota directly and indirectly. As evidence for the Study's first major conclusion (Page xx)--that biological indicators are responsive to climate change--the authors cite associations between the abundance & richness of cold-water taxa and annual air temperature. For this evidence to be convincing, the following causal linkages must be established (per Figure A-1): a. air temperature and/or precipitation varied over the time period studied, b. variation in the thermal and/or hydrological regimes of each stream is associated with variation in air temperature and/or precipitation, c. variation in biological indicators are associated with *and specific to* variation in hydrological or temperature regimes. In general, the Study demonstrated that in most of the case studies there was variation--and in many cases monotonic trends--in precipitation or air temperature over the focal time periods (item a above). Unfortunately, the Study only appears to have examined climatic data over the focal time periods (1970s onward) rather than the longer record of available data, which would have provided critical information about the historic context of climatic variability during the focal time period. However, analysis of longer-term and more local (than PRISM) hydro-climatic data would likely have revealed important insights. For example, the Study reports no trend in mean annual precipitation (A-5) and an increasing trend in mean annual temperature (Figure A-3) in the Utah mountains since 1970. Annual precipitation data since 1979 is available from a SNOTEL site in a watershed adjacent to the Weber, Utah site (**Attachment 1**). These data corroborate the finding of no trend, but also show that during the focal study period (1985-2005) the Weber River basin experienced extremes of annual precipitation--having the driest and 2nd wettest years in the last 30+ years. The last 30 years of precipitation are put in a longer-term context by examining streamflow data from the Weber River, just upstream of the site examined in this Study, and having over 100 years of record (**Attachment 2**). Annual streamflow data reveals that runoff in the Weber basin has been variable about a relatively stable mean since the 1940s, but

was much higher before that period--indicating a drier climate over the last 6 decades relative to earlier time periods. Importantly, the runoff observed during the study period (1985-2005) included one year with the highest runoff in six decades, as well as several years of rather typical low runoff.

Examination of local, long-term temperature data also provides new insights. Daily maximum temperature data were available for Coalville, Ut (about 10 km downstream of the Weber River site) from 1930. These data (**Attachment 3**) show that the observed trend in mean annual air temperature based on PRISM data (Figure A-3) through the 1980s and 1990s was well within the range of temperatures encountered in previous decades, which suggests that mean annual temperatures during the focal study period were not necessarily anomalous. However, examination of local (Coalville, Ut) mean *summer* air temperatures (**Attachment 4**)--which was unfortunately not examined in the Study--show a dramatic short-term increasing trend during the study focal period. These data show that the summertime air temperatures encountered during the focal study period were among the highest and lowest recorded in the last 80 years. Finally, since 1980, it is apparent that warmer air temperatures were associated with lower streamflows (**Attachment 5**), and in general colder years had higher streamflows. This relationship implies that it is difficult to separate the effects of purported increases in air temperature from the influence of streamflows.

In summary, examination of historic hydro-climatic data from local sources revealed that at least one site, the Weber River in Utah, experienced historically dramatic fluctuations in climatic conditions--particularly precipitation and summertime air temperature--during the focal time period of biological monitoring. It is regrettable that the Study did not provide this type of context for the other sites studied. This lack of historical context may not be a critical weakness of the Study, but it nevertheless leaves its readers without an appreciation for the climatic conditions encountered at the study sites, and it leaves the Study vulnerable to criticism that its operational definition of “climate” is limited to the last 30 years. Also unfortunately, trends in average summer air temperatures--which are most likely to cause changes in stream water temperatures--were apparently not examined in the study.

A more critical weakness of the study is its lack of evidence that stream temperature or hydrologic regimes (per Figure A-1) were associated with climatic variability (item b above). Continuous stream temperature data are notoriously scarce, especially in relatively undisturbed watersheds. The Study authors apparently examined stream temperature trends at a few USGS monitoring sites (Table A-1), but failed to adequately report the findings. Rather than R-squared values--which are not informative in trend evaluation--the Study should have reported estimates and associated uncertainty (i.e., confidence limits) of the change rates and number of years and periods of data for each site. It is also apparent that many of the sites had substantial human influences--which undoubtedly confound climate-associated trends. As a result, the stream temperature trend analysis performed in the study is weak evidence that climate-related changes in stream temperature have occurred over the course of the focus study period. The Study cites several recent reports of similar findings (e.g., Kaushal et al., 2010), and there are other studies showing relationships between air temperature and stream temperature, but these merely strengthen the *hypothesis* that stream temperatures ***may*** have increased at the study sites.

It is informative to examine the plausibility that an increase in average summer temperatures could have caused a trend in stream water temperature. Again relevant to the Weber, Utah site:

1. If we take table A-1 at face value, an average of the intermountain west streams suggests that average summer water temperatures could have increased roughly 1 degree every 10 years. This would suggest a potential for a 2 degree increase in mean summer water temperatures at the Weber River site from 1985-2005.

2. **Attachment 4** shows a 0.2 degree (F) increase / year in mean summer air temperatures from 1985-2005, and hence a 1.5 degree C total increase over that period.

3. Is it plausible that 1.5 degree warming of summer mean air temperatures is of sufficient magnitude to increase mean summer water temperature? Previous studies (Pilgrim et al.,1998, Wehrly et al.,2009) developed empirical relationships between air and stream water temperatures, which found that at an annual time scale, a one degree increase in air temperature is associated with a roughly 1 degree increase in water temperature. Given this generality, it seems probable that mean summer stream temperatures increased from 1985-2005 by at least one degree, which seems modest but is remarkably similar to the average magnitude of temperature increases actually observed in the western streams within Table A-1.

4. In conclusion, given observed trends in air (especially summer) temperatures, trends in water temperature at other streams, and empirical associations between air and water temperature, it is certainly plausible that the mean summer temperature at the Weber, Utah site increased 1-2 degrees from 1985-2005.

Had the authors presented the evidence in a way similar to what I just outlined, readers would probably have more confidence in that evidence. As currently written, the readers must make several large leaps and assumptions to believe that observed air temperatures could plausibly alter stream thermal regimes.

The most glaring omission of the Study is its lack of hydrological analysis, despite the fact that hydrologic regime change is a major pathway of climate change effects on stream biota (Figure A-1). Many of the study sites have USGS stream gaging data at or near the biological sampling locations (Appendix C), many with data records of several decades. In addition, many gages with long-term records can be found in the same regions as the study sites. It is therefore quite shocking that trends in streamflow were not examined at the study sites. For example, just upstream (~5 miles) of the Weber, Utah station, there exists a gage with >100 years of daily flow record from a relatively undeveloped watershed. Rudimentary trend analysis of these data reveal two important facts. First, as indicated by the runoff (precipitation) data discussed earlier, annual high flows declined rapidly until the 1940s, but have remained relatively unchanged (average conditions) since then. However, within the period of study at this site, abnormally high flows occurred in one year, but most years were below the long-term average. Second, monthly mean flows show gradual declining trends during the winter months over the

	last century (results not shown), which indicates declining groundwater recharge--possibly due to human activities or long-term natural processes. Of particular interest would have been a trend analysis of the hydrological characteristics discussed in Appendix J, which the Study considered to be ecologically important. In summary, the Study provides virtually no long- or short-term hydrological context at the study sites, and therefore offers no evidence of the connections between climate-streamflow-biota as illustrated in Figure A-1. In the absence of evidence about trends in the physical characteristics of the study sites, the Study relies primarily on an association between certain biological community attributes and mean annual air temperature. This evidence requires much greater scrutiny because the intermediate causal linkage (i.e., stream temperature or flow) is weak. The Study undertook a major and laudable task of developing temperature optima for many macroinvertebrate taxa. This contribution to science is potentially significant. But like many such empirically derived optima, there is a risk that factors other than stream temperature influenced the analysis. Note that the optima were developed using data from many sites, including ones with significant human influences. As a result, there are possibly confounding gradients of human impact in the data (a full treatment of this issue is given in Yuan 2007, and should have been considered in this Study). As a consequence of this problem, the optima should have been independently evaluated--that is, by evaluating how well biota-inferred water temperature predicts actual measured water temperature. This was apparently attempted, as shown in Figure A-16, but it is unclear whether these data were independent and the temperature data appear to be instantaneous, which is not really useful for this purpose. The lack of optima validation is especially critical because the major study conclusions rest on temperature-related changes in the biological communities. The critical reader is therefore left without confidence that the observed changes in the biological communities were in fact related to air temperature and not other factors. For example, the Weber, Utah site showed a strong association between “cold-water” taxa and mean annual air temperatures, but there are other plausible reasons why EPT (and possibly cold-water) taxa would decline at that site. The site lies downstream of several thousand acres of irrigated pasture land and two towns (including a wastewater treatment plant). Summer air temperatures influence irrigation diversions, which can reduce streamflow, increase water temperature, and change water chemistry. In fact, there appears to have been a strong trend in chloride concentrations at that site (Figure 2-30) which could alone account for losses of EPT taxa. To their credit, the authors acknowledge the possibility of other factors, but should have done more to control for them—such as validating whether the biological response is indeed specific to changes in stream temperature. In summary, the evidence that bioassessment indicators are effected by climate change is essentially limited to correlations between mean annual air temperature and the relative abundances of several macroinvertebrate taxa that are known to be sensitive to many human influences--not just increased water temperature. Because the purported biological response to climate is not *specific* to water temperature (item c above); and because there is no direct evidence that variability in climatic variables was related to stream temperature or stream flow, the overall evidence in support of this primary finding is weak.

	Granted, the authors acknowledge the possibility of confounding factors and show some of these data where available. But the authors also apparently overlooked key hydro-climatic information and analysis that could have strengthened the case for causality and provided critical climatic context for each site--as I attempted to illustrate for the Weber, Utah study above. Several possible corrective measures have been suggested above. First, the 30-year study period must be placed into a longer-term context using relevant climate data for a much longer time period. At a bare minimum the authors should define early in the report what is meant by "climate change" in the context of this study. For example, do the authors define climate change as annual variation in mean annual temperature and precipitation over the last 30 years? Or, do the authors define climate change from a longer-term perspective? Second, trends in stream temperature must be reported more clearly and convincingly, which includes limiting the analysis to undisturbed sites. Third, trends in streamflow characteristics (both short and long term) must be considered as a potential response to change in climate. Even if there is no evidence of trends in precipitation, it is known that trends in temperature can still influence patterns of runoff--especially with respect to snowmelt (Yarnell et al., 2010). The authors cannot forward a credible analysis of climate-related effects on stream biota without explicit examination of climate-related effects on streamflow characteristics. Finally, the authors should seek data sets (NAWQA has several datasets of paired macroinvertebrate data and long-term water temperature measurements) and independently validate the temperature optima developed in this study. Even if all other attempts to strengthen evidence are not fruitful, increased confidence that the temperature optima are specific to temperature and not confounded by other factors, would greatly strengthen the Study's primary conclusion.
Fennessy	The authors have used a variety of analytical methods to address many of the key questions related to bioassessment and climate change effects (question 1a). Table 2-1 is helpful in summarizing statistical approaches. However, in many instances the body of the report does not include enough detail on the methods used or the results obtained to make a thorough evaluation possible of the analyses (question 1b). If the goal of this report is to provide sufficient technical evidence to support programmatic modifications to address climate change effects then the evidence should be presented more clearly and comprehensively within the body of the text itself. Sorting through the appendices to find relevant data is essentially impossible. Many of the findings presented are based on data from only a small number of reference sites in one or two states; it is unclear why more data could not be used. In addition, the authors of the report use similar types of data sets from the different states, but different analytical methods are used to analyze them (e.g., ANOVA, multivariate techniques, etc.). This makes comparisons of the findings difficult. While the rationale for how the data were treated is provided in the appendices, this should be included in the report. Section 2-3 does a much better job than earlier sections of laying out the logic of the approach used in the data analysis presented in this section, greatly improving the flow of logic and clarity of this section of the report. Finally, the report would benefit from a quantitative summary of findings across sites and states to make clear what the 'weight of evidence' is with respect to the study questions. As it stands now it is difficult to synthesize the overall message of the report.

The comments below address other issues noted in Section 2 of the report. This is a large section and, as it stands, is difficult to interpret.

Figures and tables should be presented in a consistent format with complete legends (many are incomplete) and with obvious labels on the figures and complete column headings on the tables. They should also be formatted so that they will reproduce in black and white (this is not the case now). As an example, Table 2-2 shows important results of the study, and would benefit from better column headings, such as “years” instead of “groups” (column 2), and ‘Cold taxa’ and ‘Warm taxa’. It would also help the clarity of Table 2-2 if a brief explanation were provided of which taxa groups made up the rest of the community since the cold and warm taxa make up a relatively small % of the individuals in many years (Table 2-2 shows same data as Figures 2-1 and 2-2). In Figure 2-3 – 2-5, why are data shown for only one of the Maine sites? This should be explained; if it isn’t the authors risk losing credibility since some data are omitted without explanation. For these (and other) figures, the results of the statistical tests should be presented in the legends.

It is important that care be taken in making recommendations based on the results presented in many of these graphs (e.g., Figure 2-6, 2-7 and 2-8) because, while they show trends in the data, there is no statistical significance in most of these data comparisons. The comparisons of Utah sites by ecoregion, elevation, and stream order are an excellent way to stratify the data, but again, the results of any statistical tests should be included in the figure legends (e.g., Figure 2-6, 2-7 and 2-8). The authors should be clear in their conclusions that while trends in the data are sometimes seen, in some cases there are conflicting trends, and/or the trends are not significant. Finally, no evidence is provided that substantiates the link between air temperature and water temperature. Can data be provided to link the increases in air temperature to water temperature?

At many points in the text, there is not enough information provided to allow interpretation of the results. For example, the specific Indicators of Hydrologic Alteration (IHA) used in the analysis (line 479) are not listed, and on line 482 states that several low flow parameters performed well - performed will with respect to what? There are many other examples, but the text needs a thorough proof read to ensure its meaning is clear to the reader. There are also very few citations in the text relating these results to other studies.

The introduction to the report states several times that four states are the focus of the study, despite the fact that the data from Ohio, and its analysis, are nearly absent. This should either be rectified or explained.

Overall, the authors have done a good job in elucidating factors to consider when applying insights from the four states in this pilot study to other state programs, and have made helpful suggestions on which activities should be a priority in dealing with the response of communities to climate change.

For state programs, it might be helpful to link the expected responses described here to specific predictions on how climate will change in different regions in the U.S. In this way, states will be able to more finely tailor how they will use the insights derived from

	this report.
Smith	Yes and no. The evidence of change in metrics due to climate change is not clear. First, the correlation between average air temperature from PRISM and benthic metrics is sometimes moderate in size but there is not a consistency to the correlations. Patterns that occur at one station do not seem to also occur at other stations (Table 2-6, for example). Second, the project is based on data collected for a different purpose and confounding factors limit the analysis. It could be, for example, that the correlation with temperature is due to a connection with elevation (if one replaces space for time) or a factor that also varies with time. Other factors, for example, acid rain in the east, might also confound the analysis. It might have been a stronger study if the authors first reviewed the literature, postulated changes in benthic metrics then observed the changes. There seems to be a general lack of checks on assumptions. The report does point out some inadequacies with current designs that while known for years have not changed. Some of these inadequacies have been known for a while (the need to sample some sites repeatedly over time in order to assess trends).
Stevenson	This report and other scientific evidence cited in the report clearly provide sufficient technical evidence to support modifications in bioassessment programs to address climate change. We know that climate will be changing by several degrees centigrade with changes in drought and flood disturbances. There is considerable evidence in the report that invertebrate metrics respond to temperature, those metrics will respond to the projected temperature increases expected with climate change, and changes in metrics resulting from climate change would be sufficient that assessments of biological condition could change from unimpacted to impacted without changes in land use. In addition to the report, other research shows temperature has great effects on invertebrate physiology, which likely varies greatly among taxa. Thus, invertebrate metrics could respond to changes in temperature because invertebrate species composition is likely to change. There is evidence in the report at a couple sites in four states' data that historic temperature change that could be related to climate change may have affected invertebrate species composition and metrics. The evidence that inverts have responded to hydrologic alterations as projected by climate change is weaker than the evidence for temperature. While evidence in the report could be made stronger with better explanation, more thorough data analyses, and greater integration of scientific literature on invertebrates response to climate change variables, responsible resource managers should take action to account for climate change effects on biomonitoring programs because multiple lines of evidence show a reasonable probability that climate change effects on bioassessments will complicate bioassessment of local management of land use and resulting pollutants and habitat alterations. In conclusion, the science in the report shows that climate change could affect invertebrate assessments in streams; and importance of programmatic modifications to state programs is evident from the relatively poor data available to evaluate effects of climate in the four states' data. Overall, this is good report. In particular, the data analysis is extensive and supports

	multiple lines of evidence that climate change will affect bioassessment results and management. I am not convinced that the analyses were as thorough as they could be. Although many analyses were run, it seems that different statistics and variables were used with data from different locations. Analyses should be added to fill the gaps and present similar analyses, using similar variables, and for each region. The proposed adaptations for management to account for climate change effects will be useful, but they will not be sufficient to address the full range of considerations necessary for managing local effects on watersheds that will be exposed to climate change. A broader conceptual framework is needed to more completely address the needs for management adaptations and adaptations to fill those needs. Future work should: build a policy-management-assessment framework to guide the development of tools (see below); expand on the lines of analyses already conducted to have more equal and uniform treatment of all datasets; gather more data and evaluate direct stressor effects on invertebrates and how effects vary among groups, which will build on the traits database; and expand the invertebrate traits and stressors considered when thinking about effects of climate change (e.g. droughts and flood stress as well as temperature); include more rigorous cause-effect analysis of invertebrate responses to stressors, such as temperature, drought, and flood stress. In analyses of responses, consider the temporal lags in response of differing metrics, relative to stressor. The report should include rigorous, yet broadly applicable, conceptual and analytical tools for resource managers. **Specific comments related to the report and Question 1a:** There is evidence that effects of temperature change on invertebrates as projected for climate change may have been observed in analyses at a couple sites in four states' data. The evidence that inverts have responded to hydrology alterations as projected by climate change are weaker than the evidence for temperature.
Urquhart	This report shows how difficult it is to document climate change effects with data which was not gathered with that objective in mind. No additional analyses seem fruitful at this point. More relevant data is needed. The really important conclusion and only soundly defensible approach is to follow recommendation #2 (on page 7-1). (Unfortunately, the conclusions have no line numbers for reference.) This really important recommendation should be much more detailed. In fact, it should be the main point of this report!

b. The main factors to consider in order to transfer or apply insights from the four states in this pilot study to other state programs

Reviewer	Comments
Carlisle	Based on the limitations of the current draft, I could not recommend transferring insights from this study to other areas.

Fennessy	See 1.a. above.
Smith	It is a difficult question to address as the report requires a major rewrite. It is difficult to read and to connect the conclusions with the evidence. I do not see strong evidence for the insights that are presented and many are a priori insights. Obviously there is a need to measure water temperature and to design a temperature exposure metric. In general, it would be wise to keep the raw data, not just the metrics. This way, future metrics can be developed if needed.
Stevenson	The main factors to consider for transfer and application of insights from the four states to other state programs should be: 1) a framework for how climate change effects on ecological condition can be incorporated into management of state waters; 2) a conceptual model for how current management problems, which I will refer to as human activities related to land use (including generation of point and non-point source pollutants and physical habitat alterations), and climate change independently and interactively affect proximate environmental factors (e.g., temperature, droughts, floods, nutrients, sediments, and dissolved oxygen) regulating taxonomic composition and metrics of benthic invertebrates; 3) research program designs to determine predicted effects of land use and climate change effects on ecological condition; 4) monitoring program designs to assess independent and interactive effects of land use and climate change on ecological conditions; and 5) analytical tools for determination (3) and monitoring (4) and causal analysis of land use and climate change effects on ecological conditions. These elements are spelled out in a couple chapters that I lead in Barbour et al. (2004), i.e., Stevenson et al. (2004a and 2004b). From this list, 1 is lacking, 2 is incomplete, and 3-5 are developed but are largely focused on temperature. If the main factors to consider for transferring and applying insights are defined as the options for proposed programmatic modifications, then the proposed programmatic modifications are appropriate, but more analysis and testing are needed to determine which programmatic changes are appropriate under different funding availability and management goals. In a few cases, the emphasis on a specific subset of programmatic comments overemphasizes their importance within the full set of possible programmatic modifications. For example, the importance of having a long-term reference monitoring network (Sentinel Monitoring Network) is emphasized throughout the report, but not until Section 4 is the balance between long-term monitoring reference and non-reference sites discussed. There are challenges for transferring results, such as predicting which streams will be most sensitive to climate change. However, it is clear that many of the recommended adaptive management responses are possible, can be used by all states, and are important to adopt. A sound conceptual foundation in ecological first principles will increase transferability of results across regions and time. Rigorously develop the conceptual model of temperature effects on bioassessment programs, but also include the other pollutants and habitat alterations that will occur with climate change and how they interact with land use change. Although this information seems to be present in the report, it is not

	organized in a way that it will be easy for policymakers and managers to understand. Elaboration of conceptual framework for management land use change and climate change as well as a conceptual model for climate change effects on invertebrate stressors and invertebrate responses to those stressors would help organize the report, future research/monitoring, and implementation of results in management and policy. **Specific comments related to the report and Question 1b:** 2.197-222. Developing invertebrate traits based on their sensitivity and tolerance to temperature is certainly a good first step for understanding effects of climate change on invertebrates, bioassessment programs, etc. However, at this point in the report, I'm not sure what to expect the effects of elevated temperature, floods, and drought to be on invertebrates. For example, is temperature a resource stressor (*sensu* Stevenson and Sabater 2010) and really has no direct negative effect on invertebrates unless it affects dissolved oxygen or another parameter? Or do high temperatures have direct negative effects on invertebrates because they denature enzymes? Having a thorough conceptual model explained at the first principles of temperature and other stressor effects on invertebrates early in the report shows how the analytical approach is organized, enables placing results in context, and help readers remember, understand, and believe results. This is particularly true when sometimes you see effects/relationships and sometimes you don't. Variability in results is to be expected and should not compromise the certainty of the findings. Such an approach, with clear *a priori* hypotheses, will also mute criticisms of repeated hypothesis testing and outline ways to control and explain attained significance (p values) with multiple hypothesis tests.
Urquhart	Frankly, none. The difference in states' objectives and the project's objectives are so great that state data will be of little assistance in meeting the study's objective. Once some solid data is available when recommendation 2 has been implemented, some changes in multi-metric indices, as per recommendation #1 (page 7-1) could be recommended.

2. Has EPA pulled out the most important findings in the Summary for Managers and Policymakers (SMP) from the technical findings? If not, what findings do you suggest should be highlighted?

Reviewer	Comments
Carlisle	The Summary for Managers and Policy Makers (SMPM) section is too long and contains many "findings" that are not really uniquely attributable to this study. The first finding is far too strongly worded, as is the information presented as evidence. For example, this study did not really examine climate change. Rather, it examined variability in air temperatures and precipitation over the last 20-25 years. This finding should be more clearly stated as: "Some metrics used in indicators of biological integrity are associated with climatic variability." The second finding is poorly worded and difficult to follow. The writing needs to be more precise. I am very familiar with RIVPACS models but it took several readings of

	this section and I'm still not entirely clear what the point is. RIVPACS models derive expected probabilities of capture for each taxon based on relationships between environmental factors (such as average climatic conditions) and biotic structure within a population of reference sites over a fixed time period. Hence, estimates of E for both disturbed and undisturbed sites will always be based on the original time period over which the model was developed. In the future, if the climate changes as expected, the original RIVPACS model would in theory detect it as "drift" in the condition of reference sites sampled in the future. Specifically, the observed and expected communities will diverge at reference sites presumably due to climate--since we assume there are no other human influences contributing to biological changes at reference sites. So, I am confused by the recommendation that RIVPACS models be continually updated as a means to, presumably, detect changes due to climate. The third finding is nothing new to science or management and therefore not uniquely contributed by this study. The "evidence" provided for this finding does not actually present any results from this study. I suggest dropping this message. The fourth finding is poorly written--but it also is not a new finding. Again, the evidence presented contains no new information from this study. Given the finding, I would expect to have seen evidence that, through time, the biological composition of some reference sites have in fact become more similar to non-reference sites, and this change is entirely due to climate. This message should be dropped. The fifth finding, unlike the previous three, provides some evidence from actual results of this study. Unfortunately, the idea that "place matters" is worn out. Isn't this a common assumption anyway? How is this a new idea? Either spin this finding in a way that is clearly new, or drop it entirely. Findings 6-8 are also common knowledge and don't seem to present evidence that is uniquely provided by this study. It is unclear how this study contributed to these generalizations. Finding nine seems to be based on "simulated" data as in Figure 4-2, where progressively more taxa are excluded (i.e., eliminated) from metric calculations. This is weak evidence. Again, I was expecting to see actual data showing a decline in biological integrity through time at a reference site. If this was actually observed in the analysis, it is unfortunately not clearly presented as evidence. Finding ten is interesting and potentially significant, but the authors provide no evidence that this phenomenon actually occurred at the sites studied. Specifically, did the authors observe changes in the biological conditions at a site that are sufficient to cause reconsideration of its designated or attainable use? At best, the authors found subtle changes in community composition that were associated with annual variation in temperature or precipitation. And, general knowledge is given as evidence for this finding. In summary, most of the findings, as currently written, read more like a text book than solid and specific discoveries from the current study and their implications to biological assessments. I suggest reducing this section to a fraction of its current length, and

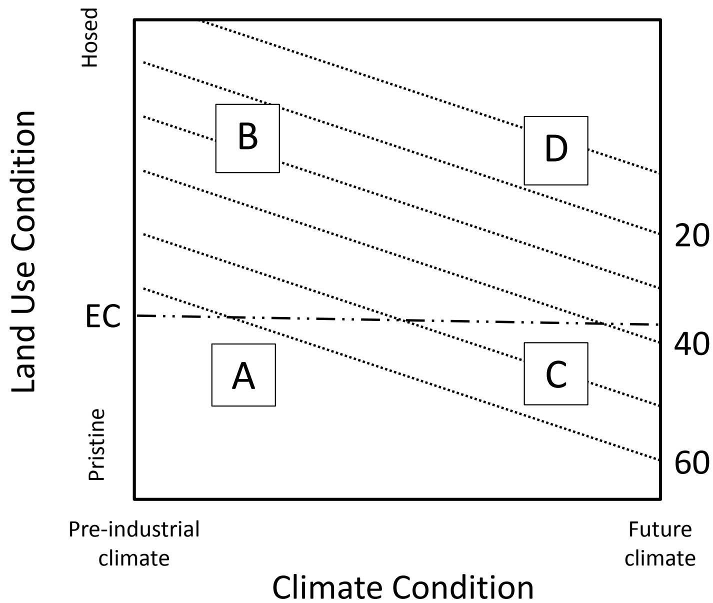
	focusing the topics on a small number of pointed findings that clearly stem from the research conducted.
Fennessy	The Summary covers the key findings of the report and is a good summary of the contents of the report. However, the conclusions as presented are much more certain than the data support. For example, many of the conclusions reached about shifts in taxonomic composition as a function of air temperatures appear to be based on five metrics that showed significant differences at two sites in UT (Weber and Virgin, lines 631-634). This isn't enough evidence to conclusively support the conclusions reached, the summary must reflect the ambiguity in the data. For example, lines 696 – 699 in the discussion of ME data, where there was a lack of association between ecological traits and community response. The results presented in many sections are either not significant (line 783 for example) or are conflicting, which would seem to limit the ability of the authors to make strong conclusions. Generally, there is a tendency in the report to over-interpret the results.
Smith	The summary suffers from a desire to find a significant conclusion when there does not seem to be one. I think it is bad science to run hundreds of tests then draw conclusions from the statistically significant ones, especially when there are not clear patterns. Again, I think more work is required. I especially did not like the frequent use of the term "vulnerability" as this is both a general term as well as a term from the climate change literature. I am not sure what it means to say the multimetrics are vulnerable to climate change. In the climate change literature vulnerability involves exposure, sensitivity and adaptive capacity so should the authors have evaluated all three of the components? Do you really mean metrics are sensitive to climate change or that climate change will confound conclusions from studies that involve reference metrics? One question that is not really addressed is "do decisions about the quality of the environment change"? If the organisms are responding in a negative way then one would expect to see false positives (declare impact when there is none) increase. It is not clear that this is the case.
Stevenson	Yes, for the findings resulting from the analyses conducted in this report. Evaluating all possible effects of climate change and interactive climate and land use change in one report is probably not practical. Clearly defining what is done in this report and what has not been done thoroughly should be clearly spelled out, so the later can be addressed in future work. Again, one practical way of doing this is starting with a management framework that includes independent assessments of climate and land use change effects and a conceptual model of how they affect ecological condition will provide the conceptual landscape in which the path of this report can be mapped and better appreciated. This should be done in the introduction of the report. An important technical finding that is not clearly explained is managers need to think about new ways for characterizing reference condition, diagnosing stressors, and developing water quality criteria with increasing effects of climate change on current bioassessment and management paradigms. This calls for a policy-management-assessment framework that explicitly includes climate change. Again, the many ways

	climate change could fit into the policy-management-assessment framework should be treated more clearly. Although elements of this policy-management-assessment framework are discussed throughout the document, they are not pulled together. **Specific comments related to a policy-management-assessment framework that includes climate change:** An important technical finding is managers need to think about new ways for characterizing reference condition, diagnosing stressors, and developing water quality criteria. Throughout the document, climate change is ambiguously treated as something either to treat separately from reference condition or to account for with reference condition. In other words, is a climate altered system still in reference condition? Of course not. If the USEPA is treating C emissions as a contaminant causing climate change impacts, then one of the impacts (in addition to human health and direct risk to well being) is degradation of ecological condition, including both water quantity and quality as well as biological condition. The report needs to deal more explicitly with how climate change could be treated relative to management goals. The challenge with treating climate change as just another set of stressors that humans cause is that management of climate change is largely out of the scope of local managers. Mercury contamination from fossil fuel emissions is similarly a transboundary and global problem calling for a modified reference approach. Of course there is the “act locally and think globally” philosophy in which local activities are hoped to have global effects on climate change, but managers have much more direct impact on local land use and best management practices than climate change stressors. This potential for management must be recognized explicitly in the policy-management-assessment framework. Potential for local management with the inevitability of climate change will require a “moving target” for management, or a moving management goal (aka expected condition (*sensu* Stevenson et al., 2004), Figure 1). Climate change impacts, locally manageable human activities and stressors, and natural variability represent three suites of factors that could be used in models, as has been described in the report. But they were not emphasized sufficiently as part of the management adaptation strategy. Climate change impacts could be treated, relative to local management goals, as natural variability is treated. In the U.S., we are usually trying to manage waters as minimally disturbed. If we can only control local land use[1] with best management practices (point and local non-point sources of contaminants), then effects of climate change can be treated as natural factors, i.e., something we need to account for when establishing our expected condition for management, but not something that we can really manage. That does not mean that we should not assess effects of climate change, or consider climate change and land use change interactive effects on ecological response, just as we don’t neglect interactive

[1] By land use, I mean human activities in general, in urban and rural, agricultural and forestry, built and unbuilt conditions.

	effects of land use change and natural factors. The key distinction here is isolating climate change management and local management of contaminants and habitat alterations that fall under the policies of the Clean Water Act (CWA). We should, therefore, assess both deviation in observed condition from the historic reference condition and the current[2] expected condition (I don't want to start calling this the current reference condition because that generates other debates). We should assess condition at a site as the deviation of observed minus expected conditions (Stevenson et al., 2004a, 2004b, and others), both historic reference condition (when possible) and current expected condition. This is very similar to the concept of tiered aquatic life uses, where we recognize natural as the anchor to the generalized stressor (human disturbance) gradient, but we don't expect to manage for natural conditions. So how do we define these two expected conditions, historic reference condition and current expected condition. We generally think of the structure and function of streams being ultimately regulated by climate and geological factors that regulate hydrology, soils, and land use. Naturally varying climatic and geological factors are often captured in "predictive models" to define expected species or metric values at a site. So current expected condition at a site is often determined as a function of elevation, latitude, and longitude (often proxies for climate), precipitation, mean annual temperature, soil permeability, groundwater loading, etc. We can measure these things now, and use models as we currently use to predict current expected condition for a site. Historic reference condition can either be documented now or modeled later, based on what we know about climate change for different regions and responses of stream structure and function to those climate changes. Key point: we should explicitly relate expected condition to the deviation between the current expected condition (with current being some specific time either now or in the future) and reference condition that was defined today (or sometime in the past). So what is the alternative? Can we have a management goal for protecting systems in the minimally disturbed condition as occurs in 2011 climate condition (which is already somewhat altered, to an extent that we cannot yet assess without future data)? No. We'd like to, but from a local perspective we cannot. So distinguishing effects of naturally occurring climate and geological conditions, from anthropogenic climate change effects, and both from land use effects is essential for management. Documenting and forecasting climate change effects will be critical for arguments to manage climate change at other policy scales and for adapting to changes in our aquatic resources that we cannot locally manage.

[2] Current here means the condition at the time of assessment, now or in the future.

	 Figure 1. Theoretical relationship between expected condition % EPT (EC) assuming climate change, % EPT, land use condition, and climate condition. Isopleths of % EPT (dotted lines) vary from 10 to 60 across gradients of land use condition ranging from natural (pristine) to highly impacted (hosed) and gradients of climate condition ranging low temperature and extreme event disturbance (pre-industrial climate) to high temperature and extreme event disturbance (future climate). % EPT in high quality waters (i.e. meeting expected condition) is predicted to change with climate change from greater than 60 to less than 40, assuming no change in land use. Minimally disturbed expected condition without climate change (today's reference condition, A) is greater than 60 % EPT, whereas in the future, the same site with minimal land use disturbance could have a % EPT near 50 (C). A disturbed site (B) today has a %EPT around 28, whereas in the future with climate change effects, it could have a % EPT close to 10. Contact me for more explanation if you are interested.
Urquhart	Both the Executive Summary and the Summary for Managers and Policymakers represent the main report well.

3. Are the analytical methods used appropriate for the questions being asked and the available data? Are the strengths and weaknesses of the methods used in the report fully described and accounted for?

Reviewer	Comments
Carlisle	The analytical approaches and their respective methods are voluminous, scattered throughout the report, and often times inadequately reported. As a result, it is extremely difficult to address this question. For example, I could not locate a single, detailed description of the methods used to generate Figures A13-A16. The accompanying text

	(Page 4-4) provided only general descriptions, then seemed to point me to Appendix D for further details. I could not readily locate any additional relevant information there. I frequently encountered this problem, which suggests an underlying issue with organization of the document. In addition the hydro-climatic evidence outlined above, I also examined the RIVPACS analysis section, and found it to be unfulfilling. Figure 3-9 suggests that for two sites on the Colorado Plateau that were presumably reference (and therefore used to develop models for E), there were many fewer expected taxa during cold years than hot years--hence O/E values are lower in cold than in hot years. The authors attribute this to the influence of climate on the models, but there is no evidence to indicate this. To make this case, the authors should have shown the list of expected taxa for each site (which would be essentially the same every year), and then show which taxa were not observed in each year. Presumably, there were several “missing” taxa during cold years--if those taxa are warm-water taxa, then the authors have evidence to support their claim that these patterns resulted from the fact that the RIVPACS models were developed during abnormally hot years, and therefore “learned” to expect warm-water taxa wherever and whenever they were applied. I found a list of cold- and warm-water taxa (Table F4-1), but it doesn’t provide the information just described. I was unable to find which specific years were classified as hot vs cold, but examination of Figure F6-6 suggests there are not strong differences in EPT richness in hot vs cold years. In summary, the few analytical methods for which concise and detailed descriptions could be found appeared to be appropriate, but there was often a disconnect between the authors’ interpretations and the underlying evidence presented. In other words, the results presented—occasionally in great detail—were were often weak evidence, whereas results that could have provided stronger evidence were often absent entirely (as I argue with the RIVPACs results above).
Fennessy	Generally the analytical approaches seem sound, and the description presented on Table 2-2 is very helpful. However, a huge variety of approaches are used, with no rationale presented for why one method was used in one section and a different method in another. The choice of methods must be made clear (and while some explanation is given in the appendices, it should also be presented in the main body of the report). In some cases the results of an analysis are not presented completely, for example, data presentation is incomplete for the CCA and NMDS analyses. For example in Figure 2-15, information on the CCA axes (axis 1 and 2 at minimum) should be provided. A brief explanation of why the different approaches were used should also be presented - why was CCA used for Utah data and NMDS used for NC data (Figure 2-16)? Section 2-5 details ‘other sources of potential spatial confounding’ (not a very clear subheading, by the way). The issue of changing land use and land cover through time is an important one, and the authors have been thorough in addressing this issue in this and other sections of the report. However, in the analysis only land cover within a 1 km buffer of the reference sites was used to indicate anthropogenic influence. A rationale for this buffer distance should be provided. Is this a common approach in stream studies? Should an analysis of the upstream catchment area be used instead? The upstream

	catchment area would seem to have a more direct influence on a given stream segment. In some cases more explanation of the magnitude of climate change necessary to generate changes in biological indicators would be helpful. For instance, for the analysis presented in Figure 3-5, what change in air temperature would be needed to generate this shift in water temperatures, and so drive the biological response modeled here. And again, early in the report it states that data from four state programs are used in the report, however, very little attention is given to Ohio data in the subsequent analysis. While this is touched on in the appendix, it should be addressed in the introductory sections with some explanation of why this is the case.
Smith	The data come from existing studies that were not designed to assess climate change. Generally “found” data is difficult to analyze as it contains confounding factors, inadequate sample sizes for the effects of interest and irregular coverage. Found data often results in confirming obvious results and less obvious results tend to be inconsistent. Most of the data that are used in the examples is based on sampling programs inspired by the EMAP program. The sampling here is intended to focus on status not trends. It is therefore not surprising that trends associated with climate change are difficult to detect and assess. Correlation analysis using Pearson correlations require linearity and independence for the measure to be useful. The authors seem to use this measure without checking these assumptions. Most of the results simply confirm the obvious. Specifically, elevation, area and stream length are important. Climate change seems to be temperature. There is a lack of analysis related to hydrologic conditions and other expected changes due to climate change. The authors should be more explicit about what they mean by climate change. The analyses seem to give weight to results that are statistically significant. Given the number of tests that are made and the relatively low correlations that occur one has to wonder if any of the results are meaningful. There are quite a few confounding factors that make the results of the analyses rather difficult to interpret. For example, in the CCA analyses, shouldn’t elevation be accounted for – i.e., A partial analysis might be appropriate. It seems that the view is that climate change affects only the reference sites. It will obviously affect the stressed sites as well. Is the real question the extent to which the stressed sites will be detected (i.e. are error rates affected)? Thus the focus should be more on the change in the power of the reference condition approach. The focus is also on means. In particular, the main effect of climate change that is addressed is associated with a rise in temperature. What is likely to be important is the variability that will increase, especially initially. While it is useful to develop an idea of temperature optimum for different taxa, the focus

	is on the optimum not the tolerance (i.e. the variance in response). The tolerance is rather critical and can be used to address some of the questions of interest. For example, if the tolerance of most taxa is small, this would suggest that temperature change will have a strong effect. On the other hand, if the tolerance is quite large then there may be a minor effect. The writing is inconsistent and I found it difficult to follow some of the material without looking at different sections of the report. Graphs also show inconsistencies, perhaps due to the fact that some were intended to be in color rather than black and white. I think the authors should use standard practices for reporting information when testing hypotheses. In particular, sample sizes are inconsistently provided.
Stevenson	The analytical methods used were appropriate for the questions being asked and the available data, but as with many elements of the report, their presentation was so unpredictable that you often found the justification for an analysis at a different spot from the analysis. So it was often hard to follow. But with limited data come limited options for analysis. A more rigorous application of causal analysis (aka Beyers 1998) would have been valuable for when only one site was analyzed at a time would have been beneficial. The challenges of the data were clearly stated. Given limited long-term data and data at reference sites, the analyses were appropriate, but in some cases they pushed the bounds. The traits analysis should include a causal analysis element, which will be describe below.
Urquhart	Analytical methods are reasonable. Some details could be argued, but overall the approach is defensible. Strengths and weakness are generally acknowledged. An attempt has been made to account for weaknesses in the available data, but it has been gathered for other objectives, so it is not as relevant as the authors would like.

4. Does the main body of the report effectively capture the more detailed information presented in the appendices and if not, what are your recommendations for improving the report or appendices?

Reviewer	Comments
Carlisle	The main body is a maze of sections that seem to contain redundant and fragmented information. The document is very difficult to read and extract the relevant information from a single location. In order to get the “whole story” about each site, the reader has to sift through virtually every chapter and appendix. I believe the main body needs a drastic reorganization. Like it or not, this study is a collection of “stories” about specific places. The pieces of the story for each site, if concisely, clearly, and logically presented, provide the necessary evidence for the interpretations and conclusions. An organizational structure focusing on sites may be much more effective than the current approach. In my attempts to fully understand the

	information obtained and analyzed and the results for the Weber River site, I literally had to pour through every section of the document and extract tidbits. I would have preferred to read through a single section in which all the evidence was presented in an orderly way to tell the story about an individual site. The section on each site should be organized in a consistent manner. This is just a suggestion, but I must admit that the current organization didn't work for me.
Fennessy	The report would be clearer as a stand-alone document (i.e., without benefit of reading the appendices first), if a more complete introduction to the methods used and rationale for site selection were provided (for example, why were the data from some stations included and not others, pg 2-3). The final conclusions of the report rest on the choices that were made with respect to sites included and analytical methods used, so elucidating how they were made is crucial to interpreting the report's conclusions. This is also true for the manner in which data is presented. For instance, results based on Utah data, presented in section 2.2.1 (line 240) are given with little explanation of why these reference stations were used. A table of the different stations with the duration of data records (as shown in the appendix) would be a welcome addition to the report. As an editorial comment, this would also be easier to read if the names of the sites were provided with the station numbers in parentheses, instead of vice versa. Likewise, a more full explanation of the development and use of "scenario- metrics" is warranted (line 229). The report would also benefit from development and presentation of conceptual models of expected ecosystem response (what are the specific predictions about how different indicators within a MMI will respond, for example, these would be related to the questions posed on lines 125-132), and at some point in the report, a summary of findings across sites and states to make clear what the 'weight of evidence' is with respect to the study questions. For instance, at line 1016, the authors provide a comparison of trends in the data, but in other places in the report, the conclusions reached based on these seem to imply much more certainty in the results.
Smith	The appendices are a bit difficult to navigate and need better linkages to the main body of the text. I have commented on them separately below.
Stevenson	The main body of the report was not as effective as it could be. The organization of the whole report is awkward, but that could be my lack of understanding about who will read the report. The executive summary was too long without clear messages highlighted in bold and plain language. I really liked the summary for managers and policymakers, because of the finding/evidence/adaptation-response section in each. Chapters 2-5, however, which relied on reports of more detailed analyses in Appendices, did not have sufficient detail to understand without using the appendices, from a scientists' background, which often brought credibility of results into question. Chapter 2 is substantially more confusing than Chapters 3-5. In addition, there's considerable redundancy with the way it was presented. Although it's probably too late for this report, future reports should be organized in more of a chronological sequence and traditional scientific presentation. Readers can always

	skip to the good parts. It seems that this organization was used to put the good parts, results and interpretation, early and in a synthesized form. The problem with that organization for many readers is details of the rationale and related analyses are important for understanding, believing and remembering the analyses.
Urquhart	First, the appendices are a jumble of information, apparently written by different people. This is a fact: Four of the appendices are state specific, for defensible reasons, have different sections, emphases, data, and analyses. Thus it is difficult to summarize them accurately in the body of the report. The summary is generally sound, except for one perspective. The individual appendices convey much more uncertainty about conclusions than the body of the report does. Appendix H deals with temporal trends in Ohio, but there is no section 2.3.4 to discuss them, and no apparent explanation for its omission.

5. The *Freshwater Biological Traits Database* was pulled out as a standalone report. Please comment on the merit of this report on its own vs. as one of the appendices of the main report. What next steps would you suggest for this report and the database?

Reviewer	Comments
Carlisle	The traits database is potentially a very valuable product from this effort. The authors brought together trait information from a variety of sources and made that information much more accessible than is currently possible. In my opinion, it should not be included as an appendix to the report, mainly because the report already has too many appendices, among which the important trait information would be lost. The traits database and web portal should be a separate entity that can stand on its own. Besides, other than the addition of temperature-based traits from the current study, the trait database itself has far broader utility than studies in climate change. Rather, it's a tool that managers and researchers can use for a wide variety of environmental studies.
Fennessy	This will be very valuable as a separate document to many practitioners in the field, including those working in State and Tribal programs, so I would say that making it available as a stand-alone document/database is valuable. I imagine there are many who could make use of the database who will also want to reference the main report's findings, but since the database is available online and it is intended to be a 'living' database, it holds promise as a valuable research tool as the issues of climate change become more pronounced. This is particularly true as trait-based approaches to MMI development and use (i.e., the development of trait-based indicators) are increasingly important as an alternative to taxonomic based approaches. A database of this sort will facilitate such research efforts.
Smith	Having a traits repository is valuable. The only comment I have is that the traits are a mix of physical and non-physical. For example, temperature is included. If potential stressors are to be included, then the database should be expanded to include other information that is available such as nutrients, heavy metals, etc.

Stevenson	This is great. Traits databases are so important for developing the kinds of indicators that I mentioned above, which have really important applications in multimetric indices, stressor diagnosis, and forecasting ecological change. I recommend that a pilot system be developed, tested with states and other users, and then completed as soon as possible. It will be important to have rules for entering data and periodic syntheses to produce a recommended and "certified" set of traits and taxa. Another element of traits analysis that is important for use in stressor diagnosis, like temperature and climate change, is a causal analysis of the traits. Some variation of the propensity score analysis conducted by Yuan (2010) would help evaluate causal traits analysis. I have conducted analyses of diatom species traits using the same kind of approach and have found that the covariation in diatom species traits for N and P can be disentangled by developing traits for one variable while stratifying the data for another. For example, N and P traits of species were highly correlated when using all data from the National Lakes Assessment. However, we if used only low P or high P lakes to develop N traits of taxa and only low N or high N lakes to develop P traits for taxa, correlations between N and P traits of taxa were greatly reduced and actually negative in one case. If we are concerned about covariation in temperature and land use in the development of taxa traits (e.g., weighted average optima) for temperature, for example, you could stratify the data by land use into several strata, calculated temperature optima for taxa within each stratum, and then correlate optima of taxa among strata. We would expect a relatively high correlation in taxa traits (e.g., WA optima) for temperature among strata, but little correlation in taxa's temperature traits and land use.
Urquhart	The *Freshwater Biological Traits Database* is necessary for manipulations needed for the document being reviewed. As it is dynamic, it does need to be online as it is. Thus the report under review should briefly describe what that database is, how it was created, what it contains, and how to access it.

6. What next steps would you suggest, based on this work, as being the most informative for EPA's Office of Water and state bioassessment/biomonitoring programs to address climate change effects?

Reviewer	Comments
Carlisle	It is difficult to provide an enlightened response to this question. In my opinion, the report still has several significant deficiencies, particularly with respect to sufficiency of evidence, clarity of presentation, and readability. I have provided suggestions to improve the evidence for climate-related biological responses. I have also provided suggestions for organization and readability. Provided these weaknesses are rectified, I believe the Summary for Managers section could be pulled out and summarized in a 4-6 page fact sheet that targets the audience of all practitioners of biological assessments. Such a fact sheet would need to look very different than the current Summary for Managers section. It would need many fewer major messages (e.g., 2 or 3), a few very simple graphs and

	figures that drive the messages home, and a case study (or two) that brings the messages "down to earth." In its current form, the report is far too technically dense and disorganized for any particular audience. In addition, its tendency for overreaching conclusions despite weak (or poorly organized) evidence, is likely to make this important work a target for those who politicize EPA efforts on climate change.
Fennessy	A concerted effort to establish protocols to gather the data needed to more fully understand the effects of climate change. Priority should be given to several of the recommendations provided in the report, including protection of reference sites and the establishment and priority sampling of sentinel sites. The findings on the vulnerability of reference sites to land use changes is perhaps the most surprising finding of the report, and one of the most important for bioassessment programs generally. The establishment of sentinel sites that could be sampled more frequently is also an important recommendation in trying to build data sets that are more suitable for detecting temporal trends in the data. One excellent suggestion is that perhaps States and Tribes could cooperate in the establishment of sentinel sites within ecoregions (for example) that cross political boundaries. This could increase the efficiency of monitoring for climate change effects (as long as sampling methods could be agreed upon). Finally, the emphasis on traits-based analysis will help build biological indicators across regions since traits are not location specific as species tend to be.
Smith	To adequately evaluate temperature change effects it seems that there is a need to have information on air and water temperature. The cost of this information is relatively cheap. Bioassessment has been viewed by too many as something that is separate from chemical programs. That biology and chemistry are not sampled together is reducing the effectiveness of both programs. a. Figure out how to get states to collaborate with other states on sampling dates and methods so that data from different states may be combined. There should be more encouragement to simultaneously monitor many of the same sites. b. Figure out how to get agencies within states to collaborate so that chemical data is collected with biological data. There is a missed opportunity to enhance sampling through cooperative programs. There are likely numerous reference sites in national forests and parks. This study needs to stress the importance of cooperative sampling.
Stevenson	Please develop more specific guidance for how climate can be integrated into a policy-management-assessment framework. Without that guidance, information will not be gathered and used as effectively as with a clear understanding of how metrics will be applied to assess, for example, deviation from historic reference condition and current reference condition.
Urquhart	EPA needs to implement recommendation #2 (page 7-1). As a part of doing that, it should assemble a panel to assist in the development of such a program. That panel should be realistic, not "pie in the sky", by being constrained with a realistic budget. The

	objectives of that project should be simple and implementable, not a charge to conduct research for 10 years!

7. Please comment on the public comments submitted for this draft report. Specifically, which comments should or should not be addressed in the final draft?

Reviewer	Comments
Carlisle	I have read the public comment provided and have nothing further to add.
Fennessy	One public comment was received. The comment states, in part, that when… "' found' data (e.g., state data sets) are to be used for the pilot studies, some discussion of the methods and data comparability issues between the data sets and the data synthesis approach applied to address these issues should be included in the document." This is a good suggestion and would help with the clarity of the results.
Smith	Via email text: I took a look at the public comment and do not have anything to comment, so do not need to change my document
Stevenson	The concerns expressed by the public comment are covered in the report or in my review.
Urquhart	As of this date (5/3/2011) public comments have not been provided to this reviewer. (Subsequent to the premeeting comments, one public comment was received and commented about in the general comment #4.)

Additional Reviewer Comments

Additional Comments Submitted by Dr. Daren M. Carlisle

Literature Cited (should be incorporated into revised report):

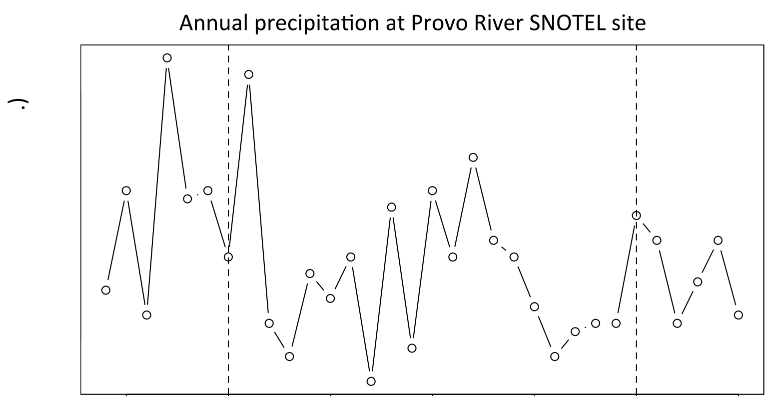
Pilgrim et al., 1998. Stream temperature correlations with air temperatures in Minnesota: Implications for climate warming. JAWRA 34:1109-1121.

Werhly et al., 2009. Comparison of statistical approaches for predicting stream temperatures across heterogenous landscapes. JAWRA 45:986-997.

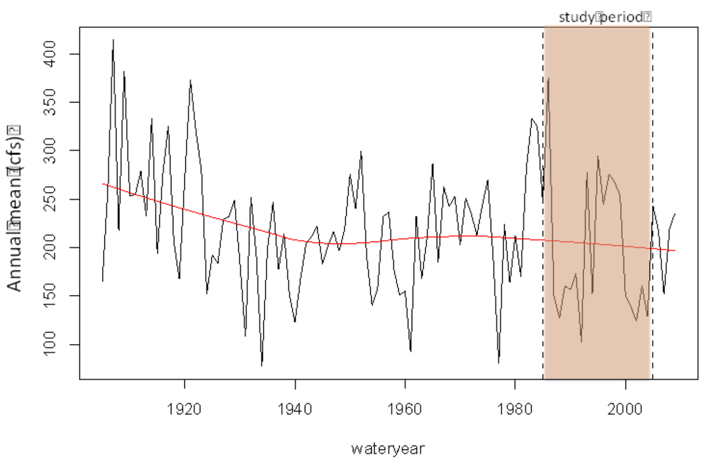
Yarnell et al., 2010. Ecology and management of the spring snowmelt recession. BioScience 60:114-127.

Yuan, L. 2007. Using biological assemblage composition to infer the values of covarying environmental factors. Freshwater Biology 52:1159-1175.

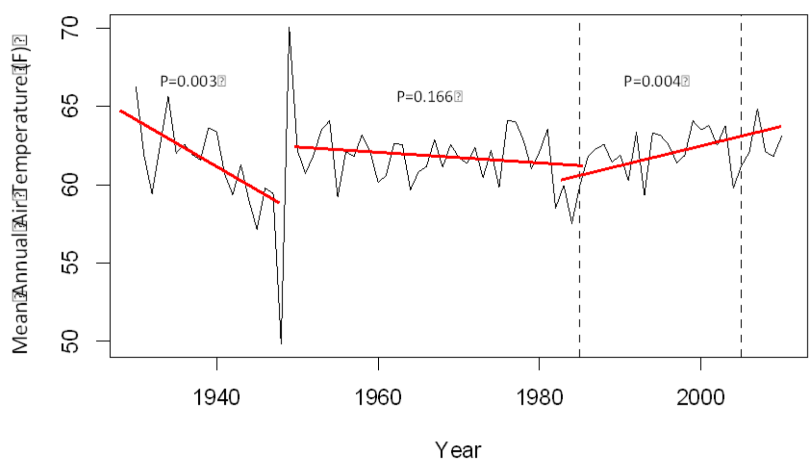
Attachment 1: Annual precipitation at SNOTEL site at Trial Lake, Utah, in Provo River Drainage (adjacent to Weber River basin). Brackets depict focal study period at Weber River site.



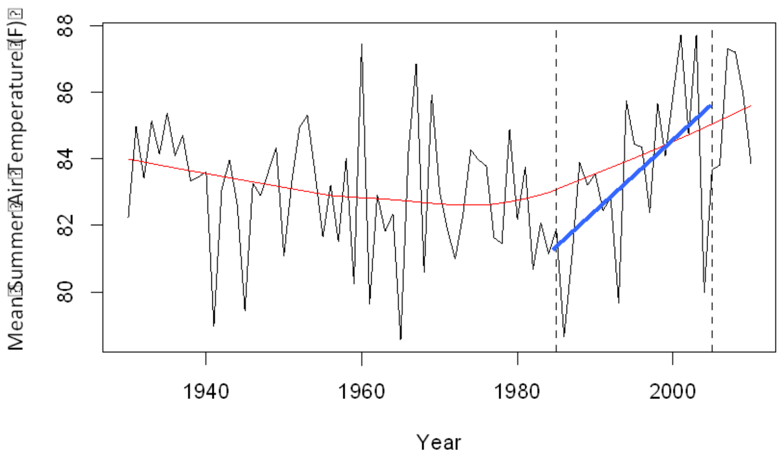
Attachment 2: Mean annual streamflows (cfs) at USGS gage 10128500, on the Weber River, Ut. Trend line is lowess smooth. Shaded bracket indicates focal study period for Weber River site.



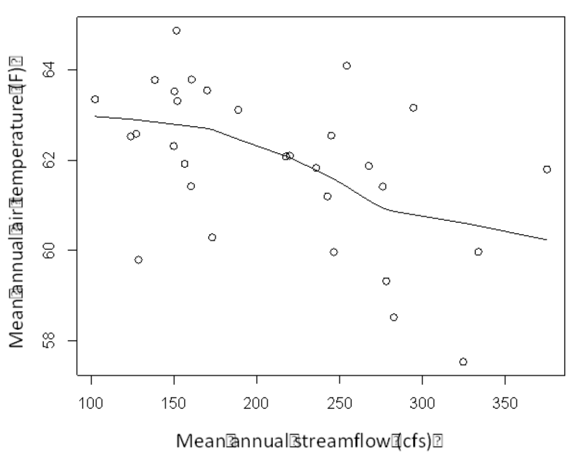
Attachment 3: Mean annual air temperature at Coalville, Utah (10 miles from Weber River site). Data from Utah Climate Center. Trend lines and P-values depict test for linear trend across years overlapped by each line. Brackets indicate focal study period for Weber River site.



Attachment 4: Mean summer air temperature at Coalville, Utah (10 miles from Weber River site). Data from Utah Climate Center. Red trend line is lowess smooth. Blue line is linear trend from 1985-2005, which is also indicated by brackets.



Attachment 5: Relation between annual air temperature and annual streamflow at the Weber, Utah site.



Additional Comments Submitted by Dr. M. Siobhan Fennessy

(Note to ERG: I realize this is a large report, but it would have been very helpful to have the figures in color, as they must be in the original. The black and white versions of many of the figures make it difficult to differentiate data points. Just a few examples include Figures 2-17, 2-18, and 2-28. In Figure 2-17 for example, it is impossible to differentiate the symbols for normal years from hottest years, so an evaluation of this result is difficult. This is true in many figures in the report.)

Section 2:

Some editorial comments have been made on Section 2 above, but in general the writing in this chapter written in a very passive voice which make it difficult to follow. Section 3 appears to be written by a different author and is more direct and clear.

Lines 1028-1030- this statement is not clear as written, why would a high degree of spatial variability between sites (in terms of stream order, elevation, etc.) reflect 'real' spatial variation in vulnerability to climate change effects.

Lines 1115-1120 – why not carry out an assessment of the responsiveness of the metrics listed here as part of this report?

Section 3:

Lines 145-147- this states that comparing the average of the first 5 years to the last 5 shows an increase of 6 EPT taxa at the Sheepscot site in Maine. However, Figure 3-3 shows that the mean for the last 5 years is heavily skewed by the data from 2005, making this statement much less convincing.

Lines 168-170 – I am bothered by statements like the one here (and found in several places in the report), which infer the authors are looking for a specific outcome. While it is reasonable to suppose that climate change will affect in-stream communities, statements like "the duration of data records was too limited to define significant trends" would be best ended with "if they are present".

Lines 300 – 303 – It would help if the authors could state the goal of this analysis in terms of protecting reference sites (or the lack of ability to do so). If climate change is a stressor and it leads to a decrease in the site condition classification score (Figure 3-5), isn't this a valid result? If there are climate change effects, how else do we document these? See also lines 498 – 500 (and if reference sites decrease in scores, non-references sites should do as well?).

Line 458 – should say "due to" instead of "do"

Line 464 – says "climate related changes in precipitation and flow could have increased cold sensitive taxa" - is there any evidence for this in this? Seems speculative.

Lines 508 –511 – not clear.

Lines 614 – 617 – this isn't totally clear, but it seems this finding conflicts with the data presented in Figure 3-9.

Section 4:

Lines 771 – Most programs conduct their sampling in order to make both spatial **and temporal** comparisons. The 5 year rotating basin approach is designed to assess watersheds, and compile data over time to track trends in stream quality in order to answer questions about the efficacy of environmental management, waste water treatment plant upgrades, etc. See also Section 4, lines 1229 – 1230, unless I am missing something, in most programs temporal patterns figure into spatial considerations.

Line 862 – 16% is a high proportion of urban land use to allow for a reference site. Some previous studies have identified thresholds of 15% impervious surface as the point where stream hydrology is altered significantly.

Line 879 – the absence of dams seems like an important criterion for reference sites, Does the land use analysis include this?

Line 902 – not clear how climate change influences reference site vulnerability – vulnerability to what exactly?

Lines 915 – 920 – don't MMI scoring schemes also allow a range of biological conditions that have regulatory significance to be identified? The case for BCG should be made more clearly here.

Lines 1031 – 1032 – not clear why different thresholds were used in evaluation of these reference sites.

Line 1058 – the reference to Figure 4-3 here appears to be incorrect?

Line 1118 – 1120 – not clear, if there is a shifting baseline problem for the reference condition, how or why would this be helped by measuring biological condition in comparison to pristine sites instead of present-day reference. Not clear how this could be done, and if so wouldn't pristine sites shift in response to the stressor of climate change as well?

Section 5:

Lines 1263 – 1265 – non-significant patterns could also be due to the fact that the patterns aren't significant. Looks like the authors are looking for a specific outcome. More support for this point is needed.

Lines 1421 – 1422 – the data that have been presented are not as definitive as this statement. The authors should take care not to make claims that the data don't support.

Line 1484 – says "establishing statistical differentiation." Not sure what this means.

Lines 1517 – 1519 – an example of how this might work would be helpful here.

Appendices:

General comment: while it is great to see detailed reporting, the appendices are a jumble with many different data sets used in different ways in different locations. For that reason they aren't nearly as informative as they could be if they were presented with a common format and a consistent approach.

Additional Comments Submitted by Dr. Eric P. Smith

Appendix A:

1. Obtaining mean annual temperature by averaging maximum and minimum temperatures is what is called a "quick and dirty" approach. This approach has to have high variability.
2. Some plots (line 84) exhibit seasonality. P-values based on simple tests are not appropriate in this case.
3. Line 106 – this figure suggest two groups of temperatures. Why?
4. If the authors are not going to check assumptions then use Kendall's correlations.
5. Why compare correlation coefficients to assess trends? Why not compare slopes or change in temperature?
6. Figure A-9 is based on daily maximum temperatures. Why are the data sparse in the later years? Why not use other methods i.e. the max over the year or the average or median of the daily maximum values? The R2 is diminished by the annual variability. Does it make sense to use R2 in this case? Report sample size.
7. Is there any sense that the inferred temperature method actually works? Isn't there variability with the method that should be reported?
8. Isn't trend in inferred temperature confounded with other factors (urbanization, for example)?

Appendix B:

The authors correctly point out the difficulties in using data from different agencies that have different methods and approaches.

Appendix C:

1. Identify the method for clustering and the distance measure.

Appendix D:

1. Benthic data are usually collected once each year. Temperature data may be available only once or at many times, possibly at a different location. How good are results given the need to summarize temperature? Do benthic taxa respond more to maximum temperature or average. Does it not depend on the taxa and perhaps time of the year?
2. North Carolina seems to have a stronger temperature relationship than Maine. Is this due to the use of presence absence data rather than abundance data? Why are relationships weak in Maine? Does this suggest the methods do not work well?
3. How is average annual air temperature calculated?
4. It would be useful to spell out what is meant by the maximum likelihood estimate as there are several ways to calculate the estimate depending on the assumptions of the model.

Appendix F:

1. Why focus on p-values? Sample sizes are rather small so p-values will not be all that useful. Pattern is more important.
2. Where is section 6?

Appendix J:

1. What is the curve in J3-2 based on? Isn't there a state effect? RI for example do not have any low density values? What are the dashed lines for?

Miscellaneous

1. Why are Kendall correlations only sometimes used?
2. The approach seems to be: run tests, interpret significant results. There seems to be little consideration to the assumptions of the tests. Without checks on assumptions, the results cannot be defensible.
3. Sample sizes are only occasionally given. They need to be presented with all the tests and analyses
4. Grouping data together increases sample size but also creates hidden biases when sample sizes, variances and confounding factors differ amongst stations.

Additional Comments Submitted by Dr. R. Jan Stevenson

Comments that should be sorted by question, depending upon how the panel interprets questions.

General comments:

I like that four geographically and climatically distinct case studies were used in the assessment of climate change effects on stream macroinvertebrate assessments.

It occurs to me that algal metrics may be less sensitive to temperature and hydrologic effects of climate change than invertebrates.

Relating comments to specific locations in the report:

1. From the summaries, page numbers are followed by paragraph and line numbers.
2. For the chapters, chapter numbers are followed by a period and then referenced to line numbers.

xv-1,1. Make it clearer, perhaps even in the title, that this is about streams and macroinvertebrates. Otherwise, explicitly address how this will relate to other resources.

xv-1-1. This is a long executive summary. Can you bring in the bullet points from xix and shorten this to a couple pages?

xv-3. About here you need to bring in concepts about how effects of climate change and local land use need to be assessed separately. See comment on policy-management-assessment framework in the face of climate change.

xv-4-8. It is not clear to me why you separated implementation from assessment design.

xvi-4-8. Use terms that are readily understood by a broad audience, especially in the Executive Summary. For example, what is "low pulse count"?

xviii. In several spots, the BCG and tiered uses are invoked as approaches "to guard against lowering water quality protective standards" (xviii-5-4), but it is not clear how. See comment on policy-management-assessment framework in the face of climate change.

xix-1-8. These bullet points are so short and filled with specialized terms that I'm not sure that policymakers, yet alone managers, will understand the point. For example:

1. xix-1-10. What are predictive models? This is used in many ways, relative to RIVPACS and modeling natural variation in multimetric expected condition. Will policy makers know what you mean?
2. xix-1-13. "Vulnerability" is used in so many contexts that its meaning is not clear here.
3. Xix-1-14. "to enable stressor diagnosis" or something that describes why.

xxii-1-4. Change "may" to "will". Statistically we know this is as certain as anything.

xxii-3&4. Is this an argument for using finer taxonomic resolution in metrics to help distinguish climate change effects from land use effects?

xxiii-1. This assumes than new models will be developed under successive climate regimes; i.e. that current expected condition is redefined in the future under different climate regimes. This may be true (see framework on policy-management-assessment), but it is not clear why from the report.

xxiii-3-1. Recalibration seems to be defining current expected condition – *sensu* my proposed policy-management-assessment framework.

xxvi-1-6 & paragraph. The RIVPACs and multimetric modeling framework of, for example, Cao et al. (2007) should be explained briefly and bundled relative to the idea of recalibration and redefining current expected condition from historic reference condition, per my proposed policy management-assessment framework.

xxviii-1-1. Distinguish climate effects and land use effects has to be a goal. The goal should not be detecting climate change effects, it should be assessing them.

xxviii-1-2. If assessing climate change effects is the goal, then targeting sensitive sites will overestimate climate change effects. A representative set of sites, from a diversity of natural climatic and geologic conditions, should be selected to assess climate change effects. Plus, this can be built into models and can be assessed with long term trends assessment.

xxviii-3-1. The rationale for protecting reference sites should include protection as part of antidegradation, as well as provide a mechanisms of assessing climate change. You should also assess climate change in trashed and moderately disturbed sites. The focus on reference sites and rationale for protecting reference sites is only part of what should be done to assess climate change effects. The arguments for using BCG and protecting reference sites is not sufficiently justified without a fuller policy-management-assessment framework. Thus the repeated inclusion of reference site monitoring and the BCG gradient could be misconstrued as promoting favorite themes without sufficient justification, which could compromise the integrity of the report.

xxix-2-2. Change this line to “use effects are great and widely distributed”. To say they have not been conceived is not correct. To say they have not been quantified may be true, but is not the point.

xxx-1-1. The importance of collecting water chemistry data and climate effects on water chemistry is not well developed in this section. The entire section is a less developed and supported than other sections. Strengthen these arguments or delete them.

xxxxi-7-1. Stream hydrology should be considered as a factor in climate change as well as temperature. We could expect shorter invertebrate generation times in streams that are more stressed by drought and flood (Riseng et al., 2006). These factors (flood/scouring and drought), in addition to temperature change, should be included more explicitly and frequently in discussion of climate change effects. Biological assessment may be particularly powerful for assessing hydrologic factors that vary at interannual scales and that are difficult to assess by measuring physical factors without continuous monitoring.

Chapter 2:

I found chapter 2 to be a very difficult chapter to read and numbers refer to the chapter and line.

2.165. Greater than what? Insert “than climate change” after effect.

2.265. It is important to distinguish how the trait based indicators can be used. Indicators based on species sensitivity or tolerance to specific environmental factors can be used for three basic efforts, measure a biodiversity response isolated to a specific stressor (as much as is possible), actually infer stressor condition, and help forecast changes in assemblages in the future. In addition, causal relationships between what you think traits are reflected and what are reflected should be confirmed using multiple lines of evidence (Beyers, 1998) and propensity analyses (Yuan 2010).

2.242-246. At this point in the report, the reader has no idea where these regions are and what to expect in results. Should we expect different results in the Wasatch Unita Mountains and Colorado Plateau? Maps of the regions and some characterizations of regional references should be developed.

2.246. Could the lack of relationships between temperature and invertebrate metrics be related to temporal lags? For example, I’d expected metrics based on individuals to be relatively responsive on short time scales, so the number of individuals of warm-water taxa could increase pretty quickly, although there is probably considerable effect of conditions the year before on number of eggs laid. Metrics based on the number of taxa may have even greater temporal lags. Taxa with short generations times could respond quickly to changes in temperature.

2.246. Were relationships of these metrics and climate change also evaluated with longer term regression models with time and long-term temperature averages, as for example in Figure 3-13? Was the same suite of analyses run on all long-term datasets (e.g., box plot comparisons (Mann-Whitney U’s?), regression, variables)? Why not just related metrics to temperature in a regression, why use a categorical approach?

2.309. Why not relate a metric (e.g., invert size, generation time) associated with flood or drought disturbance with driest/wettest years comparisons, versus temperature tolerance?

2.387. Elevation does not have a direct effect on invertebrates. Don’t you have temperature data from either air or water to show what the elevation proxy for temperature would be and whether that’s sufficient to generate an effect?

2.402. I looked back and could not find where this was described for Utah.

2.409. “The three sites” produce box plots with lots of observations in Figure 2-10? Credibility of the analysis is being compromised by some of the apparent “haphazard” grouping of data and different analytical approaches. For example, why pick 150 m as an elevation benchmark? Plus, 150 m does not seem like a lot of difference in elevation.

2.482. Should you insert “WA models for” in front of “low flow parameters”? It’s the results of the WA models, not the low flow parameters, that you are referring too, right?

2.491. Great, size matters (or generation time) with respect to hydrologic parameters.

2.494. It is not clear why you selected the variables that you did in these ordinations. For example, why does year matter? What else was in the analysis that did not matter?

2.496. References to figures and tables in the text should be parenthetic, not complete sentences that repeat part of the figure caption or table title.

2.543-544. If it is hard to draw conclusions from data. Don't present them without elaboration more about what their likely meaning is or is not.

2.558-561. Why were only reference sites with long term data used? Plus later descriptions of the Sheepscot raise questions of whether these reference sites have the same levels of human disturbance. Changes in minimally, moderately and severely impacted sites could differ.

2.569-570. Why were different analytical approaches uses in Maine and Utah?

2.577. I cannot find a justification for why Utah results would represent all western states or where results from western states were analyzed. This type of extrapolation of the results should be dropped here and in future references to New England and the Southeast.

2.583. I question the reliability of an ordination with less than 30 observations. This has about 17.

2.755-756. Credibility of findings is compromised when multiple tests were done and the reader does not know how many. Add how many metrics were tested. If 4 is not a significant number of all tests, then you should not present results of any because they could have happened by chance. By presenting them, you imply there is some certainty they mean something, when statistically there is little certainty if only 4 metrics were significant in a large number of tests.

2.1020. Please tell us how many non-reference sites had long-term data that could have been analyzed.

Tables 2-4 to 2-6. Are these P values corrected for multiple tests? If the answer is no, then these results are not particularly strong and there's no need to present them.

Pg. 2-58, ln 1-6. Why build the case that diversity decreases with all these caveats? The reason is surely that multiple cases show that decreases in cold-water taxa without commensurate increases in warm-water taxa seem to affecting metrics and are expected to be an issue with metrics in bioassessment programs as time goes forward. When presented like this, with the "one Utah station" and complexity caveats as well as after pages of correlations, the certainty of results presentations and interpretations is compromised. A conceptual model and a priori hypotheses sets up expectations for analyses of results upon which such seemingly anecdotal results have more credibility. For example, why would you expect decreases in cold-water taxa without commensurate increases in warm-water taxa? Why wouldn't you expect numbers of warm-water taxa to increase with temperature? Is environmental change too rapid for colonization? Are there no warm water taxa? Of course this is not what is observed in the Sheepscot (Figure 3-13), although the symbols are hard to see.

3.70-71. State-specific analyses do inform regional views, but not enough to use either New England and later Northeast in the title to overstate this level of transferability. Vermont is very different than Maine with respect to many of the findings demonstrated. In some cases, North Carolina may be more informative for Maine than Vermont. Delete regional references in titles.

3.190. Typo in $p<0.001$ or $p=0.01$.

3.692-693. It seems to me that some biological metrics are selected for their diagnostic value, but most are selected for their characterization of the “natural balance of flora and fauna …..” as per the definition of elements of biological integrity proposed by Karr and Dudley (1981).

3.717-722. This is a good example of the policy-management-assessment framework that should be developed around which the report should be organized. Perhaps that’s beyond the scope of the report.

3.731. What is OCH?

3.737-739. I would like to have seen the hydrology traits explored more in analyses for the report.

4.768. This section addresses collecting land use, habitat, and abiotic information such as water quality information, which is important for stressor diagnosis. Distinguishing changes in temperature, hydrology, and water chemistry resulting from climate change and land use change is critical for causal analysis and management. Relating this all to the policy-management-assessment framework is essential.

4.768. Long-term records should be developed at non-reference sites.

4.783. I am not convinced that sampling the same sites every year is necessary (see Figure 3-13). Nor is regional watershed-specific rotation satisfactory, but a more regionally distributed probabilistic sampling could work. More analysis, probably best by statistical modeling and scenario analysis, would be important to address this sampling strategy issue.

4.829. This level of methods was really helpful to understand the results presented.

4.874. Why not stream size?

4.875-876. I’m not convinced grouping by ecoregions, whether level 3 or 4, is a sound approach.

4.893. insert “to” after “power”

4.896-899. Here it states that transferring results from just a few sites to regions is “problematic”. Yes, indeed. So remove regions from titles.

4.910. We should not assume that non-reference sites will response similarly to climate change as reference sites. Later in this report, urban streams are shown to be more vulnerable than reference streams to hydrologic alteration. Nutrient loading and temperature increases with climate change could be greater for agriculturally dominated watersheds in which riparian canopy and buffers are absent than for reference sites. Thus, non-reference sites should be included in the Sentinel Network.

4.1113. or farmed

4.1142. I like this emphasis on cross-state collaboration. Since this is unlikely in most cases, maybe the federal government should run the sentinel program as part of the ongoing aquatic resources surveys. There’s no reason we only samples streams, lakes or wetlands during a single year, just as we don’t have to do rotating basins to save money.

4.1151. Why is this the first time that non-reference long-term monitoring is mentioned? The message to managers does not need to be so abbreviated that it becomes biased.

4.1159. Stressor diagnosis should be emphasized more.

4.1167. This is important. Lack of long-term trend data from non-reference sentinel locations WILL present limitations to separating effects of climate and land use change, which is critical for management.

5.1379. Can't we use regional air temperatures, which are readily available, to estimate what regional water temperatures should be based on known hydrologic differences among sites?

6.1413. Delete "threatens". Insert "challenges the current vision of".

6.1467. With TMDLs, stressor and causal diagnoses will be critical for distinguishing between management strategies for land use or effects of climate change.

6.1485. I don't think liable is the right word.

6.1481-1484. Not if both reference and non-reference change.

6.1516-1519. I'm not convinced refined aquatic life uses is nearly as important is tracking the generalized stressor gradient and human disturbance gradient to help with stressor and causal diagnoses that will be important for management strategies for land use versus climate effects.

7.1. This seems like an executive summary.

Literature Reference

Barbour, M. T., S. B. Norton, H. R. Preston, and K. W. Thornton, editors. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publication, Pensacola, Florida.

Beyers, D. W. 1998. Causal inference in environmental impact studies. Journal of the North American Benthological Society **17**:367-373.

Cao, Y. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological **26**:566-585.

Riseng, C. M., M. J. Wiley, and R. J. Stevenson. 2004. Hydrologic disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. Journal of the North American Benthological Society **23**:309-326.

Stevenson, R. J., R. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004a. Designing data collection for ecological assessments. Pages 55-84 *in* M. T. Barbour, S. B. Norton, H. R. Preston, and K. W. Thornton, editors. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publication, Pensacola, Florida.

Stevenson, R. J., B. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004b. Interpreting results of ecological assessments. In: M. T. Barbour, S. B. Norton, H. R. Preston, and K. W. Thornton, eds. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Pgs 85-111. Society of Environmental Toxicology and Chemistry, Pensacola, Florida. ISBN 1-880611-56-2.

Stevenson, R.J. and S. Sabater. 2010. Understanding effects of global change on river ecosystems: science to support policy in a changing world. Hydrobiologia 657:3-18.

Yuan, L.L. 2010. Estimating the effects of excess nutrients on stream invertebrates from observational data. Ecological Applications 20:110-125.

Additional Comments Submitted by Dr. N. Scott Urquhart

Other substantial comments:

a. What is the goal or objective of this study/report? The title and page 1-2, lines 39 & 40 imply the study is to "determine what components of bioassessment programs are threatened by climate change." Yet in a few pages the emphasis has shifted to using bioassessment data to detect the effects of climate change. Study/report goals should be stated unambiguously and early in the Executive Summary, in the SMP, and remain the focus throughout.

b. The figures need a lot of work to make them consistent from one to the next. They often contain panels A, B, etc. In most cases an abbreviated definition of A, B, etc as a part of the subfigure would greatly increase readability.

c. The authors seem to assume that randomly selected sites cannot be revisited. Exactly the opposite is true.

d. Much of the appendix and some of the main text uses English units of measure. Isn't EPA supposed to use metric measures in its reports? Why this inconsistency?

e. Many benthic macroinvertebrate collections are planned to occur during times when stream water comes mainly from base flow. As base flow arises from ground water, the temperature of such water is greatly moderated by its long passage through the ground. Such water may show very little effect from changes in air temperature. This fact appears to have been mainly ignored in this report. It is recognized at least one place (somewhere in SMP).

f. Executive Summary: Why no explicit mention of recommendation #2, (page 7-1)? This is important to mention early, because it implies the need for funds.

g. Appendices: This reviewer counted a total of 50 planned appendices, plus another report with 5 appendices. This is an overwhelming amount of supporting information. Please write a substantial introduction to the appendices which provides a very good road map as to how they link and apply to the goal of this document/study. Further, many readers might be unfamiliar with a major data source used extensively in the analyses reported in the appendices: PRISM. A clear and simple explanation could be based on http://prism.oregonstate.edu/docs/przfact.html .

h. Ohio has very extensive data sets, but they appear to have been omitted from major analyses, such as the temporal trends in the main text, and in the database. This apparently major omission needs to be addressed, either with an explanation or inclusion of more information from there.

Other Comments:

a. Abbreviations: Need to also include CC, CADDIS, and IPCC

b. In standard English usage, a comparison should have the form, **this** is compared to **that**. This report frequently leaves the reader with the responsibility to deduce "**that**". This could lead to misinterpretations of what was intended. For example, page xxi, line 2, "higher proportion" than what? Do the authors mean than at a higher elevation? Page xxiii, end of the Evidence paragraph, "larger" than what? Do the authors mean than the present differences?

c. Page xvi, line 9 “highly sensitive” Really? Elsewhere the report indicates some trait groups are somewhat, but not highly sensitive to moderate temperature changes.

d. Should most references to “states”, as users of bioassessment, refer to “states and tribes”? See, page xxiv, paragraph on “Adaptation response”.

e. Page xvii, last two sentences, last full paragraph: This important point needs more visibility here and elsewhere.

f. Page xviii, lines 18 & 19: “… so that climate effects can be tracked” → so that they are not influenced by

g. Finding 7 about abiotic data (page xxx): This reviewer has had extensive, but very disappointing experiences with such data; he has even placed and retrieved data loggers himself. Much of the biotic data happens once, or at most a few times, in a year. It sounds good to get lots of data, but there is no practical analytic technique for identifying what temperatures among thousands available drive a relatively rarely evaluated macroinvertebrate response.

h. Page 1-2, lines 39 & 40 need to be moved to Page1-1, lines 3 & 4 with subsequent information defined as context.

i. Figures 2-1 through 2-4, and subsequent ones should have panels identified similar to those in Figure 2-5. Figures should be as consistent in placement of similar items as possible. For example, 2-3 & 4 display %, number, %, number, while 2-5 displays number, %, number, %, and 2-6 displays number, number, %, %. Starting with Figure 2-7, sample sizes are mentioned, but not earlier. Give your reader a break! Reduce confusion to a minimum so readers can concentrate on content!

j. Figure 2-18. station 4951200 (Virgin) – Include the station name, as elsewhere.

k. Figure 2-23: What are the dotted curves? Confidence interval in the location of the line at a point, or a prediction interval on the location of a future point? Or something else?

l. Page 2-41, line 803: Why no mention of Ohio here? It is covered extensively in the appendices.

m. Page 3-22, line 470: … ICI and IBI improvements, OR there may be no observable effects due to climate change.

n. Page 4-4, lines 843-844: Criteria such as these need to be screened for plausibility. If conditions are too stringent, all sites might have to be in national parks, forests, or limited access areas as in the west!

o. Page 4-8, line 970: When did Maryland sneak in here? Figures relate to only North Carolina.

p. Page 4-18, line 1186-1187, &1195-1196: The first sentence is UNTRUE! The second sentence describes how to completely avoid the imagined problem.

q. Page 5-1, lines 1225-1228: The states have their responsibilities. The Climate Change Program needs to be realistic that the program’s objectives differ substantially from the states’.

r. Page 5-8, line 1358: Compositing samples is much like summing numbers; the only problem results when subsequently samples need to be split. This reviewer has looked at this problem closely. Counting is one of the major expenses in the process of evaluating benthic macroinvertebrates. Compositing with possible splitting reduces the number of samples which must be “picked.”

s. Page 6-3, lines 1474-1477. This statement is too strong or assertive, especially relative to responses being predictable.

t. Figure A-6: Differencing of variables ordinarily induces substantial correlation among the differences. Has this been accounted for in the analyses?

Details:

a. The study (page xv, line 4) What study? This study? (line 9) Does "the report" mean this study? If the intent is to refer to the present document or study, rather than some other one, identify it as **this**.

b. Figure SMP-2: "… between climate change trends" → climate change (CC above) trends

c. Page xv, line 10: "," omitted after taxa.

d. Figure 2-28: Titles interior to the figure have been squeezed to such an extent that they are unreadable.

f. Figure 3-1: Where is the PRISM mean annual air temperature?

g. Page 3-8, line 190: p=001; is this meant to be p=0.01?

h. Table 3-4: Formatting – Last column should be labeled something like Difference in Final Scores (with no horizontal line). The – signs need to be adjacent to the numbers. The latter problem repeats itself in Table 3-7.

i. Are the bar colors the same in figures 3-4 and 3-6? They should be.

j. Page 3-20, line 422: BI → NCBI?

k. Page 3-21, lines 457-463: … two possibilities exist. The first one clearly follows immediately. Where is the second? Is it the Other scenarios? If so, that implies more than two possibilities. This writing needs to be clarified.

l. Page 3-21, line 458: do → due

m. Page 3-33, line 716: Figure 3-13 should be 3-11.

n. page 4-11, lines 1021-1024: English and metric units mixed up.

o. Page 4-12 & Table 4-2: Is "BC" the Base Case?

p. page 4-12, line 1057: Figure 4-3 or perhaps 4-6?

q. Table 4-2: (…(…)) – latter) omitted

r. Page 8-18, Stoddard citation: Why capitalize Johnson?

s. Figure A-9: Why three lines? What has happened to minimum and average temperature?

t. Page B2-7, line 246: Data for North Carolina were compiled from ____ into a database. What has been left out?

u. Figure B3-15b: (… (…). Should be (…(…)).

v. Figure D-1: Vertical label is reversed and illegible; (u) → (μ); if the mean is denoted by a Greek letter, the standard deviation should be also: σ.

w. Page H-15, line 303: Where is the figure referred to here as H3?

Appendix A: Individual Reviewer Comments

COMMENTS SUBMITTED BY

Daren M. Carlisle, Ph.D.

Date: 13 May 2011
From: DM Carlisle, US Geological Survey
To: ERG
Re: Review comments on US EPA report on climate change and bioassessments

General Comments

This report tackles an issue that has far-reaching implications for how aquatic ecosystems are assessed and managed. The primary focus of this study is the potential consequences of climatic-induced shifts in the baselines and climatic-related confounding of the indicators used to assess ecological health. Although the report contains much useful information and findings relevant to this important topic, the current organizational structure renders this information extremely difficult to locate, interpret, and judge. In addition, there are several important lines of evidence on which the report is inexplicably silent—most notably hydrological variability and characterization of the longer-term climatic context of the period examined (i.e., the last 30 years).

Finally, in an effort to fully understand the depth of the analyses in the Report, I focused on a site (Weber, Utah) about which I have first-hand knowledge. I have provided several Attachments from data sets that I was able to readily obtain and that I believe would enhance the study. In doing so, I realize that scope is a limitation of every study. However, I think that examination of some additional data sets would be beneficial and strengthen the evidence for the Study's primary conclusions.

Responses to Charge Questions

1a.) Provides sufficient technical evidence...additional evidence needed?

The conceptual model depicted in Figure A-1 serves as a useful framework for organizing the evidence provided in the Study. The stated objective of the Study is to evaluate whether "components" of bioassessment programs are influenced by climate change, although most of the effort appears to be focused on biological indicators. As illustrated in Figure A-1, the effects of climate change on stream ecosystems begin with two driving factors: air temperature and precipitation. These factors in turn influence the thermal and hydrological regime of streams, which influence biota directly and indirectly.

As evidence for the Study's first major conclusion (Page xx)--that biological indicators are responsive to climate change--the authors cite associations between the abundance & richness of cold-water taxa and annual air temperature. For this evidence to be convincing, the following causal linkages must be established (per Figure A-1):

a. air temperature and/or precipitation varied over the time period studied,

b. variation in the thermal and/or hydrological regimes of each stream is associated with variation in air temperature and/or precipitation,

c. variation in biological indicators are associated with *and specific to* variation in hydrological or temperature regimes.

In general, the Study demonstrated that in most of the case studies there was variation--and in many cases monotonic trends--in precipitation or air temperature over the focal time periods (item a above). Unfortunately, the Study only appears to have examined climatic data over the focal time periods (1970s onward) rather than the longer record of available data, which would have provided critical information about the historic context of climatic variability during the focal time period.

However, analysis of longer-term and more local (than PRISM) hydro-climatic data would likely have revealed important insights. For example, the Study reports no trend in mean annual precipitation (A-5) and an increasing trend in mean annual temperature (Figure A-3) in the Utah mountains since 1970. Annual precipitation data since 1979 is available from a SNOTEL site in a watershed adjacent to the Weber, Utah site (**Attachment 1**). These data corroborate the finding of no trend, but also show that during the focal study period (1985-2005) the Weber River basin experienced extremes of annual precipitation--having the driest and 2nd wettest years in the last 30+ years. The last 30 years of precipitation are put in a longer-term context by examining streamflow data from the Weber River, just upstream of the site examined in this Study, and having over 100 years of record (**Attachment 2**). Annual streamflow data reveals that runoff in the Weber basin has been variable about a relatively stable mean since the 1940s, but was much higher before that period--indicating a drier climate over the last 6 decades relative to earlier time periods. Importantly, the runoff observed during the study period (1985-2005) included one year with the highest runoff in six decades, as well as several years of rather typical low runoff.

Examination of local, long-term temperature data also provides new insights. Daily maximum temperature data were available for Coalville, Ut (about 10 km downstream of the Weber River site) from 1930. These data (**Attachment 3**) show that the observed trend in mean annual air temperature based on PRISM data (Figure A-3) through the 1980s and 1990s was well within the range of temperatures encountered in previous decades, which suggests that mean annual temperatures during the focal study period were not necessarily anomalous. However, examination of local (Coalville, Ut) mean *summer* air temperatures (**Attachment 4**)--which was unfortunately not examined in the Study--show a dramatic short-term increasing trend during the study focal period. These data show that the summertime air temperatures encountered during the focal study period were among the highest and lowest recorded in the last 80 years. Finally, since 1980, it is apparent that warmer air temperatures were associated with lower streamflows (**Attachment 5**), and in general colder years had higher streamflows. This relationship implies that it is difficult to separate the effects of purported increases in air temperature from the influence of streamflows.

In summary, examination of historic hydro-climatic data from local sources revealed that at least one site, the Weber River in Utah, experienced historically dramatic fluctuations in climatic conditions--particularly precipitation and summertime air temperature--during the focal time period of biological monitoring. It is regrettable that the Study did not provide this type of context for the other sites studied. This lack of historical context may not be a critical weakness of the Study, but it nevertheless leaves its readers without an appreciation for the climatic conditions encountered at the study sites, and it leaves the Study vulnerable to criticism that its operational definition of “climate” is limited to the last 30 years. Also unfortunately, trends in average summer air temperatures--which are most likely to cause changes in stream water temperatures--were apparently not examined in the study.

A more critical weakness of the study is its lack of evidence that stream temperature or hydrologic regimes (per Figure A-1) were associated with climatic variability (item b above). Continuous stream temperature data are notoriously scarce, especially in relatively undisturbed watersheds. The Study authors apparently examined stream temperature trends at a few USGS monitoring sites (Table A-1), but failed to adequately report the findings. Rather than R-squared values--which are not informative in trend evaluation--the Study should have reported estimates and associated uncertainty (i.e., confidence limits) of the change rates and number of years and periods of data for each site. It is also apparent that many of the sites had substantial human influences--which undoubtedly confound climate-associated trends. As a result, the stream temperature trend analysis performed in the study is weak evidence that climate-related changes in stream temperature have occurred over the course of the focus study period. The Study cites several recent reports of similar findings (e.g., Kaushal et al. 2010), and there are other studies showing relationships between air temperature and stream temperature, but these merely strengthen the *hypothesis* that stream temperatures ***may*** have increased at the study sites.

It is informative to examine the plausibility that an increase in average summer temperatures could have caused a trend in stream water temperature. Again relevant to the Weber, Utah site:

1. If we take table A-1 at face value, an average of the intermountain west streams suggests that average summer water temperatures could have increased roughly 1 degree every 10 years. This would suggest a potential for a 2 degree increase in mean summer water temperatures at the Weber River site from 1985-2005.

2. **Attachment 4** shows a 0.2 degree (F) increase / year in mean summer air temperatures from 1985-2005, and hence a 1.5 degree C total increase over that period.

3. Is it plausible that 1.5 degree warming of summer mean air temperatures is of sufficient magnitude to increase mean summer water temperature? Previous studies (Pilgrim et al. 1998, Wehrly et al. 2009) developed empirical relationships between air and stream water temperatures, which found that at an annual time scale, a one degree increase in air temperature is associated with a roughly 1 degree increase in water temperature. Given this generality, it seems probable that mean summer stream temperatures increased from 1985-2005 by at least one degree, which seems modest but is remarkably similar to the average magnitude of temperature increases actually observed in the western streams within Table A-1.

4. In conclusion, given observed trends in air (especially summer) temperatures, trends in water temperature at other streams, and empirical associations between air and water temperature, it is certainly plausible that the mean summer temperature at the Weber, Utah site increased 1-2 degrees from 1985-2005.

Had the authors presented the evidence in a way similar to what I just outlined, readers would probably have more confidence in that evidence. As currently written, the readers must make several large leaps and assumptions to believe that observed air temperatures could plausibly alter stream thermal regimes.

The most glaring omission of the Study is its lack of hydrological analysis, despite the fact that hydrologic regime change is a major pathway of climate change effects on stream biota (Figure A-1). Many of the study sites have USGS stream gaging data at or near the biological sampling locations (Appendix C), many with data records of several decades. In addition, many gages with long-term records can be found in the same regions as the study sites. It is therefore quite shocking that trends in streamflow were not examined at the study sites. For example, just upstream (~5 miles) of the Weber, Utah station, there exists a gage with >100 years of daily flow record from a relatively undeveloped watershed. Rudimentary trend analysis of these data reveal two important facts. First, as indicated by the runoff (precipitation) data discussed earlier, annual high flows declined rapidly until the 1940s, but have remained relatively unchanged (average conditions) since then. However, within the period of study at this site, abnormally high flows occurred in one year, but most years were below the long-term average. Second, monthly mean flows show gradual declining trends during the winter months over the last century (results not shown), which indicates declining groundwater recharge--possibly due to human activities or long-term natural processes. Of particular interest would have been a trend analysis of the hydrological characteristics discussed in Appendix J, which the Study considered to be ecologically important. In summary, the Study provides virtually no long- or short-term hydrological context at the study sites, and therefore offers no evidence of the connections between climate-streamflow-biota as illustrated in Figure A-1.

In the absence of evidence about trends in the physical characteristics of the study sites, the Study relies primarily on an association between certain biological community attributes and mean annual air temperature. This evidence requires much greater scrutiny because the intermediate causal linkage (ie stream temperature or flow) is weak. The Study undertook a major and laudable task of developing temperature optima for many macroinvertebrate taxa. This contribution to science is potentially significant. But like many such empirically derived optima, there is a risk that factors other than stream temperature influenced the analysis. Note that the optima were developed using data from many sites, including ones with significant human influences. As a result, there are possibly confounding gradients of human impact in the data (a full treatment of this issue is given in Yuan 2007, and should have been considered in this Study). As a consequence of this problem, the optima should have been independently evaluated--that is, by evaluating how well biota-inferred water temperature predicts actual measured water temperature. This was apparently attempted, as shown in Figure A-16, but it is unclear whether these data were independent and the temperature data appear to be instantaneous, which is not really useful for this purpose.

The lack of optima validation is especially critical because the major study conclusions rest on temperature-related changes in the biological communities. The critical reader is therefore left without confidence that the observed changes in the biological communities were in fact related to air temperature and not other factors. For example, the Weber, Utah site showed a strong association between "cold-water" taxa and mean annual air temperatures, but there are other plausible reasons why EPT (and possibly cold-water) taxa would decline at that site. The site lies downstream of several thousand acres of irrigated pasture land and two towns (including a wastewater treatment plant). Summer air temperatures influence irrigation diversions, which can reduce streamflow, increase water temperature, and change water chemistry. In fact, there appears to have been a strong trend in chloride concentrations at that site (Figure 2-30) which could alone account for losses of EPT taxa. To their credit, the authors acknowledge the possibility of other factors, but should have done more to control for them—such as validating whether the biological response is indeed specific to changes in stream temperature.

In summary, the evidence that bioassessment indicators are effected by climate change is essentially limited to correlations between mean annual air temperature and the relative abundances of several macroinvertebrate taxa that are known to be sensitive to many human influences--not just increased water temperature. Because the purported biological response to climate is not *specific* to water temperature (item c above); and because there is no direct evidence that variability in climatic variables was related to stream temperature or stream flow, the overall evidence in support of this primary finding is weak.

Granted, the authors acknowledge the possibility of confounding factors and show some of these data where available. But the authors also apparently overlooked key hydro-climatic information and analysis that could have strengthened the case for causality and provided critical climatic context for each site--as I attempted to illustrate for the Weber, Utah study above.

Several possible corrective measures have been suggested above. First, the 30-year study period must be placed into a longer-term context using relevant climate data for a much longer time period. At a bare minimum the authors should define early in the report what is meant by "climate change" in the context of this study. For example, do the authors define climate change as annual variation in mean annual temperature and precipitation over the last 30 years? Or, do the authors define climate change from a longer-term perspective? Second, trends in stream temperature must be reported more clearly and convincingly, which includes limiting the analysis to undisturbed sites. Third, trends in streamflow characteristics (both short and long term) must be considered as a potential response to change in climate. Even if there is no evidence of trends in precipitation, it is known that trends in temperature can still influence patterns of runoff--especially with respect to snowmelt

(Yarnell et al. 2010). The authors cannot forward a credible analysis of climate-related effects on stream biota without explicit examination of climate-related effects on streamflow characteristics. Finally, the authors should seek data sets (NAWQA has several datasets of paired macroinvertebrate data and long-term water temperature measurements) and independently validate the temperature optima developed in this study. Even if all other attempts to strengthen evidence are not fruitful, increased confidence that the temperature optima are specific to temperature and not confounded by other factors, would greatly strengthen the Study's primary conclusion.

1b.) Factors to consider in applying results of this study...

Based on the limitations of the current draft, I could not recommend transferring insights from this study to other areas.

2) Has EPA pulled out the most important findings? If not, which?

The Summary for Managers and Policy Makers (SMPM) section is too long and contains many "findings" that are not really uniquely attributable to this study. The first finding is far too strongly worded, as is the information presented as evidence. For example, this study did not really examine climate change. Rather, it examined variability in air temperatures and precipitation over the last 20-25 years. This finding should be more clearly stated as: "Some metrics used in indicators of biological integrity are associated with climatic variability."

The second finding is poorly worded and difficult to follow. The writing needs to be more precise. I am very familiar with RIVPACS models but it took several readings of this section and I'm still not entirely clear what the point is. RIVPACS models derive expected probabilities of capture for each taxon based on relationships between environmental factors (such as average climatic conditions) and biotic structure within a population of reference sites over a fixed time period. Hence, estimates of E for both disturbed and undisturbed sites will always be based on the original time period over which the model was developed. In the future, if the climate changes as expected, the original RIVPACS model would in theory detect it as "drift" in the condition of reference sites sampled in the future. Specifically, the observed and expected communities will diverge at reference sites presumably due to climate--since we assume there are no other human influences contributing to biological changes at reference sites. So, I am confused by the recommendation that RIVPACS models be continually updated as a means to, presumably, detect changes due to climate.

The third finding is nothing new to science or management and therefore not uniquely contributed by this study. The "evidence" provided for this finding does not actually present any results from this study. I suggest dropping this message.

The fourth finding is poorly written--but it also is not a new finding. Again, the evidence presented contains no new information from this study. Given the finding, I would expect to have seen evidence that, through time, the biological composition of some reference sites have in fact become more similar to non-reference sites, and this change is entirely due to climate. This message should be dropped.

The fifth finding, unlike the previous three, provides some evidence from actual results of this study. Unfortunately, the idea that "place matters" is worn out. Isn't this a common assumption anyway? How is this a new idea? Either spin this finding in a way that is clearly new, or drop it entirely.

Findings 6-8 are also common knowledge and don't seem to present evidence that is uniquely provided by this study. It is unclear how this study contributed to these generalizations.

Finding nine seems to be based on "simulated" data as in Figure 4-2, where progressively more taxa are excluded (ie eliminated) from metric calculations. This is weak evidence. Again, I was expecting to see actual data showing a decline in biological integrity through time at a reference site. If this was actually observed in the analysis, it is unfortunately not clearly presented as evidence.

Finding ten is interesting and potentially significant, but the authors provide no evidence that this phenomenon actually occurred at the sites studied. Specifically, did the authors observe changes in the biological conditions at a site that are sufficient to cause reconsideration of its designated or attainable use? At best, the authors found subtle changes in community composition that were associated with annual variation in temperature or precipitation. And, general knowledge is given as evidence for this finding.

In summary, most of the findings, as currently written, read more like a text book than solid and specific discoveries from the current study and their implications to biological assessments. I suggest reducing this section to a fraction of its current length, and focusing the topics on a small number of pointed findings that clearly stem from the research conducted.

3) Are the analytical methods appropriate?

The analytical approaches and their respective methods are voluminous, scattered throughout the report, and often times inadequately reported. As a result, it is extremely difficult to address this question. For example, I could not locate a single, detailed description of the methods used to generate Figures A13-A16. The accompanying text (Page 4-4) provided only general descriptions, then seemed to point me to Appendix D for further details. I could not readily locate any additional relevant information there. I frequently encountered this problem, which suggests an underlying issue with organization of the document.

In addition the hydro-climatic evidence outlined above, I also examined the RIVPACS analysis section, and found it to be unfulfilling. Figure 3-9 suggests that for two sites on the Colorado Plateau that were presumably reference (and therefore used to develop models for E), there were many fewer expected taxa during cold years than hot years--hence O/E values are lower in cold than in hot years. The authors attribute this to the influence of climate on the models, but there is no evidence to indicate this. To make this case, the authors should have shown the list of expected taxa for each site (which would be essentially the same every year), and then show which taxa were not observed in each year. Presumably, there were several "missing" taxa during cold years--if those taxa are warm-water taxa, then the authors have evidence to support their claim that these patterns resulted from the fact that the RIVPACS models were developed during abnormally hot years, and therefore "learned" to expect warm-water taxa wherever and whenever they were applied. I found a list of cold- and warm-water taxa (Table F4-1), but it doesn't provide the information just described. I was unable to find which specific years were classified as hot vs cold, but examination of Figure F6-6 suggests there are not strong differences in EPT richness in hot vs cold years.

In summary, the few analytical methods for which concise and detailed descriptions could be found appeared to be appropriate, but there was often a disconnect between the authors' interpretations and the underlying evidence presented. In other words, the results presented—occasionally in great detail—were were often weak evidence, whereas results that could have provided stronger evidence were often absent entirely (as I argue with the RIVPACs results above).

4) Does the main body effectively capture the more detailed information?

The main body is a maze of sections that seem to contain redundant and fragmented information. The document is very difficult to read and extract the relevant information from a single location. In order to get the "whole story" about each site, the reader has to sift through virtually every chapter and appendix.

I believe the main body needs a drastic reorganization. Like it or not, this study is a collection of "stories" about specific places. The pieces of the story for each site, if concisely, clearly, and logically presented, provide the necessary evidence for the interpretations and conclusions. An organizational structure focusing on sites may be much more effective than the current approach. In my attempts to fully understand the information obtained and analyzed and the results for the Weber River site, I literally had to pour through every section of the document and extract tidbits. I would have preferred to read through a single section in which all the evidence was presented in an orderly way to tell the story about an individual site. The section on each site should be organized in a consistent manner. This is just a suggestion, but I must admit that the current organization didn't work for me.

5) Freshwater Biological Traits database?

The traits database is potentially a very valuable product from this effort. The authors brought together trait information from a variety of sources and made that information much more accessible than is currently possible. In my opinion, it should not be included as an appendix to the report, mainly because the report already has too many appendices, among which the important trait information would be lost. The traits database and web portal should be a separate entity that can stand on its own. Besides, other than the addition of temperature-based traits from the current study, the trait database itself has far broader utility than studies in climate change. Rather, it's a tool that managers and researchers can use for a wide variety of environmental studies.

6) Next steps relative to EPA and state programs?

It is difficult to provide an enlightened response to this question. In my opinion, the report still has several significant deficiencies, particularly with respect to sufficiency of evidence, clarity of presentation, and readability. I have provided suggestions to improve the evidence for climate-related biological responses. I have also provided suggestions for organization and readability. Provided these weaknesses are rectified, I believe the Summary for Managers section could be pulled out and summarized in a 4-6 page fact sheet that targets the audience of all practitioners of biological assessments. Such a fact sheet would need to look very different than the current Summary for Managers section. It would need many fewer major messages (e.g., 2 or 3), a few very simple graphs and figures that drive the messages home, and a case study (or two) that brings the messages "down to earth."

In its current form, the report is far too technically dense and disorganized for any particular audience. In addition, its tendency for overreaching conclusions despite weak (or poorly organized) evidence, is likely to make this important work a target for those who politicize EPA efforts on climate change.

7) Public Comments

I have read the public comment provided and have nothing further to add.

Literature Cited (should be incorporated into revised report)

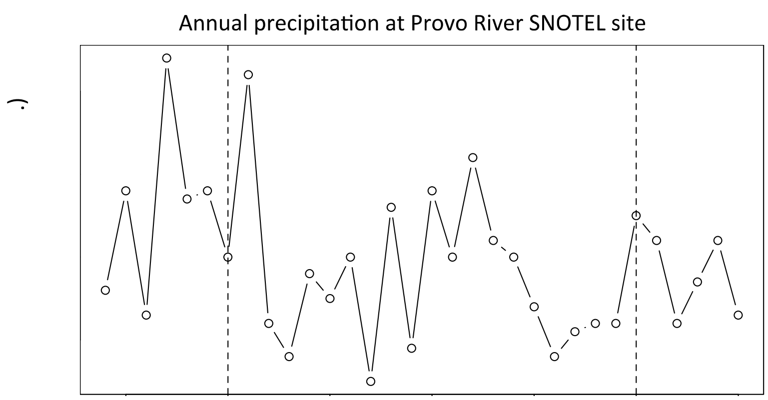
Pilgrim et al. 1998. Stream temperature correlations with air temperatures in Minnesota: Implications for climate warming. JAWRA 34:1109-1121.

Werhly et al. 2009. Comparison of statistical approaches for predicting stream temperatures across heterogenous landscapes. JAWRA 45:986-997.

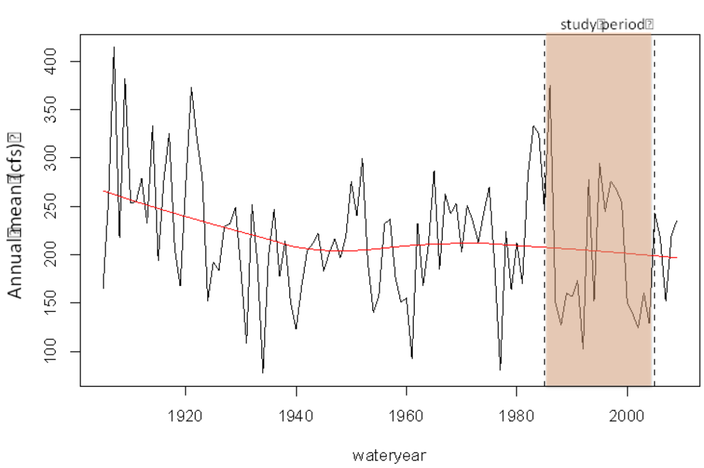
Yarnell et al. 2010. Ecology and management of the spring snowmelt recession. BioScience 60:114-127.

Yuan, L. 2007. Using biological assemblage composition to infer the values of covarying environmental factors. Freshwater Biology 52:1159-1175.

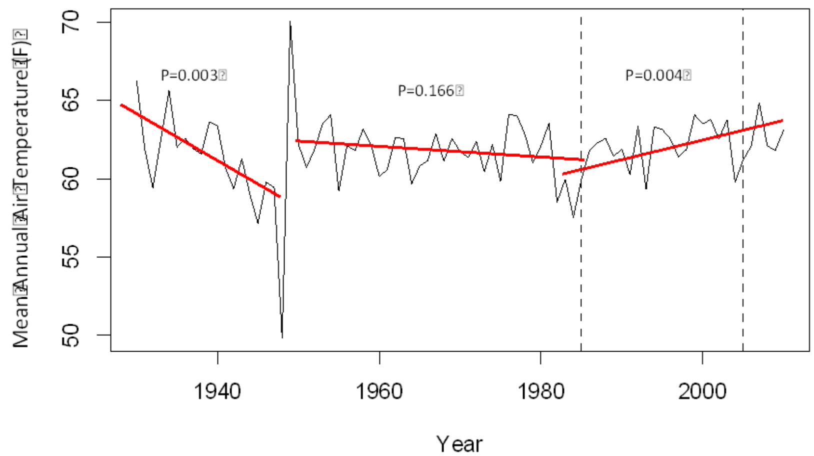
Attachment 1: Annual precipitation at SNOTEL site at Trial Lake, Utah, in Provo River Drainage (adjacent to Weber River basin). Brackets depict focal study period at Weber River site.



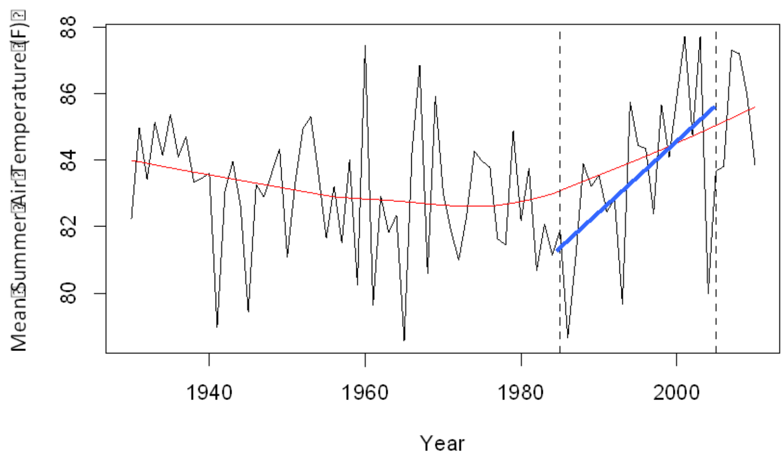
Attachment 2: Mean annual streamflows (cfs) at USGS gage 10128500, on the Weber River, Ut. Trend line is lowess smooth. Shaded bracket indicates focal study period for Weber River site.



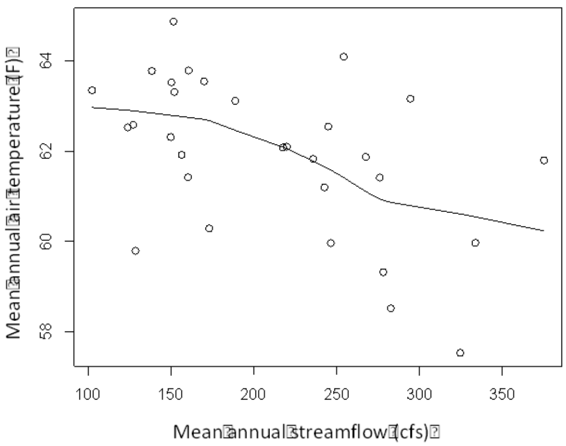
Attachment 3: Mean annual air temperature at Coalville, Utah (10 miles from Weber River site). Data from Utah Climate Center. Trend lines and P-values depict test for linear trend across years overlapped by each line. Brackets indicate focal study period for Weber River site.



Attachment 4: Mean summer air temperature at Coalville, Utah (10 miles from Weber River site). Data from Utah Climate Center. Red trend line is lowess smooth. Blue line is linear trend from 1985-2005, which is also indicated by brackets.



Attachment 5: Relation between annual air temperature and annual streamflow at the Weber, Utah site.



COMMENTS SUBMITTED BY

M. Siobhan Fennessy, Ph.D.

Review of the EPA report: Implications of Climate Change for Bioassessment Programs and Approaches to Account for Effects.

Siobhan Fennessy, Kenyon College

The primary goal of this study is to 'determine what components of bioassessment programs are threatened by (the stressor of) climate change'. Sections 2 – 4 of the report describe the data analyses and models used to address climate change effects, and Sections 5 – 7 summarize the state or ability of bioassessment programs to detect climate change, and make recommendations for environmental management. The comments that follow address the questions posed to reviewers, followed by editorial comments on the report.

1) **Based on your knowledge of bioassessment/biomonitoring programs, biological indicators, and climate change effects, please comment on the report with respect to:**

    a.) **Providing sufficient technical evidence to support programmatic modifications to address climate change effects; what additional steps, data, or analyses would improve the evidence?**

    b.) **The main factors to consider in order to transfer or apply insights from the four states in this pilot study to other state programs**

The authors have used a variety of analytical methods to address many of the key questions related to bioassessment and climate change effects (question 1a). Table 2-1 is helpful in summarizing statistical approaches. However, in many instances the body of the report does not include enough detail on the methods used or the results obtained to make a thorough evaluation possible of the analyses (question 1b). If the goal of this report is to provide sufficient technical evidence to support programmatic modifications to address climate change effects then the evidence should be presented more clearly and comprehensively within the body of the text itself. Sorting through the appendices to find relevant data is essentially impossible. Many of the findings presented are based on data from only a small number of reference sites in one or two states; it is unclear why more data could not be used. In addition, the authors of the report use similar types of data sets from the different states, but different analytical methods are used to analyze them (e.g., ANOVA, multivariate techniques, etc.). This makes comparisons of the findings difficult. While the rationale for how the data were treated is provided in the appendices, this should be included in the report. Section 2-3 does a much better job than earlier sections of laying out the logic of the approach used in the data analysis presented in this section, greatly improving the flow of logic and clarity of this section of the report.

Finally, the report would benefit from a quantitative summary of findings across sites and states to make clear what the 'weight of evidence' is with respect to the study questions. As it stands now it is difficult to synthesize the overall message of the report.

The comments below address other issues noted in Section 2 of the report. This is a large section and, as it stands, is difficult to interpret.

Figures and tables should be presented in a consistent format with complete legends (many are incomplete) and with obvious labels on the figures and complete column headings on the tables. They should also be formatted so that they will reproduce in black and white (this is not the case now). As an example, Table 2-2 shows important results of the study, and would benefit from better column headings, such as "years" instead of "groups" (column 2), and 'Cold taxa' and 'Warm taxa'. It would also help the clarity of Table 2-2 if a brief explanation were provided of which taxa groups made up the rest of the community since the cold and warm taxa make up a relatively small % of the individuals in many years (Table 2-2 shows same data as Figures 2-1 and 2-2). In Figure 2-3 – 2-5, why are data shown for only one of the Maine sites? This should be explained; if it isn't the authors risk losing credibility since some data are omitted without explanation. For these (and other) figures, the results of the statistical tests should be presented in the legends.

It is important that care be taken in making recommendations based on the results presented in many of these graphs (e.g., Figure 2-6, 2-7 and 2-8) because, while they show trends in the data, there is no statistical significance in most of these data comparisons. The comparisons of Utah sites by ecoregion, elevation, and stream order are an excellent way to stratify the data, but again, the results of any statistical tests should be included in the figure legends (e.g., Figure 2-6, 2-7 and 2-8). The authors should be clear in their conclusions that while trends in the data are sometimes seen, in some cases there are conflicting trends, and/or the trends are not significant. Finally, no evidence is provided that substantiates the link between air temperature and water temperature. Can data be provided to link the increases in air temperature to water temperature?

At many points in the text, there is not enough information provided to allow interpretation of the results. For example, the specific Indicators of Hydrologic Alteration (IHA) used in the analysis (line 479) are not listed, and on line 482 states that several low flow parameters performed well - performed will with respect to what? There are many other examples, but the text needs a thorough proof read to ensure its meaning is clear to the reader. There are also very few citations in the text relating these results to other studies.

The introduction to the report states several times that four states are the focus of the study, despite the fact that the data from Ohio, and its analysis, are nearly absent. This should either be rectified or explained.

Overall, the authors have done a good job in elucidating factors to consider when applying insights from the four states in this pilot study to other state programs, and have made helpful suggestions on which activities should be a priority in dealing with the response of communities to climate change.

For state programs, it might be helpful to link the expected responses described here to specific predictions on how climate will change in different regions in the U.S. In this way, states will be able to more finely tailor how they will use the insights derived from this report.

2) Has EPA pulled out the most important findings in the Summary for Managers and Policymakers (SMP) from the technical findings? If not, what findings do you suggest should be highlighted?

The Summary covers the key findings of the report and is a good summary of the contents of the report. However, the conclusions as presented are much more certain than the data support. For example, many of the conclusions reached about shifts in taxonomic composition as a function of air temperatures appear to be based on five metrics that showed significant differences at two sites in UT (Weber and Virgin, lines 631-634). This isn't enough evidence to conclusively support the conclusions reached, the summary must reflect the ambiguity in the data. For example, lines 696 – 699 in the discussion of ME data, where there was a lack of association between ecological traits and community response. The results presented in many sections are either not significant (line 783 for example) or are conflicting, which would seem to limit the ability of the authors to make strong conclusions. Generally, there is a tendency in the report to over-interpret the results.

3) Are the analytical methods used appropriate for the questions being asked and the available data? Are the strengths and weaknesses of the methods used in the report fully described and accounted for?

Generally the analytical approaches seem sound, and the description presented on Table 2-2 is very helpful. However, a huge variety of approaches are used, with no rationale presented for why one method was used in one section and a different method in another. The choice of methods must be made clear (and while some explanation is given in the appendices, it should also be presented in the main body of the report). In some cases the results of an analysis are not presented completely, for example, data presentation is incomplete for the CCA and NMDS analyses. For example in Figure 2-15, information on the CCA axes (axis 1 and 2 at minimum) should be provided. A brief explanation of why the different approaches were used should also be presented - why was CCA used for Utah data and NMDS used for NC data (Figure 2-16)?

Section 2-5 details 'other sources of potential spatial confounding' (not a very clear subheading, by the way). The issue of changing land use and land cover through time is an important one, and the authors have been thorough in addressing this issue in this and other sections of the report. However, in the analysis only land cover within a 1 km buffer of the reference sites was used to indicate anthropogenic influence. A rationale for this buffer distance should be provided. Is this a common approach in stream studies? Should an analysis of the upstream catchment area be used instead? The upstream catchment area would seem to have a more direct influence on a given stream segment.

In some cases more explanation of the magnitude of climate change necessary to generate changes in biological indicators would be helpful. For instance, for the analysis presented in Figure 3-5, what change in air temperature would be needed to generate this shift in water temperatures, and so drive the biological response modeled here.

And again, early in the report it states that data from four state programs are used in the report, however, very little attention is given to Ohio data in the subsequent analysis. While this is touched on in the appendix, it should be addressed in the introductory sections with some explanation of why this is the case.

4) Does the main body of the report effectively capture the more detailed information presented in the appendices and if not, what are your recommendations for improving the report or appendices?

The report would be clearer as a stand-alone document (i.e., without benefit of reading the appendices first), if a more complete introduction to the methods used and rationale for site selection were provided (for example, why were the data from some stations included and not others, pg 2-3). The final conclusions of the report rest on the choices that were made with respect to sites included and analytical methods used, so elucidating how they were made is crucial to interpreting the report's conclusions. This is also true for the manner in which data is presented. For instance, results based on Utah data, presented in section 2.2.1 (line 240) are given with little explanation of why these reference stations were used. A table of the different stations with the duration of data records (as shown in the appendix) would be a welcome addition to the report. As an editorial comment, this would also be easier to read if the names of the sites were provided with the station numbers in parentheses, instead of vice versa. Likewise, a more full explanation of the development and use of "scenario-metrics" is warranted (line 229).

The report would also benefit from development and presentation of conceptual models of expected ecosystem response (what are the specific predictions about how different indicators within a MMI will respond, for example, these would be related to the questions posed on lines 125-132), and at some point in the report, a summary of findings across sites and states to make clear what the 'weight of evidence' is with respect to the study questions. For instance, at line 1016, the authors provide a comparison of trends in the data, but in other places in the report, the conclusions reached based on these seem to imply much more certainty in the results.

5) The Freshwater Biological Traits Database was pulled out as a standalone report. Please comment on the merit of this report on its own vs. as one of the appendices of the main report. What next steps would you suggest for this report and the database?

This will be very valuable as a separate document to many practitioners in the field, including those working in State and Tribal programs, so I would say that making it available as a stand-alone document/database is valuable. I imagine there are many who could make use of the database who will also want to reference the main report's findings, but since the database is available online and it is intended to be a 'living' database, it holds promise as a valuable research tool as the issues of climate change become more pronounced. This is particularly true as trait-based approaches to MMI development and use (i.e., the development of trait-based indicators) are increasingly important as an alternative to taxonomic based approaches. A database of this sort will facilitate such research efforts.

6) What next steps would you suggest, based on this work, as being the most informative for EPA's Office of Water and state bioassessment/biomonitoring programs to address climate change effects?

A concerted effort to establish protocols to gather the data needed to more fully understand the effects of climate change. Priority should be given to several of the recommendations provided in the report, including protection of reference sites and the establishment and priority sampling of sentinel sites. The findings on the vulnerability of reference sites to land use changes is perhaps the most surprising finding of the report, and one of the most important for bioassessment programs generally. The establishment of sentinel sites that could be sampled more frequently is also an important recommendation in trying to build data sets that are more suitable for detecting temporal trends in the data. One excellent suggestion is that perhaps States and Tribes could cooperate in the establishment of sentinel sites within ecoregions (for example) that cross political boundaries. This could increase the efficiency of monitoring for climate change effects (as long as sampling methods could be agreed upon). Finally, the emphasis on traits-based analysis will help build biological indicators across regions since traits are not location specific as species tend to be.

7) Please comment on the public comments submitted for this draft report. Specifically, which comments should or should not be addressed in the final draft? (public comments will be compiled and provided to reviewers by ERG)

One public comment was received. The comment states, in part, that when… "' found' data (e.g. state data sets) are to be used for the pilot studies, some discussion of the methods and data comparability issues between the data sets and the data synthesis approach applied to address these issues should be included in the document." This is a good suggestion and would help with the clarity of the results.

Additional Comments on the Report

(Note to ERG: I realize this is a large report, but it would have been very helpful to have the figures in color, as they must be in the original. The black and white versions of many of the figures make it difficult to differentiate data points. Just a few examples include Figures 2-17, 2-18, and 2-28. In Figure 2-17 for example, it is impossible to differentiate the symbols for normal years from hottest years, so an evaluation of this result is difficult. This is true in many figures in the report.)

Section 2:

Some editorial comments have been made on Section 2 above, but in general the writing in this chapter written in a very passive voice which make it difficult to follow. Section 3 appears to be written by a different author and is more direct and clear.

Lines 1028-1030- this statement is not clear as written, why would a high degree of spatial variability between sites (in terms of stream order, elevation, etc.) reflect 'real' spatial variation in vulnerability to climate change effects.

Lines 1115-1120 – why not carry out an assessment of the responsiveness of the metrics listed here as part of this report?

Section 3:

Lines 145-147- this states that comparing the average of the first 5 years to the last 5 shows an increase of 6 EPT taxa at the Sheepscot site in Maine. However, Figure 3-3 shows that the mean for the last 5 years is heavily skewed by the data from 2005, making this statement much less convincing.

Lines 168-170 – I am bothered by statements like the one here (and found in several places in the report), which infer the authors are looking for a specific outcome. While it is reasonable to suppose that climate change will affect in-stream communities, statements like "the duration of data records was too limited to define significant trends" would be best ended with "if they are present".

Lines 300 – 303 – It would help if the authors could state the goal of this analysis in terms of protecting reference sites (or the lack of ability to do so). If climate change is a stressor and it leads to a decrease in the site condition classification score (Figure 3-5), isn't this a valid result? If there are climate change effects, how else do we document these? See also lines 498 – 500 (and if reference sites decrease in scores, non-references sites should do as well?).

Line 458 – should say "due to" instead of "do"

Line 464 – says "climate related changes in precipitation and flow could have increased cold sensitive taxa" - is there any evidence for this in this? Seems speculative.

Lines 508 –511 – not clear.

Lines 614 – 617 – this isn't totally clear, but it seems this finding conflicts with the data presented in Figure 3-9.

Section 4:

Lines 771 – Most programs conduct their sampling in order to make both spatial **and temporal** comparisons. The 5 year rotating basin approach is designed to assess watersheds, and compile data over time to track trends in stream quality in order to answer questions about the efficacy of environmental management, waste water treatment plant upgrades, etc. See also Section 4, lines 1229 – 1230, unless I am missing something, in most programs temporal patterns figure into spatial considerations.

Line 862 – 16% is a high proportion of urban land use to allow for a reference site. Some previous studies have identified thresholds of 15% impervious surface as the point where stream hydrology is altered significantly.

Line 879 – the absence of dams seems like an important criterion for reference sites, Does the land use analysis include this?

Line 902 – not clear how climate change influences reference site vulnerability – vulnerability to what exactly?

Lines 915 – 920 – don't MMI scoring schemes also allow a range of biological conditions that have regulatory significance to be identified? The case for BCG should be made more clearly here.

Lines 1031 – 1032 – not clear why different thresholds were used in evaluation of these reference sites.

Line 1058 – the reference to Figure 4-3 here appears to be incorrect?

Line 1118 – 1120 – not clear, if there is a shifting baseline problem for the reference condition, how or why would this be helped by measuring biological condition in comparison to pristine sites instead of present-day reference. Not clear how this could be done, and if so wouldn't pristine sites shift in response to the stressor of climate change as well?

Section 5:

Lines 1263 – 1265 – non-significant patterns could also be due to the fact that the patterns aren't significant. Looks like the authors are looking for a specific outcome. More support for this point is needed.

Lines 1421 – 1422 – the data that have been presented are not as definitive as this statement. The authors should take care not to make claims that the data don't support.

Line 1484 – says "establishing statistical differentiation." Not sure what this means.

Lines 1517 – 1519 – an example of how this might work would be helpful here.

Appendices:

General comment; while it is great to see detailed reporting, the appendices are a jumble with many different data sets used in different ways in different locations. For that reason they aren't nearly as informative as they could be if they were presented with a common format and a consistent approach.

COMMENTS SUBMITTED BY

Eric P. Smith, Ph.D.

General Comments

The purpose of this report is to evaluate the potential effects of climate change on benthic monitoring and its use for evaluation of water quality. The researchers used four sampling programs to try to find consistent changes in macroinvertebrate indices associated with changes in climate. This type of data is what I call “found” data in that it was not collected to answer the question of interest. Therefore my expectations are low as there are confounding factors that cannot be easily controlled. While the researchers have done an admirable job of trying to address a difficult problem there are several issues that severely limit the utility of the work and its conclusions

1. There is an excessive focus on the use of p-values and significance as the measure of the importance of a relationship between a benthic metric and climate change. I suggest the authors consider using an approach that is more connected to the science and expected changes or look at other statistical measures such as slope and size of correlation. At the least, there should be an adjustment of the p-values for the number of tests that are run.
2. Climate change has many components and expected changes. The authors seem to focus on defining climate change as temperature change. Aren’t there other components that merit evaluation such as precipitation in the east and fire in the west? Correlation analysis also is susceptible to spurious relationships. I suspect correlations that are based on temperature are confounded with urbanization. It will be hard to argue that the changes are due just to temperature.
3. A further difficulty is the temperature metric is not well connected to the site since it is not directly measured at the site. Use of average temperature seems somewhat arbitrary. Even if temperature data are available it is not clear how to combine the information in temperature into a metric that is connected to benthic metrics. Temperature data tends to be measured over a short time interval while benthic samples are taken in either spring or fall. What is the appropriate temperature metric? If PRISM data is used, uncertainty associated with the resulting temperature values need to be discussed and possibly estimated. How good are the estimate temperatures from Prism?
4. In my opinion, the authors overstate the conclusions associated with the data analysis. With found data and observational data one must be careful to avoid the trap of correlation versus causation. Many of the conclusions are obvious and not a result of the analysis but seems to be prior views on climate change. I don’t see where the support is for the statement that benthic metrics are vulnerable to climate change. Is this based on 1 significant correlation?

Questions

1) Comment on

a.) Does the report provide sufficient technical evidence to support programmatic change?

Yes and no.

The evidence of change in metrics due to climate change is not clear. First, the correlation between average air temperature from PRISM and benthic metrics is sometimes moderate in size but there is not a consistency to the correlations. Patterns that occur at one station do not seem to also occur at other stations

(Table 2-6, for example). Second, the project is based on data collected for a different purpose and confounding factors limit the analysis. It could be, for example, that the correlation with temperature is due to a connection with elevation (if one replaces space for time) or a factor that also varies with time. Other factors, for example, acid rain in the east, might also confound the analysis. It might have been a stronger study if the authors first reviewed the literature, postulated changes in benthic metrics then observed the changes.

There seems to be a general lack of checks on assumptions.

The report does point out some inadequacies with current designs that while known for years have not changed. Some of these inadequacies have been known for a while (the need to sample some sites repeatedly over time in order to assess trends).

b.) What are the main factors to consider in order to transfer insights from the four states to other states?

It is a difficult question to address as the report requires a major rewrite. It is difficult to read and to connect the conclusions with the evidence. I do not see strong evidence for the insights that are presented and many are a priori insights.

Obviously there is a need to measure water temperature and to design a temperature exposure metric.

In general, it would be wise to keep the raw data, not just the metrics. This way, future metrics can be developed if needed.

2) Has EPA pulled out the most important findings for managers and policymakers? If not what else is valuable?

The summary suffers from a desire to find a significant conclusion when there does not seem to be one. I think it is bad science to run hundreds of tests then draw conclusions from the statistically significant ones, especially when there are not clear patterns.

Again, I think more work is required. I especially did not like the frequent use of the term “vulnerability” as this is both a general term as well as a term from the climate change literature. I am not sure what it means to say the multimetrics are vulnerable to climate change. In the climate change literature vulnerability involves exposure, sensitivity and adaptive capacity so should the authors have evaluated all three of the components? Do you really mean metrics are sensitive to climate change or that climate change will confound conclusions from studies that involve reference metrics?

One question that is not really addressed is “do decisions about the quality of the environment change”? If the organisms are responding in a negative way then one would expect to see false positives (declare impact when there is none) increase. It is not clear that this is the case.

3) Are the analytical methods appropriate?

The data come from existing studies that were not designed to assess climate change. Generally “found” data is difficult to analyze as it contains confounding factors, inadequate sample sizes for the effects of interest and irregular coverage. Found data often results in confirming obvious results and less obvious results tend to be inconsistent.

Most of the data that are used in the examples is based on sampling programs inspired by the EMAP program. The sampling here is intended to focus on status not trends. It is therefore not surprising that trends associated with climate change are difficult to detect and assess.

Correlation analysis using Pearson correlations require linearity and independence for the measure to be useful. The authors seem to use this measure without checking these assumptions.

Most of the results simply confirm the obvious. Specifically, elevation, area and stream length are important. Climate change seems to be temperature. There is a lack of analysis related to hydrologic conditions and other expected changes due to climate change. The authors should be more explicit about what they mean by climate change.

The analyses seem to give weight to results that are statistically significant. Given the number of tests that are made and the relatively low correlations that occur one has to wonder if any of the results are meaningful.

There are quite a few confounding factors that make the results of the analyses rather difficult to interpret. For example, in the CCA analyses, shouldn't elevation be accounted for – ie. A partial analysis might be appropriate.

It seems that the view is that climate change affects only the reference sites. It will obviously affect the stressed sites as well. Is the real question the extent to which the stressed sites will be detected (i.e. are error rates affected)? Thus the focus should be more on the change in the power of the reference condition approach.

The focus is also on means. In particular, the main effect of climate change that is addressed is associated with a rise in temperature. What is likely to be important is the variability that will increase, especially initially.

While it is useful to develop an idea of temperature optimum for different taxa, the focus is on the optimum not the tolerance (i.e. the variance in response). The tolerance is rather critical and can be used to address some of the questions of interest. For example, if the tolerance of most taxa is small, this would suggest that temperature change will have a strong effect. On the other hand, if the tolerance is quite large then there may be a minor effect.

The writing is inconsistent and I found it difficult to follow some of the material without looking at different sections of the report. Graphs also show inconsistencies, perhaps due to the fact that some were intended to be in color rather than black and white. I think the authors should use standard practices for reporting information when testing hypotheses. In particular, sample sizes are inconsistently provided.

4) Does the main body capture the details in the appendix?

The appendices are a bit difficult to navigate and need better linkages to the main body of the text. I have commented on them separately below.

5) The Freshwater database was pulled as a standalone report. Comment on its merit

Having a traits repository is valuable. The only comment I have is that the traits are a mix of physical and non-physical. For example, temperature is included. If potential stressors are to be included, then the database should be expanded to include other information that is available such as nutrients, heavy metals, etc.

6) What next steps would you suggest?

To adequately evaluate temperature change effects it seems that there is a need to have information on air and water temperature. The cost of this information is relatively cheap. Bioassessment has been viewed by too many as something that is separate from chemical programs. That biology and chemistry are not sampled together is reducing the effectiveness of both programs.

1. Figure out how to get states to collaborate with other states on sampling dates and methods so that data from different states may be combined. There should be more encouragement to simultaneously monitor many of the same sites.
2. Figure out how to get agencies within states to collaborate so that chemical data is collected with biological data. There is a missed opportunity to enhance sampling through cooperative programs. There are likely numerous reference sites in national forests and parks. This study needs to stress the importance of cooperative sampling.

7) Comments on the public comments

Via email text: I took a look at the public comment and do not have anything to comment, so do not need to change my document

Appendix A

1. Obtaining mean annual temperature by averaging maximum and minimum temperatures is what is called a “quick and dirty” approach. This approach has to have high variability.
2. Some plots (line 84) exhibit seasonality. P-values based on simple tests are not appropriate in this case.
3. Line 106 – this figure suggest two groups of temperatures. Why?
4. If the authors are not going to check assumptions then use Kendall’s correlations.
5. Why compare correlation coefficients to assess trends? Why not compare slopes or change in temperature?
6. Figure A-9 is based on daily maximum temperatures. Why are the data sparse in the later years? Why not use other methods i.e. the max over the year or the average or median of the daily maximum values? The R2 is diminished by the annual variability. Does it make sense to use R2 in this case? Report sample size.
7. Is there any sense that the inferred temperature method actually works? Isn’t there variability with the method that should be reported?
8. Isn’t trend in inferred temperature confounded with other factors (urbanization, for example).

Appendix B

The authors correctly point out the difficulties in using data from different agencies that have different methods and approaches.

Appendix C

1. Identify the method for clustering and the distance measure.

Appendix D

1. Benthic data are usually collected once each year. Temperature data may be available only once or at many times, possibly at a different location. How good are results given the need to summarize temperature? Do benthic taxa respond more to maximum temperature or average. Does it not depend on the taxa and perhaps time of the year?
2. North Carolina seems to have a stronger temperature relationship than Maine. Is this due to the use of presence absence data rather than abundance data? Why are relationships weak in Maine? Does this suggest the methods do not work well?
3. How is average annual air temperature calculated?
4. It would be useful to spell out what is meant by the maximum likelihood estimate as there are several ways to calculate the estimate depending on the assumptions of the model.

Appendix F

1. Why focus on p-values? Sample sizes are rather small so p-values will not be all that useful. Pattern is more important.
2. Where is section 6?

Appendix J

1. What is the curve in J3-2 based on? Isn't there a state effect? RI for example do not have any low density values? What are the dashed lines for?

Miscellaneous

1. Why are Kendall correlations only sometimes used?
2. The approach seems to be: run tests, interpret significant results. There seems to be little consideration to the assumptions of the tests. Without checks on assumptions, the results cannot be defensible.
3. Sample sizes are only occasionally given. They need to be presented with all the tests and analyses
4. Grouping data together increases sample size but also creates hidden biases when sample sizes, variances and confounding factors differ amongst stations.

COMMENTS SUBMITTED BY

R. Jan Stevenson, Ph.D.

Review of Hamilton et al. (draft), "Implications for Climate Change for Bioassessment Programs and Approaches to Account for Effects"

By R. Jan Stevenson
Michigan State University

1) Based on your knowledge of bioassessment/biomonitoring programs, biological indicators, and climate change effects, please comment on the report with respect to:

a.) Providing sufficient technical evidence to support programmatic modifications to address climate change effects; what additional steps, data, or analyses would improve the evidence?

Summary Question 1a.

This report and other scientific evidence cited in the report clearly provide sufficient technical evidence to support modifications in bioassessment programs to address climate change. We know that climate will be changing by several degrees centigrade with changes in drought and flood disturbances. There is considerable evidence in the report that invertebrate metrics respond to temperature, those metrics will respond to the projected temperature increases expected with climate change, and changes in metrics resulting from climate change would be sufficient that assessments of biological condition could change from unimpacted to impacted without changes in land use. In addition to the report, other research shows temperature has great effects on invertebrate physiology, which likely varies greatly among taxa. Thus, invertebrate metrics could respond to changes in temperature because invertebrate species composition is likely to change. There is evidence in the report at a couple sites in four states' data that historic temperature change that could be related to climate change may have affected invertebrate species composition and metrics. The evidence that inverts have responded to hydrologic alterations as projected by climate change is weaker than the evidence for temperature. While evidence in the report could be made stronger with better explanation, more thorough data analyses, and greater integration of scientific literature on invertebrates response to climate change variables, responsible resource managers should take action to account for climate change effects on biomonitoring programs because multiple lines of evidence show a reasonable probability that climate change effects on bioassessments will complicate bioassessment of local management of land use and resulting pollutants and habitat alterations. In conclusion, the science in the report shows that climate change could affect invertebrate assessments in streams; and importance of programmatic modifications to state programs is evident from the relatively poor data available to evaluate effects of climate in the four states' data.

Overall, this is good report. In particular, the data analysis is extensive and supports multiple lines of evidence that climate change will affect bioassessment results and management. I am not convinced that the analyses were as thorough as they could be. Although many analyses were run, it seems that different statistics and variables were used with data from different locations. Analyses should be added to fill the gaps and present similar analyses, using similar variables, and for each region. The proposed adaptations for management to account for climate change effects will be useful, but they will not be sufficient to address the full range of considerations necessary for managing local effects on watersheds that will be exposed to climate change. A broader conceptual framework is needed to more completely address the needs for management adaptations and adaptations to fill those needs.

Future work should: build a policy-management-assessment framework to guide the development of tools (see below); expand on the lines of analyses already conducted to have more equal and uniform treatment of all datasets; gather more data and evaluate direct stressor effects on invertebrates and how effects vary among groups, which will build on the traits database; and expand the invertebrate traits and stressors considered when thinking about effects of climate change (e.g., droughts and flood stress as well as temperature); include more rigorous cause-effect analysis of invertebrate responses to stressors, such as temperature, drought, and flood stress. In analyses of responses, consider the temporal lags in response of differing metrics, relative to stressor. The report should include rigorous, yet broadly applicable, conceptual and analytical tools for resource managers.

Specific comments related to the report and Question 1a:

There is evidence that effects of temperature change on invertebrates as projected for climate change may have been observed in analyses at a couple sites in four states' data.

The evidence that inverts have responded to hydrology alterations as projected by climate change are weaker than the evidence for temperature.

b.) The main factors to consider in order to transfer or apply insights from the four states in this pilot study to other state programs.

Summary Question 1b.

The main factors to consider for transfer and application of insights from the four states to other state programs should be: 1) a framework for how climate change effects on ecological condition can be incorporated into management of state waters; 2) a conceptual model for how current management problems, which I will refer to as human activities related to land use (including generation of point and non-point source pollutants and physical habitat alterations), and climate change independently and interactively affect proximate environmental factors (e.g. temperature, droughts, floods, nutrients, sediments, and dissolved oxygen) regulating taxonomic composition and metrics of benthic invertebrates; 3) research program designs to determine predicted effects of land use and climate change effects on ecological condition; 4) monitoring program designs to assess independent and interactive effects of land use and climate change on ecological conditions; and 5) analytical tools for determination (3) and monitoring (4) and causal analysis of land use and climate change effects on ecological conditions. These elements are spelled out in a couple chapters that I lead in Barbour et al. (2004), i.e. Stevenson et al. (2004a and 2004b). From this list, 1 is lacking, 2 is incomplete, and 3-5 are developed but are largely focused on temperature.

If the main factors to consider for transferring and applying insights are defined as the options for proposed programmatic modifications, then the proposed programmatic modifications are appropriate, but more analysis and testing are needed to determine which programmatic changes are appropriate under different funding availability and management goals. In a few cases, the emphasis on a specific subset of programmatic comments overemphasizes their importance within the full set of possible programmatic modifications. For example, the importance of having a long-term reference monitoring network (Sentinel Monitoring Network) is emphasized throughout the report, but not until Section 4 is the balance between long-term monitoring reference and non-reference sites discussed. There are challenges for transferring results, such as predicting which streams will be most sensitive to climate change. However, it is clear that many of the recommended adaptive management responses are possible, can be used by all states, and are important to adopt.

A sound conceptual foundation in ecological first principles will increase transferability of results across regions and time. Rigorously develop the conceptual model of temperature effects on bioassessment programs, but also include the other pollutants and habitat alterations that will occur with climate change and how they interact with land use change. Although this information seems to be present in the report, it is not organized in a way that it will be easy for policymakers and managers to understand. Elaboration of conceptual framework for management land use change and climate change as well as a conceptual model for climate change effects on invertebrate stressors and invertebrate responses to those stressors would help organize the report, future research/monitoring, and implementation of results in management and policy.

Specific comments related to the report and Question 1b:

2.197-222. Developing invertebrate traits based on their sensitivity and tolerance to temperature is certainly a good first step for understanding effects of climate change on invertebrates, bioassessment programs, etc. However, at this point in the report, I'm not sure what to expect the effects of elevated temperature, floods, and drought to be on invertebrates. For example, is temperature a resource stressor (*sensu* Stevenson and Sabater 2010) and really has no direct negative effect on invertebrates unless it affects dissolved oxygen or another parameter? Or do high temperatures have direct negative effects on invertebrates because they denature enzymes? Having a thorough conceptual model explained at the first principles of temperature and other stressor effects on invertebrates early in the report shows how the analytical approach is organized, enables placing results in context, and help readers remember, understand, and believe results. This is particularly true when sometimes you see effects/relationships and sometimes you don't. Variability in results is to be expected and should not compromise the certainty of the findings. Such an approach, with clear *a priori* hypotheses, will also mute criticisms of repeated hypothesis testing and outline ways to control and explain attained significance (p values) with multiple hypothesis tests.

2) Has EPA pulled out the most important findings in the Summary for Managers and Policymakers (SMP) from the technical findings? If not, what findings do you suggest should be highlighted?

Summary Question 2:

Yes, for the findings resulting from the analyses conducted in this report. Evaluating all possible effects of climate change and interactive climate and land use change in one report is probably not practical. Clearly defining what is done in this report and what has not been done thoroughly should be clearly spelled out, so the later can be addressed in future work. Again, one practical way of doing this is starting with a management framework that includes independent assessments of climate and land use change effects and a conceptual model of how they affect ecological condition will provide the conceptual landscape in which the path of this report can be mapped and better appreciated. This should be done in the introduction of the report.

An important technical finding that is not clearly explained is managers need to think about new ways for characterizing reference condition, diagnosing stressors, and developing water quality criteria with increasing effects of climate change on current bioassessment and management paradigms. This calls for a policy-management-assessment framework that explicitly includes climate change. Again, the many ways climate change could fit into the policy-management-assessment framework should be treated more clearly. Although elements of this policy-management-assessment framework are discussed throughout the document, they are not pulled together.

Specific comments related to a policy-management-assessment framework that includes climate change:

An important technical finding is managers need to think about new ways for characterizing reference condition, diagnosing stressors, and developing water quality criteria. Throughout the document, climate change is ambiguously treated as something either to treat separately from reference condition or to account for with reference condition. In other words, is a climate altered system still in reference condition? Of course not. If the USEPA is treating C emissions as a contaminant causing climate change impacts, then one of the impacts (in addition to human health and direct risk to well being) is degradation of ecological condition, including both water quantity and quality as well as biological condition. The report needs to deal more explicitly with how climate change could be treated relative to management goals.

The challenge with treating climate change as just another set of stressors that humans cause is that management of climate change is largely out of the scope of local managers. Mercury contamination from fossil fuel emissions is similarly a transboundary and global problem calling for a modified reference approach. Of course there is the "act locally and think globally" philosophy in which local activities are hoped to have global effects on climate change, but managers have much more direct impact on local land use and best management practices than climate change stressors. This potential for management must be recognized explicitly in the policy-management-assessment framework.

Potential for local management with the inevitability of climate change will require a "moving target" for management, or a moving management goal (aka expected condition (*sensu* Stevenson et al. 2004), Figure 1). Climate change impacts, locally manageable human activities and stressors, and natural variability represent three suites of factors that could be used in models, as has been described in the report. But they were not emphasized sufficiently as part of the management adaptation strategy. Climate change impacts could be treated, relative to local management goals, as natural variability is treated. In the US, we are usually trying to manage waters as minimally disturbed. If we can only control local land use[1] with best management practices (point and local non-point sources of contaminants), then effects of climate change can be treated as natural factors, i.e. something we need to account for when establishing our expected condition for management, but not something that we can really manage. That does not mean that we should not assess effects of climate change, or consider climate change and land use change interactive effects on ecological response, just as we don't neglect interactive effects of land use change and natural factors.

The key distinction here is isolating climate change management and local management of contaminants and habitat alterations that fall under the policies of the Clean Water Act (CWA).

We should, therefore, assess both deviation in observed condition from the historic reference condition and the current[2] expected condition (I don't want to start calling this the current reference condition because that generates other debates). We should assess condition at a site as the deviation of observed minus expected conditions (Stevenson et al. 2004a, 2004b, and others), both historic reference condition (when possible) and current expected condition. This is very similar to the concept of tiered aquatic life uses, where we recognize natural as the anchor to the generalized stressor (human disturbance) gradient, but we don't expect to manage for natural conditions.

So how do we define these two expected conditions, historic reference condition and current expected condition. We generally think of the structure and function of streams being ultimately regulated by climate and geological factors that regulate hydrology, soils, and land use. Naturally varying climatic and geological factors are often captured in "predictive models" to define expected species or metric values at a site. So current expected condition at a site is often determined as a function of elevation, latitude, and longitude (often proxies for climate), precipitation, mean annual temperature, soil permeability, groundwater loading, etc. We can measure these things now, and use models as we currently use to predict current expected condition for a site. Historic reference condition can either be documented now or modeled later, based on what we know about climate change for different regions and responses of stream structure and function to those climate changes.

---

[1] By land use, I mean human activities in general, in urban and rural, agricultural and forestry, built and unbuilt conditions.
[2] Current here means the condition at the time of assessment, now or in the future.

Key point: we should explicitly relate expected condition to the deviation between the current expected condition (with current being some specific time either now or in the future) and reference condition that was defined today (or sometime in the past).

So what is the alternative? Can we have a management goal for protecting systems in the minimally disturbed condition as occurs in 2011 climate condition (which is already somewhat altered, to an extent that we cannot yet assess without future data)? No. We'd like to, but from a local perspective we cannot. So distinguishing effects of naturally occurring climate and geological conditions, from anthropogenic climate change effects, and both from land use effects is essential for management. Documenting and forecasting climate change effects will be critical for arguments to manage climate change at other policy scales and for adapting to changes in our aquatic resources that we cannot locally manage.

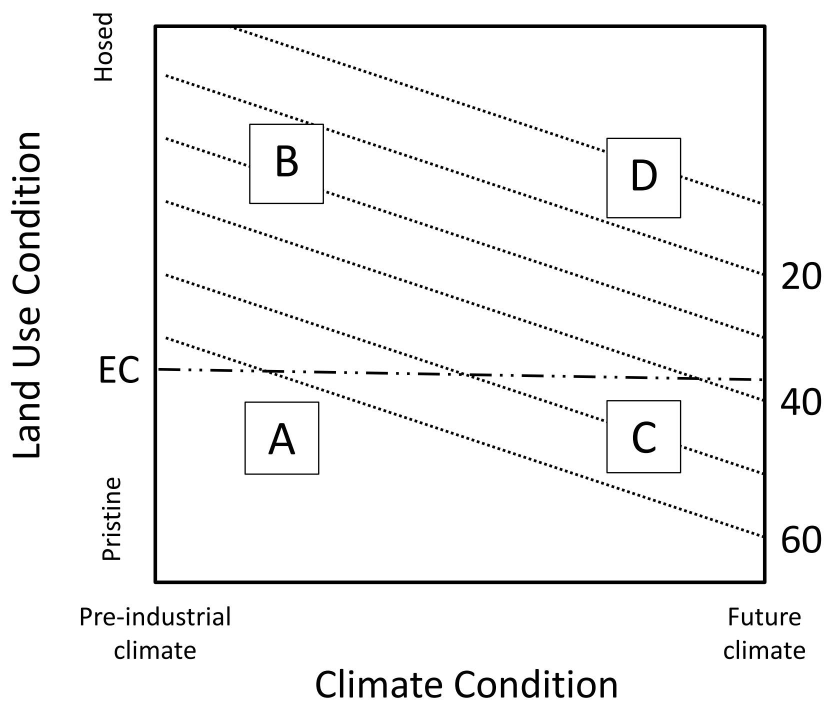


Figure 1. Theoretical relationship between expected condition % EPT (EC) assuming climate change, % EPT, land use condition, and climate condition. Isopleths of % EPT (dotted lines) vary from 10 to 60 across gradients of land use condition ranging from natural (pristine) to highly impacted (hosed) and gradients of climate condition ranging low temperature and extreme event disturbance (pre-industrial climate) to high temperature and extreme event disturbance (future climate). % EPT in high quality waters (i.e. meeting expected condition) is predicted to change with climate change from greater than 60 to less than 40, assuming no change in land use. Minimally disturbed expected condition without climate change (today's reference condition, A) is greater than 60 % EPT, whereas in the future, the same site with minimal land use disturbance could have a % EPT near 50 (C). A disturbed site (B) today has a %EPT around 28, whereas in the future with climate change effects, it could have a % EPT close to 10. Contact me for more explanation if you are interested.

3) Are the analytical methods used appropriate for the questions being asked and the available data? Are the strengths and weaknesses of the methods used in the report fully described and accounted for?

Summary Question 3:

The analytical methods used were appropriate for the questions being asked and the available data, but as with many elements of the report, their presentation was so unpredictable that you often found the justification for an analysis at a different spot from the analysis. So it was often hard to follow. But with limited data come limited options for analysis. A more rigorous application of causal analysis (aka Beyers 1998) would have been valuable for when only one site was analyzed at a time would have been beneficial. The challenges of the data were clearly stated. Given limited long-term data and data at reference sites, the analyses were appropriate, but in some cases they pushed the bounds.

The traits analysis should include a causal analysis element, which will be describe below.

4) Does the main body of the report effectively capture the more detailed information presented in the appendices and if not, what are your recommendations for improving the report or appendices?

Summary for Question 4:

The main body of the report was not as effective as it could be. The organization of the whole report is awkward, but that could be my lack of understanding about who will read the report. The executive summary was too long without clear messages highlighted in bold and plain language. I really liked the summary for managers and policymakers, because of the finding/evidence/adaptation-response section in each. Chapters 2-5, however, which relied on reports of more detailed analyses in Appendices, did not have sufficient detail to understand without using the appendices, from a scientists' background, which often brought credibility of results into question. Chapter 2 is substantially more confusing than Chapters 3-5. In addition, there's considerable redundancy with the way it was presented.

Although it's probably too late for this report, future reports should be organized in more of a chronological sequence and traditional scientific presentation. Readers can always skip to the good parts. It seems that this organization was used to put the good parts, results and interpretation, early and in a synthesized form. The problem with that organization for many readers is details of the rationale and related analyses are important for understanding, believing and remembering the analyses.

5) The *Freshwater Biological Traits Database* was pulled out as a standalone report. Please comment on the merit of this report on its own vs. as one of the appendices of the main report. What next steps would you suggest for this report and the database?

This is great. Traits databases are so important for developing the kinds of indicators that I mentioned above, which have really important applications in multimetric indices, stressor diagnosis, and forecasting ecological change. I recommend that a pilot system be developed, tested with states and other users, and then completed as soon as possible. It will be important to have rules for entering data and periodic syntheses to produce a recommended and "certified" set of traits and taxa.

Another element of traits analysis that is important for use in stressor diagnosis, like temperature and climate change, is a causal analysis of the traits. Some variation of the propensity score analysis conducted by Yuan (2010) would help evaluate causal traits analysis. I have conducted analyses of diatom species traits using the same kind of approach and have found that the covariation in diatom species traits for N and P can be disentangled by developing traits for one variable while stratifying the data for another. For example, N and P traits of species were highly correlated when using all data from the National Lakes Assessment. However, we if used only low P or high P lakes to develop N traits of taxa and only low N or high N lakes to develop P traits for taxa, correlations between N and P traits of taxa were greatly reduced and actually negative in one case.

If we are concerned about covariation in temperature and land use in the development of taxa traits (e.g., weighted average optima) for temperature, for example, you could stratify the data by land use into several strata, calculated temperature optima for taxa within each stratum, and then correlate optima of taxa among strata. We would expect a relatively high correlation in taxa traits (e.g., WA optima) for temperature among strata, but little correlation in taxa's temperature traits and land use.

6) What next steps would you suggest, based on this work, as being the most informative for EPA's Office of Water and state bioassessment/biomonitoring programs to address climate change effects?

Please develop more specific guidance for how climate can be integrated into a policy-management-assessment framework. Without that guidance, information will not be gathered and used as effectively as with a clear understanding of how metrics will be applied to assess, for example, deviation from historic reference condition and current reference condition.

7) Please comment on the public comments submitted for this draft report. Specifically, which comments should or should not be addressed in the final draft? (public comments will be compiled and provided to reviewers by ERG)

The concerns expressed by the public comment are covered in the report or in my review.

Comments that should be sorted by questions, depending upon how the panel interprets questions.

General comments:

I like that four geographically and climatically distinct case studies were used in the assessment of climate change effects on stream macroinvertebrate assessments.

It occurs to me that algal metrics may be less sensitive to temperature and hydrologic effects of climate change than invertebrates.

Relating comments to specific locations in the report:

1. From the summaries, page numbers are followed by paragraph and line numbers.
2. For the chapters, chapter numbers are followed by a period and then referenced to line numbers.

xv-1,1. Make it clearer, perhaps even in the title, that this is about streams and macroinvertebrates. Otherwise, explicitly address how this will relate to other resources.

xv-1-1. This is a long executive summary. Can you bring in the bullet points from xix and shorten this to a couple pages?

xv-3 About here you need to bring in concepts about how effects of climate change and local land use need to be assessed separately. See comment on policy-management-assessment framework in the face of climate change.

xv-4-8 It is not clear to me why you separated implementation from assessment design.

xvi-4-8 Use terms that are readily understood by a broad audience, especially in the Executive Summary. For example, what is “low pulse count”?

xviii In several spots, the BCG and tiered uses are invoked as approaches “to guard against lowering water quality protective standards” (xviii-5-4), but it is not clear how. See comment on policy-management-assessment framework in the face of climate change.

xix-1-8 These bullet points are so short and filled with specialized terms that I’m not sure that policymakers, yet alone managers, will understand the point. For example:

1. xix-1-10 What are predictive models? This is used in many ways, relative to RIVPACS and modeling natural variation in multimetric expected condition. Will policy makers know what you mean?
2. xix-1-13 “Vulnerability” is used in so many contexts that its meaning is not clear here.
3. Xix-1-14 “to enable stressor diagnosis” or something that describes why.

xxii-1-4 Change “may” to “will”. Statistically we know this is as certain as anything.

xxii-3&4 Is this an argument for using finer taxonomic resolution in metrics to help distinguish climate change effects from land use effects?

xxiii-1 This assumes than new models will be developed under successive climate regimes; i.e. that current expected condition is redefined in the future under different climate regimes. This may be true (see framework on policy-management-assessment), but it is not clear why from the report.

xxiii-3-1 Recalibration seems to be defining current expected condition – *sensu* my proposed policy-management-assessment framework.

xxvi-1-6 & paragraph. The RIVPACs and multimetric modeling framework of, for example, Cao et al. (2007) should be explained briefly and bundled relative to the idea of recalibration and redefining current expected condition from historic reference condition, per my proposed policy management-assessment framework.

xxviii-1-1 Distinguish climate effects and land use effects has to be a goal. The goal should not be detecting climate change effects, it should be assessing them.

xxviii-1-2 If assessing climate change effects is the goal, then targeting sensitive sites will overestimate climate change effects. A representative set of sites, from a diversity of natural climatic and geologic conditions, should be selected to assess climate change effects. Plus, this can be built into models and can be assessed with long term trends assessment.

xxviii-3-1 The rationale for protecting reference sites should include protection as part of antidegradation, as well as provide a mechanism of assessing climate change. You should also assess climate change in trashed and moderately disturbed sites. The focus on reference sites and rationale for protecting reference sites is only part of what should be done to assess climate change effects. The arguments for using BCG and protecting reference sites is not sufficiently justified without a fuller policy-management-assessment framework. Thus the repeated inclusion of reference site monitoring and the BCG gradient could be misconstrued as promoting favorite themes without sufficient justification, which could compromise the integrity of the report.

xxix-2-2 Change this line to “use effects are great and widely distributed”. To say they have not been conceived is not correct. To say they have not been quantified may be true, but is not the point.

xxx-1-1 The importance of collecting water chemistry data and climate effects on water chemistry is not well developed in this section. The entire section is a less developed and supported than other sections. Strengthen these arguments or delete them.

xxxxi-7-1 Stream hydrology should be considered as a factor in climate change as well as temperature. We could expect shorter invertebrate generation times in streams that are more stressed by drought and flood (Riseng et al. 2006). These factors (flood/scouring and drought), in addition to temperature change, should be included more explicitly and frequently in discussion of climate change effects. Biological assessment may be particularly powerful for assessing hydrologic factors that vary at interannual scales and that are difficult to assess by measuring physical factors without continuous monitoring.

Chapter 2.

I found chapter 2 to be a very difficult chapter to read and

Numbers refer to the chapter and line.

2.165 Greater than what? Insert “than climate change” after effect.

2.265 It is important to distinguish how the trait based indicators can be used. Indicators based on species sensitivity or tolerance to specific environmental factors can be used for three basic efforts, measure a biodiversity response isolated to a specific stressor (as much as is possible), actually infer stressor condition, and help forecast changes in assemblages in the future. In addition, causal relationships between what you think traits are reflected and what are reflected should be confirmed using multiple lines of evidence (Beyers, 1998) and propensity analyses (Yuan 2010).

2.242-246 At this point in the report, the reader has no idea where these regions are and what to expect in results. Should we expect different results in the Wasatch Unita Mountains and Colorado Plateau? Maps of the regions and some characterizations of regional references should be developed.

2.246 Could the lack of relationships between temperature and invertebrate metrics be related to temporal lags? For example, I’d expected metrics based on individuals to be relatively responsive on short time scales, so the number of individuals of warm-water taxa could increase pretty quickly, although there is probably considerable effect of conditions the year before on number of eggs laid. Metrics based on the number of taxa may have even greater temporal lags. Taxa with short generations times could respond quickly to changes in temperature.

2.246 Were relationships of these metrics and climate change also evaluated with longer term regression models with time and long-term temperature averages, as for example in Figure 3-13? Was the same suite of analyses run on all long-term datasets (e.g. box plot comparisons (Mann-Whitney U's?), regression, variables)? Why not just related metrics to temperature in a regression, why use a categorical approach?

2.309 Why not relate a metric (e.g., invert size, generation time) associated with flood or drought disturbance with driest/wettest years comparisons, versus temperature tolerance?

2.387 Elevation does not have a direct effect on invertebrates. Don't you have temperature data from either air or water to show what the elevation proxy for temperature would be and whether that's sufficient to generate an effect?

2.402 I looked back and could not find where this was described for Utah.

2.409 "The three sites" produce box plots with lots of observations in Figure 2-10? Credibility of the analysis is being compromised by some of the apparent "haphazard" grouping of data and different analytical approaches. For example, why pick 150 m as an elevation benchmark? Plus, 150 m does not seem like a lot of difference in elevation.

2.482 Should you insert "WA models for" in front of "low flow parameters"? It's the results of the WA models, not the low flow parameters, that you are referring too, right?

2.491 Great, size matters (or generation time) with respect to hydrologic parameters.

2.494 It is not clear why you selected the variables that you did in these ordinations. For example, why does year matter? What else was in the analysis that did not matter?

2.496. References to figures and tables in the text should be parenthetic, not complete sentences that repeat part of the figure caption or table title.

2.543-544. If it is hard to draw conclusions from data. Don't present them without elaboration more about what their likely meaning is or is not.

2.558-561 Why were only reference sites with long term data used? Plus later descriptions of the Sheepscot raise questions of whether these reference sites have the same levels of human disturbance. Changes in minimally, moderately and severely impacted sites could differ.

2.569-570. Why were different analytical approaches uses in Maine and Utah?

2.577 I cannot find a justification for why Utah results would represent all western states or where results from western states were analyzed. This type of extrapolation of the results should be dropped here and in future references to New England and the Southeast.

2.583. I question the reliability of an ordination with less than 30 observations. This has about 17.

2.755-756. Credibility of findings is compromised when multiple tests were done and the reader does not know how many. Add how many metrics were tested. If 4 is not a significant number of all tests, then you should not present results of any because they could have happened by chance. By presenting them, you imply there is some certainty they mean something, when statistically there is little certainty if only 4 metrics were significant in a large number of tests.

2.1020 Please tell us how many non-reference sites had long-term data that could have been analyzed.

Tables 2-4 to 2-6. Are these P values corrected for multiple tests? If the answer is no, then these results are not particularly strong and there's no need to present them.

Pg. 2-58, ln 1-6. Why build the case that diversity decreases with all these caveats? The reason is surely that multiple cases show that decreases in cold-water taxa without commensurate increases in warm-water taxa seem to affecting metrics and are expected to be an issue with metrics in bioassessment programs as time goes forward. When presented like this, with the "one Utah station" and complexity caveats as well as after pages of correlations, the certainty of results presentations and interpretations is compromised. A conceptual model and a priori hypotheses sets up expectations for analyses of results upon which such seemingly anecdotal results have more credibility. For example, why would you expect decreases in cold-water taxa without commensurate increases in warm-water taxa? Why wouldn't you expect numbers of warm-water taxa to increase with temperature? Is environmental change too rapid for colonization? Are there no warm water taxa? Of course this is not what is observed in the Sheepscot (Figure 3-13), although the symbols are hard to see.

3.70-71. State-specific analyses do inform regional views, but not enough to use either New England and later Northeast in the title to overstate this level of transferability. Vermont is very different than Maine with respect to many of the findings demonstrated. In some cases, North Carolina may be more informative for Maine than Vermont. Delete regional references in titles.

3.190 Typo in $p<0.001$ or $p=0.01$.

3.692-693. It seems to me that some biological metrics are selected for their diagnostic value, but most are selected for their characterization of the "natural balance of flora and fauna ….." as per the definition of elements of biological integrity proposed by Karr and Dudley (1981).

3.717-722 This is a good example of the policy-management-assessment framework that should be developed around which the report should be organized. Perhaps that's beyond the scope of the report.

3.731 What is OCH?

3.737-739 I would like to have seen the hydrology traits explored more in analyses for the report.

4.768 This section addresses collecting land use, habitat, and abiotic information such as water quality information, which is important for stressor diagnosis. Distinguishing changes in temperature, hydrology, and water chemistry resulting from climate change and land use change is critical for causal analysis and management. Relating this all to the policy-management-assessment framework is essential.

4.768. Long-term records should be developed at non-reference sites.

4.783. I am not convinced that sampling the same sites every year is necessary (see Figure 3-13). Nor is regional watershed-specific rotation satisfactory, but a more regionally distributed probabilistic sampling could work. More analysis, probably best by statistical modeling and scenario analysis, would be important to address this sampling strategy issue.

4.829 This level of methods was really helpful to understand the results presented.

4.874 Why not stream size?

4.875-876. I'm not convinced grouping by ecoregions, whether level 3 or 4, is a sound approach.

4.893 insert "to" after "power"

4.896-899. Here it states that transferring results from just a few sites to regions is "problematic". Yes, indeed. So remove regions from titles.

4.910. We should not assume that non-reference sites will response similarly to climate change as reference sites. Later in this report, urban streams are shown to be more vulnerable than reference streams to hydrologic alteration. Nutrient loading and temperature increases with climate change could be greater for agriculturally dominated watersheds in which riparian canopy and buffers are absent than for reference sites. Thus, non-reference sites should be included in the Sentinel Network.

4.1113 or farmed

4.1142 I like this emphasis on cross-state collaboration. Since this is unlikely in most cases, maybe the federal government should run the sentinel program as part of the ongoing aquatic resources surveys. There's no reason we only samples streams, lakes or wetlands during a single year, just as we don't have to do rotating basins to save money.

4.1151 Why is this the first time that non-reference long-term monitoring is mentioned? The message to managers does not need to be so abbreviated that it becomes biased.

4.1159 Stressor diagnosis should be emphasized more.

4.1167 This is important. Lack of long-term trend data from non-reference sentinel locations WILL present limitations to separating effects of climate and land use change, which is critical for management.

5.1379 Can't we use regional air temperatures, which are readily available, to estimate what regional water temperatures should be based on known hydrologic differences among sites?

6.1413 Delete "threatens". Insert "challenges the current vision of".

6.1467 With TMDLs, stressor and causal diagnoses will be critical for distinguishing between management strategies for land use or effects of climate change.

6.1485 I don't think liable is the right word.

6.1481-1484 Not if both reference and non-reference change.

6.1516-1519 I'm not convinced refined aquatic life uses is nearly as important is tracking the generalized stressor gradient and human disturbance gradient to help with stressor and causal diagnoses that will be important for management strategies for land use versus climate effects.

7.1 This seems like an executive summary.

Literature Reference

Barbour, M. T., S. B. Norton, H. R. Preston, and K. W. Thornton, editors. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publication, Pensacola, Florida.

Beyers, D. W. 1998. Causal inference in environmental impact studies. Journal of the North American Benthological Society **17**:367-373.

Cao, Y. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological **26**:566-585.

Riseng, C. M., M. J. Wiley, and R. J. Stevenson. 2004. Hydrologic disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. Journal of the North American Benthological Society **23**:309-326.

Stevenson, R. J., R. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004a. Designing data collection for ecological assessments. Pages 55-84 *in* M. T. Barbour, S. B. Norton, H. R. Preston, and K. W. Thornton, editors. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publication, Pensacola, Florida.

Stevenson, R. J., B. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004b. Interpreting results of ecological assessments. In: M. T. Barbour, S. B. Norton, H. R. Preston, and K. W. Thornton, eds. Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Pgs 85-111. Society of Environmental Toxicology and Chemistry, Pensacola, Florida. ISBN 1-880611-56-2.

Stevenson, R.J. and S. Sabater. 2010. Understanding effects of global change on river ecosystems: science to support policy in a changing world. Hydrobiologia 657:3-18.

Yuan, L.L. 2010. Estimating the effects of excess nutrients on stream invertebrates from observational data. Ecological Applications 20:110-125.

COMMENTS SUBMITTED BY

N. Scott Urquhart, Ph.D.

GENERAL COMMENTS – POST MEETING

A. The report under review represents an important and substantial effort. The effort should be applauded. It demonstrates how difficult it is to extract meaningful findings from data gathered for another purpose. The report ignores its potential reader. More detail on that below.

B. Virtually every comment made by the other reviewers is valid and supported by this reviewer. In particular Eric Smith's comments about the distortion of significance levels by the application of thousands of statistical tests identifies an important limitation in the results.

C. The main conclusion from this study/report should be that data relevant to the impact of global change on bioassessment is practically unavailable, especially on the spatial scale of the entire US. If the relevant personnel at EPA agree with this evaluation, and the need for such data is judged to be important, then EPA needs to initiate a program to gather the relevant data on a national scale, and with consistently applied collection and evaluation protocols. By starting soon, the needed calibration data should start to become available within a decade.

D. PRESENTATION: Look at how EPA has presented its REPORT ON THE ENVIRONMENT (ROE). (http://www.epa.gov/roe/). It has a solid report as a large pdf file, but with web access to extensive supporting data, as well as briefer summaries for nontechnical audiences. This reviewer suggests the ROE material and presentation should serve as a model for how this report and its supporting information could be presented. The important point here is that the rewrite should consider the potential reader of the document and supporting information. Organize to communicate. Explain to the reader how the material, especially what was in the appendices in the draft, is organized and to which points it is related. One member of the target audience made observations about the sort of information which would be of interest to people in positions such as hers.

ALL OF THE FOLLOWING OBSERVATIONS WERE MADE BY URQUHART PRIOR TO THE MEETING. HE SEES NO REASON TO CHANGE THEM.

Charge Questions

1) Based on your knowledge of bioassessment/biomonitoring programs, biological indicators, and climate change effects, please comment on the report with respect to:

   a). Providing sufficient technical evidence to support programmatic modifications to address climate change effects; what additional steps, data, or analyses would improve the evidence?

   RESPONSE: This report shows how difficult it is to document climate change effects with data which was not gathered with that objective in mind. No additional analyses seem fruitful at this point. More relevant data is needed. The really important conclusion and only soundly defensible approach is to follow recommendation #2 (on page 7-1). (

Unfortunately, the conclusions have no line numbers for reference.) This really important recommendation should be much more detailed. In fact, it should be the main point of this report!

b). The main factors to consider in order to transfer or apply insights from the four states in this pilot study to other state programs.

RESPONSE: Frankly, none. The difference in states' objectives and the project's objectives are so great that state data will be of little assistance in meeting the study's objective. Once some solid data is available when recommendation 2 has been implemented, some changes in multi-metric indices, as per recommendation #1 (page 7-1) could be recommended.

2) Has EPA pulled out the most important findings in the Summary for Managers and Policymakers (SMP) from the technical findings? If not, what findings do you suggest should be highlighted?

RESPONSE: Both the Executive Summary and the Summary for Managers and Policymakers represent the main report well.

3) Are the analytical methods used appropriate for the questions being asked and the available data? Are the strengths and weaknesses of the methods used in the report fully described and accounted for?

RESPONSE: Analytical methods are reasonable. Some details could be argued, but overall the approach is defensible. Strengths and weakness are generally acknowledged. An attempt has been made to account for weaknesses in the available data, but it has been gathered for other objectives, so it is not as relevant as the authors would like.

4) Does the main body of the report effectively capture the more detailed information presented in the appendices and if not, what are your recommendations for improving the report or appendices?

RESPONSE: First, the appendices are a jumble of information, apparently written by different people. This is a fact: Four of the appendices are state specific, for defensible reasons, have different sections, emphases, data, and analyses. Thus it is difficult to summarize them accurately in the body of the report. The summary is generally sound, except for one perspective. The individual appendices convey much more uncertainty about conclusions than the body of the report does. Appendix H deals with temporal trends in Ohio, but there is no section 2.3.4 to discuss them, and no apparent explanation for its omission.

5) The *Freshwater Biological Traits Database* was pulled out as a standalone report. Please comment on the merit of this report on its own vs. as one of the appendices of the main report. What next steps would you suggest for this report and the database?

   RESPONSE: The *Freshwater Biological Traits Database* is necessary for manipulations needed for the document being reviewed. As it is dynamic, it does need to be online as it is. Thus the report under review should briefly describe what that database is, how it was created, what it contains, and how to access it.

6) What next steps would you suggest, based on this work, as being the most informative for EPA's Office of Water and state bioassessment/biomonitoring programs to address climate change effects?

   RESPONSE: EPA needs to implement recommendation #2 (page 7-1). As a part of doing that, it should assemble a panel to assist in the development of such a program. That panel should be realistic, not "pie in the sky", by being constrained with a realistic budget. The objectives of that project should be simple and implementable, not a charge to conduct research for 10 years!

7) Please comment on the public comments submitted for this draft report. Specifically, which comments should or should not be addressed in the final draft? (public comments will be compiled and provided to reviewers by ERG)

   RESPONSE: As of this date (5/3/2011) public comments have not been provided to this reviewer. (Subsequent to the premeeting comments, one public comment was received and commented about in the general comment #4.)

Reviewer Comments

8) **Other substantial comments**:

   a. What is the goal or objective of this study/report? The title and page 1-2, lines 39 & 40 imply the study is to "determine what components of bioassessment programs are threatened by climate change." Yet in a few pages the emphasis has shifted to using bioassessment data to detect the effects of climate change. Study/report goals should be stated unambiguously and early in the Executive Summary, in the SMP, and remain the focus throughout.

   b. The figures need a lot of work to make them consistent from one to the next. They often contain panels A, B, etc. In most cases an abbreviated definition of A, B, etc as a part of the subfigure would greatly increase readability.

c. The authors seem to assume that randomly selected sites cannot be revisited. Exactly the opposite is true.

d. Much of the appendix and some of the main text uses English units of measure. Isn't EPA supposed to use metric measures in its reports? Why this inconsistency?

e. Many benthic macroinvertebrate collections are planned to occur during times when stream water comes mainly from base flow. As base flow arises from ground water, the temperature of such water is greatly moderated by its long passage through the ground. Such water may show very little effect from changes in air temperature. This fact appears to have been mainly ignored in this report. It is recognized at least one place (somewhere in SMP).

f. Executive Summary: Why no explicit mention of recommendation #2, (page 7-1)? This is important to mention early, because it implies the need for funds.

g. Appendices: This reviewer counted a total of 50 planned appendices, plus another report with 5 appendices. This is an overwhelming amount of supporting information. Please write a substantial introduction to the appendices which provides a very good road map as to how they link and apply to the goal of this document/study. Further, many readers might be unfamiliar with a major data source used extensively in the analyses reported in the appendices: PRISM. A clear and simple explanation could be based on http://prism.oregonstate.edu/docs/przfact.html.

h. Ohio has very extensive data sets, but they appear to have been omitted from major analyses, such as the temporal trends in the main text, and in the database. This apparently major omission needs to be addressed, either with an explanation or inclusion of more information from there.

9**) Other Comments:**

a. Abbreviations: Need to also include CC, CADDIS, and IPCC

b. In standard English usage, a comparison should have the form, **this** is compared to **that**. This report frequently leaves the reader with the responsibility to deduce "**that**". This could lead to misinterpretations of what was intended. For example, page xxi, line 2, "higher proportion" than what? Do the authors mean than at a higher elevation? Page xxiii, end of the Evidence paragraph, "larger" than what? Do the authors mean than the present differences?

c. Page xvi, line 9 "highly sensitive" Really? Elsewhere the report indicates some trait groups are somewhat, but not highly sensitive to moderate temperature changes.

d. Should most references to “states”, as users of bioassessment, refer to “states and tribes”? See, page xxiv, paragraph on “Adaptation response”.

e. Page xvii, last two sentences, last full paragraph: This important point needs more visibility here and elsewhere.

f. Page xviii, lines 18 & 19: “… so that climate effects can be tracked” → so that they are not influenced by

g. Finding 7 about abiotic data (page xxx): This reviewer has had extensive, but very disappointing experiences with such data; he has even placed and retrieved data loggers himself. Much of the biotic data happens once, or at most a few times, in a year. It sounds good to get lots of data, but there is no practical analytic technique for identifying what temperatures among thousands available drive a relatively rarely evaluated macroinvertebrate response.

h. Page 1-2, lines 39 & 40 need to be moved to Page1-1, lines 3 & 4 with subsequent information defined as context.

i. Figures 2-1 through 2-4, and subsequent ones should have panels identified similar to those in Figure 2-5. Figures should be as consistent in placement of similar items as possible. For example, 2-3 & 4 display %, number, %, number, while 2-5 displays number, %, number, %, and 2-6 displays number, number, %, %. Starting with Figure 2-7, sample sizes are mentioned, but not earlier. Give your reader a break! Reduce confusion to a minimum so readers can concentrate on content!

j. Figure 2-18. station 4951200 (Virgin) – Include the station name, as elsewhere.

k. Figure 2-23: What are the dotted curves? Confidence interval in the location of the line at a point, or a prediction interval on the location of a future point? Or something else?

l. Page 2-41, line 803: Why no mention of Ohio here? It is covered extensively in the appendices.

m. Page 3-22, line 470: … ICI and IBI improvements, OR there may be no observable effects due to climate change.

n. Page 4-4, lines 843-844: Criteria such as these need to be screened for plausibility. If conditions are too stringent, all sites might have to be in national parks, forests, or limited access areas as in the west!

o. Page 4-8, line 970: When did Maryland sneak in here? Figures relate to only North Carolina.

p. Page 4-18, line 1186-1187, &1195-1196: The first sentence is UNTRUE! The second sentence describes how to completely avoid the imagined problem.

q. Page 5-1, lines 1225-1228: The states have their responsibilities. The Climate Change Program needs to be realistic that the program's objectives differ substantially from the states'.

r. Page 5-8, line 1358: Compositing samples is much like summing numbers; the only problem results when subsequently samples need to be split. This reviewer has looked at this problem closely. Counting is one of the major expenses in the process of evaluating benthic macroinvertebrates. Compositing with possible splitting reduces the number of samples which must be "picked."

s. Page 6-3, lines 1474-1477. This statement is too strong or assertive, especially relative to responses being predictable.

t. Figure A-6: Differencing of variables ordinarily induces substantial correlation among the differences. Has this been accounted for in the analyses?

10**) Details:**

a. The study (page xv, line 4) What study? This study? (line 9) Does "the report" mean this study? If the intent is to refer to the present document or study, rather than some other one, identify it as **this**.

b. Figure SMP-2: "… between climate change trends" → climate change (CC above) trends

c. Page xv, line 10: "," omitted after taxa.

d. Figure 2-28: Titles interior to the figure have been squeezed to such an extent that they are unreadable.

f. Figure 3-1: Where is the PRISM mean annual air temperature?

g. Page 3-8, line 190: p=001; is this meant to be p=0.01?

h. Table 3-4: Formatting – Last column should be labeled something like Difference in Final Scores (with no horizontal line). The – signs need to be adjacent to the numbers. The latter problem repeats itself in Table 3-7.

i. Are the bar colors the same in figures 3-4 and 3-6? They should be.

j. Page 3-20, line 422: BI → NCBI?

k. Page 3-21, lines 457-463: … two possibilities exist. The first one clearly follows immediately. Where is the second? Is it the Other scenarios? If so, that implies more than two possibilities. This writing needs to be clarified.

l. Page 3-21, line 458: do → due

m. Page 3-33, line 716: Figure 3-13 should be 3-11.

n. page 4-11, lines 1021-1024: English and metric units mixed up.

o. Page 4-12 & Table 4-2: Is “BC” the Base Case?

p. page 4-12, line 1057: Figure 4-3 or perhaps 4-6?

q. Table 4-2: (…(…)) – latter) omitted

r. Page 8-18, Stoddard citation: Why capitalize Johnson?

s. Figure A-9: Why three lines?   What has happened to minimum and average temperature?

t. Page B2-7, line 246: Data for North Carolina were compiled from ____ into a database. What has been left out?

u. Figure B3-15b: (… (…). Should be (…(…)).

v. Figure D-1: Vertical label is reversed and illegible; (u) → (μ); if the mean is denoted by a Greek letter, the standard deviation should be also: σ.

w. Page H-15, line 303: Where is the figure referred to here as H3?

Appendix B: List of Reviewers

Peer Review Workshop for EPA's Draft Report: Implications of Climate Change for Bioassessment Programs and Approaches to Account for Effects

The Navy League Building
Arlington, VA
Wednesday, May 11, 2011

Reviewers

Daren M. Carlisle, Ph.D.
Chief, Ecology National Synthesis
U.S. Geological Survey
National Water-Quality Assessment Program
12201 Sunrise Valley Drive
Reston, VA 20192
703-648-6890
Email: dcarlisle@usgs.gov

M. Siobhan Fennessy, Ph.D.
Professor, Department of Biology
Kenyon College
Gambir, OH 43022
740-427-5455
Email: fennessym@kenyon.edu

Eric P. Smith, Ph.D.
Professor and Department Head
Department of Statistics
Virginia Polytechnic Institute and State University
406A Hutcheson Hall
Blacksburg, VA 24061
540-231-5657
Email: epsmith@vt.edu

R. Jan Stevenson, Ph.D. (Chair)
Professor, Department of Zoology
Michigan State Unversity
East Lansing, MI 48824-1115
517-432-8083
Email: rjstev@msu.edu

N. Scott Urquhart, Ph.D.
Statisical Consultant
Professor of Statistics (retired)
Colorado State University
Windsor, CO 80550-4627
Email: nsu@peakpeak.com

Appendix C: List of Observers

United States
Environmental Protection Agency
National Center for Environmental Assessment

Peer Review Workshop for EPA's Draft Report: Implications of Climate Change for Bioassessment Programs and Approaches to Account for Effects

The Navy League Building
Arlington, VA
Wednesday, May 11, 2011

Final Observers

LeAnne E. Astin
Ecologist II
Fairfax County Stormwater Planning
12000 Gov. Ctr. Pkwy.
Fairfax, VA 22035
Phone: (703) 324-5879
Fax: (703) 802-5955
Email:leanne.astin@fairfaxcounty.gov

Britta Bierwagen
Physical Scientist
U.S. Environmental Protection Agency
1200 Pennsylvania Ave., NW
MC 8601P
Washington, DC 20460
Phone: (703) 347-8613
Fax: (703) 347-8694
Email: bierwagen.britta@epa.gov

Jan Connery
Meeting Facilitator
Eastern Research Group, Inc.
110 Hartwell Avenue
Lexington, MA 02421
Phone: (781) 674-7322
Fax: (781) 674-2851
Email: Jan.Connery@erg.com

Anna Hamilton
Tetra Tech, Inc.
502 W. Cordova Rd., Suite C
Santa Fe, NM 87505
Phone: (505) 982-0583
Fax: (505) 982-0584
Email: Anna.Hamilton@tetratech.com

Richard Mitchell
Biologist
U.S. Environmental Protection Agency
1200 Pennsylvania Ave, MC 4503-T
Washington, DC 20460-0001
Phone: (202) 566-0644
Fax: (202) 566-1331
Email: mitchell.richard@epa.gov

Rachael Novak
Physical Scientist
U.S. Environmental Protection Agency,
Office of Water
1200 Pennsylvania Ave. NW, MC 4304T
Washington, DC 20460
Phone: (202) 566-2385
Fax: (202) 566-1437
Email: novak.rachael@epa.gov

Ellen Tarquinio
National Rivers and Streams Assessment Lead
U.S. Environmental Protection Agency
Office of Water
1200 Pennsylvania Ave, MC 4503T
Washington, DC 20460
Phone: (202) 566-2267
Fax: (202) 566-1437
Email: tarquinio.ellen@epa.gov

Diedre Watkins
Program Manager
Lewis-Williams Conference & Logistics
Management, LLC (ERG Subcontractor)
1299 Lamberton Drive, Suite 205
Silver Spring, MD 20902
Phone: (301) 593-2800
Fax: (301) 593-5800
Email: dwatkins@lclmllc.com

Appendix D: Agenda

United States
Environmental Protection Agency
National Center for Environmental Assessment

Peer Review Workshop for EPA's Draft Report: Implications of Climate Change for Bioassessment Programs and Approaches to Account for Effects

The Navy League Building
Arlington, VA
May 11, 2011

Agenda

8:00 a.m.	Registration/Check in
8:30 a.m.	**Welcome, Introductions, Meeting Purpose & Agenda** .... *Jan Connery, ERG (contractor)*
8:40 a.m.	**EPA Welcome Remarks**..*Britta Bierwagen, EPA NCEA*
8:50 a.m.	**Public Comment** .... *Jan Connery*
9:00 a.m.	**General Questions** ....*R. Jan Stevenson, Ph.D.* (*Chair) & Panel*
	1) Based on your knowledge of bioassessment/biomonitoring programs, biological indicators, and climate change effects, please comment on the report with respect to: a) Providing sufficient technical evidence to support programmatic modifications to address climate change effects; what additional steps, data, or analyses would improve the evidence? b) The main factors to consider in order to transfer or apply insights from the four states in this pilot study to other state programs.
10:10 p.m.	BREAK
10:30 a.m.	2) Has EPA pulled out the most important findings in the Summary for Managers and Policymakers (SMP) from the technical findings? If not, what findings do you suggest should be highlighted?
11:00 a.m.	3) Are the analytical methods used appropriate for the questions being asked and the available data? Are the strengths and weaknesses of the methods used in the report fully described and accounted for?
11:45 a.m.	LUNCH

1:00 p.m. 4) Does the main body of the report effectively capture the more detailed information presented in the appendices and if not, what are your recommendations for improving the report or appendices?

1:45 p.m. 5) The *Freshwater Biological Traits Database* was pulled out as a standalone report. Please comment on the merit of this report on its own vs. as one of the appendices of the main report. What next steps would you suggest for this report and the database?

2:45 p.m. BREAK

3:00 p.m. 6) What next steps would you suggest, based on this work, as being the most informative for EPA's Office of Water and state bioassessment/biomonitoring programs to address climate change effects?

3:50 p.m. 7) Please comment on the public comments submitted for this draft report. Specifically, which comments should or should not be addressed in the final draft?

4:00 p.m. **Reviewer Final Comments**..................................... *R. Jan Stevenson, Ph.D. (Chair) & Panel*

4:20 p.m. **Closing Remarks**.. *Jan Connery*

4:30 p.m. ADJOURN