APPENDIX C

METHODS AND RESULTS DETAILS FOR CASE STUDY 1: THE POWER OF BIOLOGICAL ASSESSMENTS TO DETECT CLIMATE CHANGE

C.1. DETAILED METHODS FOR CASE STUDY 1

C.1.1. Analysis Approach

Power, as it relates to a monitoring program, is the ability to detect a real effect that has occurred. In statistical jargon, power is defined as the probability of rejecting a false null hypothesis. The more power a test has, the more likely one is to correctly infer that a real change has actually occurred. Power is related to type II error (β), which is the probability of accepting a null hypothesis (no change) when it is false (things changed). Thus, power = 1-β.

This study focuses on climate change effects associated with temperature, principally because 1) the goal is to demonstrate a process for calculating the capacity of a program to detect change and not to predict all the effects of climate change and 2) while few data exist on climate change effects on aquatic assemblages in general, there are more data on temperature effects than hydrologic effects. As discussed in Section 4.1, this case study attempts to answer two questions:

- How long must monitoring be conducted to have a fixed probability of detecting a change in the mean native taxa richness of the reference site population?
- How long must monitoring be conducted to have a fixed probability of detecting a change in mean native taxa richness for a particular site?

Predicted macroinvertebrate taxa loss rates due to temperature increases were derived from the literature based on observed changes in taxa richness associated with temperature increases.

For this study, native or expected taxa richness is considered rather than total richness; species replacement is not being considered. Total richness may not change if species replacement rates are high. For example, it is possible that stenothermal species (those with a narrow range of temperature tolerance) that are lost will be replaced over time by eurythermal taxa (those with wide temperature tolerances). However, native taxa are expected to be lost from many streams (e.g., Moore et al., 1997; Xenopolous et al., 2005; Parmesan, 2006), and native taxa richness based on current climate will decrease. It is this loss rate that is considered here.

Many other ecological responses are expected but are not being considered in this study. For example, responses such as density and range shifts, changes in timing of important life history stages and phenology, morphological, physiological, and behavioral changes, and changes in gene frequencies (Schindler, 1997; Hogg et al., 1998; Walther et al., 2002; Root et al., 2003; Parmesan, 2006) are not being considered. Taxa richness, a very common component metric evaluated in bioassessment programs and incorporated in multimetric indices, is evaluated for signs of bioassessment program vulnerability.

C.1.2. Ability to Detect Change–Power Analysis

Power analyses requires several critical components: the sample size (N), the variability in an observed factor (taxa richness in this case) (σ), the effect size (how much of a change one wants to detect, δ), and the significance level (type I error, α). In this case study, the power analysis approach was used to evaluate how much of a change in taxa richness (the effect size or minimal detectable difference) can be detected in a typical biomonitoring program. This detectable difference will then be compared to expected taxa loss from climate change to see how long (in years) the program must monitor to have assurance (at a set probability level) of being able to detect a taxa loss signal resulting only from climate change, all else being equal. For this application of power analysis, we must fix the desired level of power (1 - β) as well. The case study demonstrates how changing some of these components can increase or decrease the ability to detect a climate change effect.

The equations for calculating effect size for comparing two paired population means can be found in many statistical textbooks. The basic formula, assuming normally distributed populations, is:

$$\delta = \frac{(Z_{1-\alpha/2} + Z_{1-\beta}) \times \sigma}{\sqrt{n}}$$ (after Snedecor and Cochran, 1980)

where Z_{α} and Z_{β} are the z-scores (probability levels) for the desired type I (α) and type II (β) error rates.

The power equation above requires knowledge of two critical variables, population variance (σ^2) and effect size. For this case study, variance is estimated using existing monitoring data for sites sampled repeatedly over several years during an index period. Such data give an estimate of the natural variability in biological condition through time, assuming minimal external changes. Knowledge about what taxa changes (effect size) might be expected in response to climate change is also needed. This value can be derived from existing literature on taxa loss in relation to temperature changes.

C.1.3. Data Sets Analyzed

C.1.3.1. MBSS Biological Data

For this case study, the Maryland Biological Stream Survey (MBSS) dataset was used to estimate population variance (σ^2). The MBSS biological monitoring program approach and sampling methods are described in Appendix B.

A series of 28 MBSS sentinel sites have been monitored annually for the last 6 years, including sampling for benthic macroinvertebrates and fish using the same protocols described in Appendix B. This 6-year period fortuitously includes both a dry and a wet climate cycle. These repeat visit data are used to calculate an estimate of σ^2, variance in taxa richness. While this case study compares expected effects across various regions of the US, the MBSS derived variance will be used for all regions as the estimate of variance associated with biomonitoring.

Relative variability of taxa richness is assumed to be constant over time. This is likely not true as both the mean and variation in biological condition of sites may change with warming water temperatures. For simplicity, it is assumed that the variance in taxa richness associated with the MBSS program can be extrapolated in time and across different regions.

C.1.3.2. Information on Taxa Loss Rates

Macroinvertebrate taxa (i.e., genus or species, reflecting practical taxonomic limitations) loss rates were obtained from two sources. A study in France examined taxa loss associated with climate changes in the upper Rhone River over a 20 year period (Daufresne et al., 2003). The other source was the Clean Water Act 316(a) program, which includes studies on thermal discharges (e.g., Lehigh University, 1960; Morgan and Coutant, 1972; Coutant and Pfuderer, 1973; Talmage and Coutant, 1980; Energy Citations Database (http://www.osti.gov/energycitations/index.jsp)).

Daufresne et al. (2003) observed a loss of 7 macroinvertebrate taxa in streams associated with a 1.5 ºC increase over the period 1980–1999; this will be referred to as the high taxa loss rate. This equals a loss rate of roughly 4.6 taxa per °C. In reality, taxa loss rates are likely to occur episodically over time, for instance when particular species thresholds are reached. It is understood that by looking at total change in number of species over a sufficiently long time period and relating this to total observed temperature change, we are essentially estimating average species loss per unit of temperature change, and making an assumption of a linear loss rate is only to allow projection of further losses in the future.

The second estimate was derived from the literature associated with thermal discharge studies associated with the Clean Water Act 316 program. Most of the 316(a) studies focused on fish effects and many were physiological. One study (Lehigh University, 1960) included macroinvertebrate effects and this study found a loss rate of approximately 1 taxon per °C over the range 22–28 degree °C; this will be referred to as the low taxa loss rate. This represents a fairly high thermal range, but these studies were designed to investigate effects of thermal effluent, not effects of climate change. Nevertheless, the results are considered applicable. There are likely more 316(a) studies with invertebrate data, but these individual studies, for the most part, are not published in standard scientific citation databases and can be hard to locate (but see the Energy Citations Database (http://www.osti.gov/energycitations/index.jsp)).

Predicted fish taxa loss rates were also considered from the thermal discharge literature. A study of thermal effluent on the Wabash River found a loss rate of 3.6 fish taxa per °C increase in temperature (Gammon, 1973). This may be on the high side for loss rates, but it was one of the few data-based values found within a reasonable temperature range.

C.1.3.3. Prediction of Expected Taxa Losses with Projected Temperature Increases

Estimates of taxa loss rates were coupled with projected temperature increases to model the expected rate of taxa loss per year due to climate change. Projected temperature increases due to climate change for each region of the United States were taken from the National Assessment Synthesis Team (NAST) summary report (2001). NAST (2001) relied mainly on results from two coupled atmosphere/ocean general circulation models (AO-GCMs) which were used to estimate projected temperature increases for various regions of the United States (Table C-1). Although the biological data are from the Mid-Atlantic region, we also investigated how projected climate changes in other regions affected taxa loss rates.

We estimated a rate of temperature increase per year using the climate projection data and compared low and high values. We then linked the temperature rates to taxa loss rates derived from the literature to estimate taxa loss per year; again, low and high rates were used for comparison.

Table C-1 - Average annual temperature increases expected by region of the US (NAST, 2001).

	Average Annual Temperature (° C) Increases by 2100	
Region	**Min**	**Max**
Northeast/Mid-Atlantic	2.6 (Hadley)	5 (Canadian)
Southeast	2.3 (Hadley)	5.5 (Canadian)
Midwest	3 (Hadley)	6 (Canadian)
Great Plains	3 (Hadley)	6.5 (Canadian)
West	4 (Hadley)	5.5 (Canadian)
Pacific Northwest	2.7 (by 2050) (Hadley)	3.2 (by 2050) (Canadian)

Linear projections of climate change effects are used in this case study as a basis for estimating ability to detect climate-induced changes after various monitoring periods. It is likely that climate will actually change in a non-linear fashion with periods of fast change followed by periods of slower changes (Alley et al., 2007). There is little way to predict this course, however, so the linear assumption is the more conservative approach and is a common assumption used in the literature (e.g., Najjar et al., 2000). Thermal effects on taxa are also assumed to result in losses at a linear rate. Given the complexity of biotic interactions as well as temperature effects, this is not thought to be an accurate representation of reality. For example, loss of keystone taxa may precipitate abrupt and dramatic changes on stream communities as well as on stream processes (Power et al., 1985; Power, 1990; Pringle et al., 1993; Flecker, 1996). However, the simplifying assumption allows us to model changes into the future.

C.1.4. Specific Analyses

For each of the two questions being assessed (see Section C.1.1), different confidence levels were investigated (*i.e.*, α and β were varied). Effects of sample size were also investigated for both questions. To address the first question of detecting change in a reference population, the number of sampling sites, N, was either fixed per year (at N=40 for the reference population) or cumulative (N increasing by 40 each year for the reference population) and assumed a comparison of year one versus a cumulative pooled sample size through time. For the second question of detecting a change at a particular site, N was either fixed at one of three different levels (N=5, 10, and 20) or N was cumulative, increasing by 5, 10, or 20 for the respective analysis runs.

The MBSS-derived variance estimates were used for each region. Z scores were varied by changing the type I and type II error rates to illustrate the effects of these choices on the ability to detect a climate change effect. The outputs from these analyses are time series of taxa loss rates predicted from climate change effects. These outputs are compared with minimal detectable effect sizes to illustrate the length of time required to detect a climate change effect on taxa richness under various conditions (taxa loss rates, temperature scenarios, and error rates).

C.2. RESULTS

C.2.1. Question 1 - How Long Must We Monitor to Have a Fixed Probability of Detecting a Change in the Mean Native Taxa Richness of the Reference Site Population?

If a population of reference sites (N=40) is sampled each year, an average macroinvertebrate taxa richness in reference streams can be calculated. For comparing any two samples of N=40 sites, there is a fixed difference in mean taxa richness (effect size) at which significance can be detected with a specified power. For $\alpha=\beta=0.05$ (95% confidence, 95% power), and the Maryland data, the effect size is 4.5 taxa. Thus, to have a 95% probability of detecting a significant ($p < 0.05$) taxa loss between 2 samples of 40 sites, requires a mean difference of 4.5 taxa. At high taxa loss rates and under the higher estimate for warming in the Northeast/Mid-Atlantic region, it will take 15 years to achieve a mean loss of 4.5 taxa (Figure C.1), assuming that: 1) the same 40 sites are sampled each year; 2) samples from a site are not treated as cumulative through time; and 3) the analysis uses type I and type II error rates of 0.05. This value is derived by identifying the point where the effect size line (hatched) crosses the taxa loss rate line (solid) (Figure C.1).

Figure C.1 illustrates a variety of scenarios representing different confidence levels and either fixed or cumulative sample sizes. For example, relaxing the confidence level decreases the time to detect a change. Increasing α and β from 0.05 to 0.20 (reducing both confidence and power), reduces the time to achieve 80% probability of detecting a significant climate change effect ($p<0.2$) to approximately 8 years. If a 1 in 5 (rather than a 1 in 20) chance that statistically significant results are due to random chance alone is acceptable, a taxa change attributable to climate change could be detected in half the time. This is the type of trade-off that is important for programs to consider.

Table C.2 - The time (years) to achieve a fixed probability of detecting a statistically significant effect of temperature increases on macroinvertebrate and fish taxa loss across different regions under maximum and minimum temperature projections. These data are for question 1 and assume a fixed sample size of N=40 reference sites sampled each year. Data are shown for different taxa loss rates and for different confidence levels.

		Northeast/Mid-Atlantic	Southeast	Midwest	Great Plains	West	Pacific Northwest
Maximum Predicted Temperature Increase by 2100							
Macroinvertebrates - High Taxa Loss Rate (4.6 per degree C)							
	$\alpha=\beta=0.95$	**15**	**14**	**13**	**---**	**13**	**---**
	$\alpha=\beta=0.8$	**8**	**7**	**7**	**---**	**7**	**---**
Macroinvertebrates - Low Taxa Loss Rate (1 per degree C)							
	$\alpha=\beta=0.95$	**70**	**64**	**58**	**---**	**57**	**---**
	$\alpha=\beta=0.8$	**36**	**33**	**30**	**---**	**29**	**---**
Fish Taxa Loss Rate (3.6 per degree C)							
	$\alpha=\beta=0.95$	**20**	**18**	**17**	**---**	**16**	**---**
	$\alpha=\beta=0.8$	**10**	**9**	**9**	**---**	**9**	**---**
Minimum Predicted Temperature Increase by 2100							
Macroinvertebrates - High Taxa Loss Rate (4.6 per degree C)							
	$\alpha=\beta=0.95$	**29**	**33**	**38**	**19**	**17**	**14**
	$\alpha=\beta=0.8$	**15**	**17**	**19**	**10**	**9**	**7**
Macroinvertebrates - Low Taxa Loss Rate (1 per degree C)							
	$\alpha=\beta=0.95$	**>100**	**>100**	**>100**	**88**	**79**	**64**
	$\alpha=\beta=0.8$	**69**	**78**	**89**	**45**	**41**	**33**
Fish Taxa Loss Rate (3.6 per degree C)							
	$\alpha=\beta=0.95$	**38**	**42**	**49**	**25**	**22**	**18**
	$\alpha=\beta=0.8$	**19**	**22**	**25**	**13**	**12**	**9**

Decreasing the projected effect size (δ), either by using a lower rate of projected temperature increase or a lower rate of taxa loss per degree, will increase the time until a climate change effect can be detected, all else being equal. This is somewhat out of the control of programs in this context, since one cannot manipulate either of these factors. However, it is worth mentioning because some regions may have more specific or defensible data about these rates that would affect the projections for that region.

Lastly, the population variance (σ^2) was fixed in this study based on the MBSS data. Different states have different species composition in their rivers and streams with different inherent variability and/or may use other sampling protocols with inherently different variability. If replicated site data are available, these could be used to calculate a state specific estimate of variance to be applied in these analyses. Regardless of the estimate used, the greater the variance, the longer it will take to detect an effect.

The process described here can be applied to any assessment program using either assumptions listed here or better values derived from more regionally precise estimates. More specific state or regional estimates of population variance, climate projections, and/or predicted taxa responses to temperature increases will improve the outcome. Note that a similar process could also be applied for hydrologic effects, especially drought frequencies, if the effects of hydrologic alteration on taxa richness or any other assemblage attribute can be derived from the literature or other data sources.

There are several implications of this research for monitoring program design. The choice of probabilistic designs or targeted designs depends on the questions being asked. Probabilistic designs are good for asking a variety of questions. For example, the average condition of streams is best assessed by randomly sampling stream populations across a region. Estimating the taxa richness among reference sites is best assessed using random samples across the reference stream population. Our estimates of sample size and power are based on paired tests, that is, comparing the same set of reference sites over time, but selected randomly the first time. Paired tests are much more powerful than drawing new independent samples every year because site-to-site differences are removed from the variance term, leaving only differences over time within sites. A new random site selection each year would result in lower power (i.e., longer times to detection at specified power) than estimated here. The probabilistic resampling scenario would require use of the overall population variance of the reference sites, instead of the interannual measurement error as we used here. With regard to climate change effects, probabilistic sampling across reference sites would be ideal using as large a sample size as possible; this sampling would have benefits for biocriteria programs as well independent of climate detection. For example, probabilistic designs are best for identifying trends in average resource condition and for generating a variety of inferences about any number of resource elements (e.g., taxa distributions and population sizes) across large spatial scales.

Targeted site selection, however, is often needed to answer specific questions. These include whether a specific site (watershed or reach of stream) is meeting its designated use or permit requirements. In this case, the sampling frame is the site, but random samples are still ideally taken from within that sampling frame. For the second question considered in this case study, it was assumed that the replicated samples were randomly selected from the site. Another question that benefits from targeted site selection is what the effect of a specific land use is on stream condition. It is often best to identify specific sites along a gradient to test a hypothesis related to the effects of a particular land use, for example, on stream condition. This may be important for studying how land use will interact with climate change to affect stream condition. Note that probabilistic designs can also be used to answer these questions, but the gradient may not be sampled completely with random selections.

If a biocriteria program is contemplating designing monitoring to detect climate change effects, there are several points to consider. First, protecting reference streams emerges as an important concept, especially considering that reference sites are the sentinels that will be used to gauge climate change effects as well as the relative effects of climate change on other stressors (see Case Study 2, Chapter 3). Ongoing monitoring of reference site populations is also an important aspect of program design, though this may be constrained by availability of resources (money and manpower). The larger the sample size and the more frequent the sampling, the stronger the ability to detect changes will be. Sampling these reference sites probabilistically will also provide the least biased estimate of average condition. The use of rotating designs may optimize resources, since crews can stay within defined areas, travel can be limited, and total numbers of samples collected and processed each year is reduced by focusing on a subset of basins. This approach also means that reference sites within any one basin will only be sampled once every several years, increasing the time it will take to obtain replicate samples needed to define climate change-associated trends.

Incorporating more specific regional expectations would improve the analysis. Applying power analysis using specific data on population variance and especially on regionally-specific climate projections will also improve evaluation of program vulnerability to climate change effects. In addition, it is important to review other assumptions (discussed above) inherent in this analysis approach, including taxa loss expectations, possible expectations for other community responses that could be tested, and assumptions of constant variance over time, before analyses are undertaken. If more defensible assumptions can be identified, they should be used.