5. Benthic Macroinvertebrates as Indicators of Watershed Condition

A. Overview

Sampling and identifying benthic macroinvertebrates (BMIs) and determining occurrence and relative abundance of different types of BMIs can be a useful component of assessment. There are several primary approaches for doing this that have been developed around the country and California, with the main two being the US Environmental Protection Agency's method (Barbour et al., 1999) and the related methods developed by the California Department of Fish and Game, Aquatic Bioassessment Laboratory (ABL, Harrington and Born, 2003; http://www.dfg.ca.gov/cabw/csbp_2003.pdf & http://www.dfg.ca.gov/cabw/Field/csbpwforms.html).

The "California Stream Bioassessment Procedure" (CSBP) described in Harrington and Born (2003) and in the draft update CSBP (Aquatic Bioassessment Laboratory, 2006) includes measuring metrics for BMIs to determine impacts from both non-point and point sources of pollution. The term metric refers to an attribute of a system or thing that you can measure and use to make inferences about condition of that system or thing. An example of a BMI metric is "taxa richness" – the number of different kinds of BMIs in a sampled stream.

The procedure (CSBP) consists of the following steps: 1) identifying the type of impacts to be studied through a BMI community assessment; 2) selecting a number of sites in a waterway appropriate for understanding the degree of impacts; 3) describing physical habitat and chemical water quality characteristics; 4) sampling BMIs from benthic sediments at each site; 5) identifying the BMIs to either an "Order" (Level 1), "Family" (Level 2), or "Genus/species" (Level 3) taxonomic level; 6) recording the various metrics of BMI community structure; 7) analyzing basic statistical properties of the metrics (e.g., mean and variation); and 8) comparing metrics among reference and impacted waterways or reaches and/or among waterways and expected values.

The CSBP is considered by the authors and others who have used it to be an appropriate tool for advanced citizen monitoring programs in order to identify waterways that may be experiencing impacts from water or land use. Site selection and sampling are the least time-consuming of the methodological steps and identification of the BMIs is the most time-

Chapter Guide

consuming. Whether a local group, organization, or agency decides to use this approach, or to hire someone to do it, a similar knowledge of the protocol is needed in order to use the data and knowledge developed appropriately.

B. Questions that can be answered using method

An important concept for understanding the use of this approach is “ecosystem indicators”, which are things in an ecosystem that can be measured that indicate the condition of the ecosystem and sometimes the cause of change in condition (see Chapter 1, CWAM Vol. 2). BMIs are considered to be ecosystem indicators because, like any living thing, they respond to changes in their environment and, as a group, respond in different ways from each other. Certain BMI species are tolerant to a wide range of conditions and may be poor indicators of specific conditions, but their presence indicates that a degraded condition could be present. Other species are intolerant of a wide range of conditions and their presence indicates that a certain condition could be present (e.g., degraded benthic water and substrate quality). The relative proportions of certain groups of related species, or certain functional feeding groups (e.g., shredders), can indicate the kinds of conditions prevailing in a waterway.

BMIs are considered to be “integrators” of waterways and watershed conditions in that they respond as individuals and populations to the sum of habitat conditions in which they live. Over short time frames (days), individuals of a particular group may die in response to a significant pollution event (e.g., pollution runoff from the first major storm of the season). Over moderate time frames (months, years), individuals and therefore populations of certain species may not thrive under sustained poor conditions (e.g., excessive fine sediment or high temperatures) and their numbers may dwindle and other species may do better. Over long time frames (decades), whole species may be lost or gained in waterways undergoing sustained impacts (e.g., from dams or logging) and change the nature of the aquatic ecosystem until a sustained recovery becomes possible.

The reason for conducting an investigation of BMIs will influence the approach taken for site selection and the metrics analyzed and looked at most thoroughly. The two main types of investigations pursued with the CSBP are for point sources and non-point sources of pollution (Chapter 8, Harrington and Born, 2003). Sites selected for point sources investigations will include at least a site (riffle) above and a site below the point of disturbance. Other sites may give more information about the extent of the impact or the reference (undisturbed) condition in the waterway. Sites selected for non-point source investigations are chosen to represent groups of streams or reaches within a watershed. They may be randomly chosen from throughout the watershed, or chosen randomly from within groups of similar streams. The process of site selection for non-point source pollution is best done under the guidance of someone with geographic and statistical skills.

The protocols described in the CBSP result in many metrics (9 for Level 1 and 19 for Level 2) which can then be used subsequently to describe condition of the BMI community. You will probably collect all of the data described in the CBSP, thus the selection of the Level of taxonomic detail needed will depend on the question being asked about watershed condition. In general, more information is usually better, so a Level 2 or 3 analysis will be preferable. A Level 1 analysis may be useful for describing extreme differences among waterways or extreme changes due to a land or water use. Because much of the cost associated with BMI investigations can come with expert identification to the Family level, a choice between Level 1 and Level 2 or 3 could be a fiscally-based one. However, there is no point in conducting a Level

1 investigation to study watershed condition and impacts if the question requires a more sensitive analysis (Level 3), unless just going through the exercise is useful from a community or group education perspective.

The metrics obtained from analyzing the inventories of identified invertebrates can provide several types of information useful to a watershed assessment. 1) Where in the watershed are there potential impacts of land and/or water use on aquatic biota? 2) How much of an impact is occurring? And 3) When is the impact occurring in terms of longevity if measurements are repeated over several years or longer. The table below shows the relationships between the metrics listed and watershed processes/activities they are relevant to understanding.

C. Description of method

At the time of first writing, only the 2003 version of the California Stream Bioassessment Protocol was available and in use by many watershed groups and taxonomists. Since then, a new CSBP has been released, with a draft publication date of 2006. Both versions are used in the descriptions below to facilitate an understanding of data collected under the 2003 protocol and planning for collection under the 2006 version.

1) Identify the type of impacts to be studied through a BMI community assessment

Determining what impacts or conditions in the watershed you want to assess with a BMI survey should originate from the questions identified at the beginning of or during the assessment and an understanding of the limitations and opportunities from studying BMIs. Examples of types of impacts and the CSBP (Chapter 8) survey approach to take (in parentheses) are:

- A fixed source of pollution that periodically or regularly delivers water-borne pollutants to a waterway (point source sampling).
- Diffuse sources of pollution affecting a given waterway (non-point source sampling).
- Range of conditions from very disturbed to non-disturbed across watershed (ambient sampling)

Step 1 Write out the impacts or conditions you want to study using BMIs and the reasons you think BMIs are an appropriate indicator for the investigation.

2) Select a number of sites in a waterway appropriate for understanding the degree of impacts

Chapter 8 of the CSBP (Harrington and Born, 2003) describes methods for deciding which sites will be appropriate for your impact or condition questions. These methods are a “decision framework” because the authors could not anticipate all conditions the assessor could face. There are many ways that sites can be selected, from using solely professional opinion to using computer/GIS models to aid decision-making. For non-point source or ambient sampling in medium-sized (>10,000 to 100,000 acres) complex (multiple and fragmented ownerships) watersheds, it is important to use existing information on the spatial characteristics of land-use and land-cover to aid site selection. For smaller watersheds, there may be sufficient professional knowledge to select sites for sampling.

Point-source sampling design requires at least the selection of sites above (control/reference) and below (impacted) the potential disturbance. Other sites within the waterway or in nearby similar waterways may be selected to increase knowledge about reference or non-disturbed conditions. Other sites below the potential disturbance may also be selected to understand the linear extent of the disturbance down the waterway. For the 2003 CSBP, each site will include 3 transects (lines across the waterway) and 3 sampling locations per transect, for a total of 9 sampling locations. At each transect, you will combine the 3 sets of BMIs collected at each of the three locations into one composite sample. For Level 1 and 2 investigations, you will need at least 100 BMIs per sample, so for each riffle/site, you will need 300 BMIs. For Level 3 investigations, you will need at least 300 BMIs per sample, so for each riffle/site, you will need 900 BMIs. What this means is that the number of sites chosen will directly affect the amount of time and potentially money spent on BMI identification.

Non-point source sampling design (2003 CSBP) requires the selection of 3 transects within a reach containing ≥ 5 riffles. The reaches will be chosen using one of the following strategies: a) Stratified random selection – which means that all sub-watersheds and reaches will be combined into groups of like and sampling reaches and sites selected randomly from within each group (CMAP uses this approach). First group all reaches in the study area (watershed) into categories based on one or more of physical (geology, soils, slopes, precipitation), biological (fish populations, vegetation), and land-use (ownership, past/current land-use, zoned/permitted use) patterns. Once you have grouped the reaches, you would select a sub-set within each group to represent the entire group. How the groups are described will determine the number of groups. b) Non-random selection based on physical/habitat characteristics – which means that stream order and channel type is used to group reaches and one riffle is chosen from each reach. c) Non-random selection based on access – which means that all accessible reaches are identified for each sub-watershed and sampling sets of 3 riffles chosen for upper, middle, and lower reaches within each sub-watershed.

For the 2006 CSBP, each ambient monitoring (non-point source) site is a 150 m long reach (or 250 m if your creek is greater than 10 m wide) and is divided into 11 equidistant transects. Ten more transects are interspersed within these 11, to give a total of 21 transects per reach/site.

Ambient bioassessment sampling means that you would identify a set of sites using the non-point source sampling approach as well as including sites that are above and below chronic problems, suspected problems, or land-uses of concern.

Step 2 Select sites (riffles) according to the types of impacts/conditions you wish to study.

3) Describe physical habitat and chemical water quality characteristics

There are two main types of habitat quality that are measured with the 2003 & 2006 CSBP – water quality and physical habitat quality. Chemical measurements are taken prior to BMI sampling. Physical habitat measurements can be taken after biological sampling.

Water Quality

The four water characteristics measured at the time of BMI sampling are: temperature, specific conductance, dissolved oxygen, and pH. Standard EPA or SWRCB protocols should be used for these measurements.

Physical Habitat Quality

a) Measure reach length, riffle length, average riffle width, riffle slope, and average riffle depth.
b) Estimate riffle velocity by measuring the rate of movement of a floating object, or (2006 CSBP) measure flow velocity using a flow meter at a single representative transect.
c) Estimate or measure canopy cover over the riffle by eye or using a spherical densiometer.
d) Estimate substrate complexity and embeddedness in the entire riffle length.
e) Estimate proportion of riffle in sediment categories ranging from fines to large boulders.
f) (2006 CSBP) Measure particle size frequencies (Wolman, 1954 technique, 5 particles) for each of the 11 major transects (n=55 particles) and 10 inter-transects (n=50 particles). Sample a single particle at each bank and at ¼, ½, and ¾ the width of the creek.
g) (2006 CSBP) Estimate percent that the sampled particles at each transect are embedded in fine sediments.
h) (2006 CSBP) Use inter-transect distances and elevation changes to calculate average reach slope.
i) (2006 CSBP) Record any of the various categories of human activities present in the riparian centered on each transect.
j) (2006 CSBP) Record size, type, and condition of riparian vegetation and bank stability.
k) (2006 CSBP) Record in-stream habitat complexity, including natural and human elements.
l) (2006 CSBP) Record particle size frequencies at the 10 inter-transects.
m) (2006 CSBP) Record flow-based habitat types at each of the 11 major transects.
n) (2006 CSBP) Measure bank-full width and multiple depths at a single representative transect.

The data from these measurements should be entered into the California Bioassessment Worksheet, available in the CSBP (Harrington and Born, 2003).

Step 3 Measure water quality and physical habitat attributes in the sampling riffles and reaches.

4) Sample BMIs from benthic sediments at each site

For the 2003 CSBP, the selection of locations to sample within a riffle sampling site varies slightly between point source and non-point source sampling designs. This difference is described in (2) above.

The process of collection is essentially the same for any of the methods described in the CSBP. a) Place a D-shaped kick net (500 micron mesh size) on the bottom perpendicular to flow. b) Disturb an area immediately upstream of the net measuring 1 foot (net width) by 1 foot (2006 CSBP) or 2 feet (2003 CSBP) and several inches deep. c) Sort the sample to remove debris. d) Place the sample into 95% ethanol to kill and preserve organisms. e) Combine the collections from the transect into one composite sample.

For the 2006 CSBP, two types of sampling are conducted – multiple habitat (MH) and targeted riffle composite (TRC). For MH, an independent sub-sample is taken at each of the 11 major transects. The position of the 1 square foot sampling area is alternately placed at distances from the bank of 25% stream-width, 50% stream-width, and 75% stream-width. BMIs from each sampled area are stored as an independent sample and then combined into a composite sample. For TRC, 8 randomly-chosen sites are sampled among targeted riffles or moving water habitats. Sampling is conducted using the same net for all sites until all 8 sub-sampling collections are present in the bottom of the net. The net is emptied into a single container.

> ***Step 4*** Collect BMIs with a kick-net at sampling locations, preserve BMIs, and combine samples from one transect into a single sample (2003 CSBP) or from multiple transects into a single sample (2006 CSBP).

5) Identify the BMIs to either an "Order" (Level 1), "Family" (Level 2), or "Genus/species" (Level 3) taxonomic level

The collected BMIs can be identified by a professional aquatic invertebrate biologist, or by a well-trained citizen monitoring team. The level of identification will be determined by the professional or volunteer training and will in turn determine the uses of the data. If you are paying for the identification, this step is one of the more expensive in the protocol. A well-trained cadre of citizen monitors can expect to identify BMIs to the Family level (Level 2), and could employ a professional for the identification of BMIs to the genus/species (Level 3) and/or to check volunteer identification at the family level.

The 2003 CSBP (Chapters 9 & 11) describes in detail the chain of custody requirements, equipment needed, time required, and quality assurance/quality control steps for laboratory identification. The short version is:

- ☐ you need various forms to describe and track what you did with the samples,
- ☐ you need containers, forceps, taxonomic key, and microscopes
- ☐ you need about 300 hours of identification time for 6 samples, to Level 2 taxonomy
- ☐ you need dedicated laboratory and storage space for identification and sample storage

> ***Step 5*** Sort and identify BMIs to the Family or Genus/Species level, depending on need.

6) Record the various metrics of BMI community structure

There are four types of biological metrics in the CSBP: "richness" – measure of diversity of taxa, "composition" – measure of proportion of particular taxa, "tolerance" – sensitivity of taxa to disturbance, and "functional feeding groups/trophic groups" – indicator of proportions of different types of feeding among taxa (Chapter 10, 2003 CSBP). Each of the metrics is thought to respond in a particular way to human and natural disturbance. Individually, the metrics are not as robust as having several or many metrics for a given waterway/watershed and should be used together. The table below shows the metrics, what they mean, and how the metric changes in response to disturbance (from tables and text Chapter 10, pages 2 – 6).

Table 1 CSBP Levels 1 and 2 benthic macroinvertebrate taxonomic identification. Metrics without grey highlighting are common for both Levels. Metrics highlighted in light gray are for Level 2 only. Metrics highlighted with gray hatching are Level 1 only.

Biological Metrics	Description	Response to Disturbance
Richness Measures		
Taxa Richness	Total number of taxa found	Decrease

Ephemeroptera Taxa	Number of taxa within this group (mayflies)	Decrease
Plecoptera Taxa	Number of taxa within this group (stoneflies)	Decrease
Trichoptera Taxa	Number of taxa within this group (caddisflies)	Decrease
EPT Taxa	Total number of taxa in the Ephemeroptera, Plecoptera, and Trichoptera groups	Decrease
Composition Measures		
EPT Index	Percent of all benthic macroinvertebrate taxa that are EPT taxa	Decrease
Sensitive EPT Index	Percent of all benthic macroinvertebrate taxa that are EPT taxa with tolerance values of 0 through 3 (“t-values” given in CSBP, Appendix C)	Decrease
Percent Hydropsychidae	Percent of all benthic macroinvertebrate taxa that are Hydropsychidae taxa (more tolerant caddisfly taxon)	Increase
Percent Baetidae	Percent of all benthic macroinvertebrate taxa that are Baetidae taxa (moderately tolerant mayfly taxon)	Increase
Tolerance/Intolerance Measures		
Tolerance Value	Weighted tolerance value for whole sample (number of organisms per taxa times t-value for taxa; sum this value for all taxa in sample; divide by total number of organisms in sample)	Increase
Percent Intolerant Organisms	Percent of all benthic macroinvertebrates in sample that are intolerant of disturbance (have tolerance values of 0, 1, or 2)	Decrease
Percent Tolerant Organisms	Percent of all benthic macroinvertebrates in sample that are very tolerant of disturbance (have tolerance values of 8, 9, or 10)	Increase
Percent Dominant Taxa	Percent of all benthic macroinvertebrates in sample that are in the most abundant taxon	Increase
Percent Stoneflies	Percent of all benthic macroinvertebrates in sample that are stoneflies	Decrease
Trophic Measures		
Percent Collectors	Percent of all benthic macroinvertebrates in sample that are in the functional feeding group “collectors” (CG), as designated in Appendix C, CSBP.	Increase
Percent Filterers	Percent of all benthic macroinvertebrates in sample that are in the functional feeding group “filterer” (FC), as designated in Appendix C, CSBP.	Increase
Percent Scrapers	Percent of all benthic macroinvertebrates in sample that are in the functional feeding group “scrapers” (SC), as designated in Appendix C, CSBP.	Variable
Percent Predators	Percent of all benthic macroinvertebrates in sample that are in the functional feeding group “predators” (P), as designated in Appendix C, CSBP.	Variable
Percent Shredders	Percent of all benthic macroinvertebrates in sample that are in the functional feeding group “shredders” (SH), as designated in Appendix C, CSBP.	Decrease

Step 6 Calculate the metrics appropriate for the taxonomic level used in identification.

7) Analyze basic statistical properties of the metrics (e.g., mean and variation) and comparing metrics among reference and impacted waterways or reaches and/or among waterways and expected values (e.g., correlation analysis).

The calculation of the mean and variation for BMI data is the same as for any quantitative data. The 2003 CSBP describes using the Student t-test and Analysis of Variance (ANOVA) as ways to compare two sets of data (e.g., upstream and downstream of a point source of pollution). The use of these tests should be limited when using BMI metric data because of the nature of the data and the impact of sampling design and natural variation on the numbers obtained. The values and metrics calculated include absolute numbers, proportions (percentages), and weighted proportions. The number and distribution of sampling sites between disturbed and control conditions, and the number of these sites compared to all possible sites within the study area impact the carrying out statistical analyses and interpreting results. In addition, because there is natural variation in the physical nature of riffles, both within riffles and among riffles, the samples collected potentially represent different natural physical conditions, which may confound their comparison with other samples from nearby riffles. The US EPA has an online "statistical primer" for use with biological and other data (http://www.epa.gov/bioindicators/primer/index.html) that includes basic definitions and descriptions of different analytical choices.

The mean and standard deviation values for the three samples composing a metric such as EPT taxa can be calculated and the mean for one riffle compared to the mean for another riffle using the Student t-test or among multiple riffles using ANOVA. As described in the CWAM, Chapter 5, there are many resources available for conducting statistical analyses. If you use MS Excel, the Help menu in this program can guide you through conducting simple comparisons of samples, correlation/covariance analysis, and regression analysis. The most critical aspect of conducting statistical analysis with your BMI metrics is that you are confident that the question you are asking will be addressed by the statistical test, the quality of the data, and the amount of data available.

The 2003 CSBP describes one approach to comparing BMI metric data (comparison of means). It also mentions trends analysis and comparison of values measured with "biocriteria" for waterways. As discussed in the CWAM (Chapter 5), trends analysis of natural systems is a very involved process, requiring knowledge of the system and very good knowledge of statistics. Most environmental processes have some periodicity or cycling associated with them, which is often due to short and long-term climatic changes. Therefore, any trends analysis for BMIs must go beyond linear regression or similar analysis and include the potential effects of cycling changes in the environment (e.g., the "El Nino" cycle). In addition, human land and water use and management can have cycles that are at odds or different from natural cycles. In the case of waterways, flows may be much higher than natural in the summer to provide for irrigation water, or power generation. This will impact water temperatures and other in-stream processes.

Step 7 Select statistical analyses for the bioassessment data that support addressing the assessment questions.

D. Minimum requirements

There are several minimum requirements for the quality of BMI data included in a watershed assessment. Meeting these benchmarks will maximize the robustness of decisions made using these data.

1) Number of sites

The term sites here means riffles that are sampled using the 2003 CSBP approach. For each riffle, there are 3 transects containing 3 sampling locations. Remember, the samples within each transect are combined.

The absolute minimum number of sites you would use for sampling BMIs using the 2003 CSBP is 2 within the point-source investigation framework. Even then, it would be better to have >1 upstream sites for reference and >1 downstream sites to understand natural variability and to understand the extent of impacts. A reasonable standard would be ≥3 sites above and ≥3 sites below the point of disturbance, which will improve your ability to control for natural variability among your control or reference sites.

For non-point source investigations, the choice of number of sites is less straightforward and may be just as dependent on available funding and expertise as the question you are trying to answer. Under one approach in the 2003 CSBP, random stratified sampling (Step 2 above), you would first group all reaches in the study area (watershed) into categories of “likes”. Once you have grouped the reaches, you would select a sub-set within each group to represent the entire group. This method may be more appropriate where there are many sub-watersheds and reaches to sample and choosing a representative sample is key. There are also non-random approaches to choosing sampling reaches, which are also based on physical and biological attributes of the reaches and often access to the reaches. This approach requires a field-intensive pre-survey and may be more appropriate for smaller watersheds with fewer possible sampling reaches. For either random or non-random site selection approaches, the number of sites is hard to predict.

The best approach to take would be to estimate the variability in BMIs sampled and use that estimate and the number of BMIs to be sampled per site to estimate the number of samples needed per condition (group of reaches) using statistical power analysis and sample size analysis (see Chapter 1, Vol. II).

2) Site distribution

The distribution of sites throughout your area of interest is important because they allow you to associate conditions in the watershed with BMI metrics at particular sites on particular waterways. The distribution of sites for point source pollution investigation is straightforward and is defined by the 2003 CSBP as being at least immediately above the point source and as near below the point source as is feasible. You should also choose other sites further upstream and further downstream to replicate your sampling at the immediate upstream and downstream sites.

For non-point source disturbance investigations, the best distribution of sites may be less intuitively obvious. In the most complex circumstances, that is, watersheds with many natural and human-induced conditions, you will need to employ a strategy such as the stratified random

The benchmark for appropriate site distribution is basically that you have sufficient sampling sites to represent the population of conditions that you want to assess using the BMIs as indicators of watershed and waterway condition. This is best determined using statistical analysis of sample size, which can be related to the number of sites sampled.

sampling approach described in the 2003 CSBP (see #1 above and Step 2 above).

3) Timing of sampling

Many BMIs are larval forms of terrestrial and aquatic invertebrates. They change form during larval development and after metamorphosis to the adult stage. This can complicate identification and also can make the timing of sampling important. Certain species will emerge from waterways and metamorphose on the edge of waterways, then fly away. If sampling for these species occurs when a cohort has already emerged, then they will potentially be found in smaller numbers than if sampled prior to emergence. Many species that will emerge do so in the spring and early summer. By late fall, few species are emerging and many that did so during the summer have already laid their eggs back into the waterway.

Timing BMI collection consistently (for a given waterway) during the late spring and fall will increase the among-year comparability of the data obtained (2003 CSBP, page 8-6).

4) BMI identification

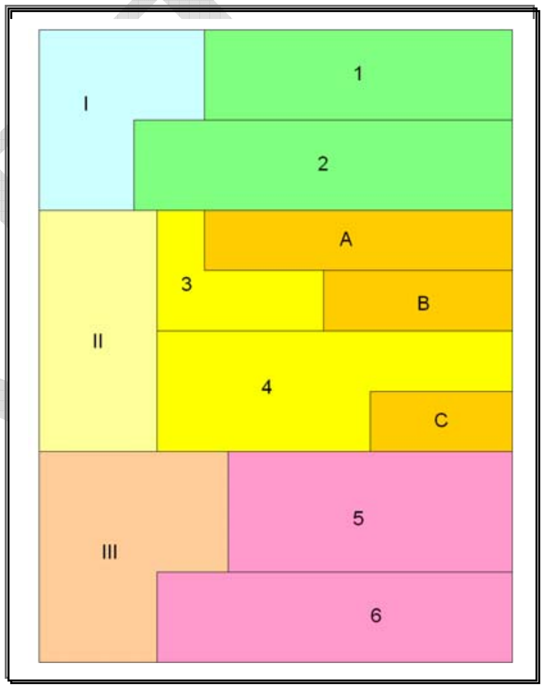
Chapters 12 to 18 of the 2003 CSBP describe the identification of BMIs collected during a monitoring/assessment program. There are two main ways that you can ensure that your identification process is giving consistently reliable and accurate results (see chapters 9 and 11). One is to save 10 to 20% of all of your sampled and identified BMIs, or 100% of a few samples, for confirmation identification by a professional/expert taxonomist. This is a way to calibrate your identification process, especially if you are using trained volunteers. Another way to ensure accuracy is to have all of the identification of BMIs be done by an expert taxonomist with verifiable credentials. Both ways cost, with the amount dependent upon the number of BMIs and the rate of the expert.

E. Spatial and temporal scale

There are a variety of scales over which BMI data can be used to assess waterway and watershed conditions. Obviously bioassessment accuracy increases with a higher intensity of sampling, more specific taxonomy, and the knowledge of influencing factors that can be considered the disturbance under investigation, or the environmental variables for which you would want to control. The case that may appear simplest on the surface (point-source investigation), still includes potential upstream impacts and influences, climatic variability, and dependence on sampling intensity. Here are some things to consider in relation to scale:

1) Single reach investigations You may be investigating a point source of pollution or the recovery of a reach after restoration through engineering/horticulture or management/ownership change. Having several sites above (control) and below (treatment) the reach of interest will help control for variation outside the scope of the point source or restoration. Things you can control for in this way include climatic variation, other land and water use practices above the reference and treatment sites, and natural disturbance above the sites.

2) Multiple reaches or watershed non-point source investigations Investigations of multiple sources and types of impacts to waterways involves several spatial scales, from points to rivers and sub-watersheds. A watershed-wide investigation will include many possible scales to make conclusions, depending on sampling intensity (number of sites and longevity of program). If you have more than 3 sites for a reach or creek, you may be able to draw conclusions about an average condition compared to a reference or previous condition. Because there will be 3 sampling transects per site, you may also be able to draw conclusions about changes from upstream to downstream, with the more sites the better. In the box below is a model watershed showing the different scales and potential scales for which BMI data and metrics could be useful. For Sub-watershed I you may have metrics for the mainstem of I and for creeks 1 and 2. The metrics for 1 and 2 could be evaluated separately and aggregated, if they are statistically similar, to contribute to a metric for the whole sub-watershed. For sub-watershed II, you could similarly compare A and B and if similar, they could be aggregated to contribute to a metric for 3, or maintained as distinct metrics. You could also compare A, B, and C and, if similar, aggregate them as metrics for a certain stream order or type. Among all 3 sub-watersheds (I, II, and III), you could compare the stream watersheds of a similar type or order (e.g., 1 to 6) and either identify one of them as a reference for the others, or develop a combined metric for all of them or a sub-group of them. Finally, you could develop a tiered or hierarchical metric system where you attribute condition scores based on an aggregation of BMI metric scores from the top of the sub-watershed down. For example, if scores for A and B are “high” and for C “low”, for 3 “high” and for 4 “medium”, then for II, the contribution of these subsidiary creek watersheds to condition could be “medium to high”.



In some ways the temporal scale is more difficult to manage for bioassessment indicators, primarily because it can take many years to develop a statistically meaningful indication of change. You might have many sites that satisfy statistical requirements for the spatial scale of your study, but it may still take decades to measure recovery of a system from the impacts to human activity, for example. Fortunately (or unfortunately), it may take considerably less time to measure the evolving degradation of a waterway’s biota, or the existing degraded condition. This contrast points out one of the most critical things to consider when designing your use of BMI data and metrics – what is the question you are trying to address and therefore what timescale is relevant.

F. What do the measurements mean?

The variety of metrics available from BMI investigations and what they mean are shown generally in Step 6 above. These biological metrics are designed to indicate a response to disturbance of some kind. Collectively they can provide a condition score of sorts, as is done with the “index of biotic integrity” (IBI) approach that has been used throughout the West with fish and macroinvertebrate assemblages.

Comparing to References

There is necessarily a lot of inference when deciding what individual BMI community metrics mean in the context of watershed condition. For metrics to be meaningful they either have to be related to a standard (e.g., drinking water standard), reference healthy condition (e.g., undisturbed waterway), or the least disturbed condition in the study area. There is a solid literature regarding appropriate use of the concept of “reference sites” before moving forward with identifying and sampling sites (Hughes et al., 1986; Hawkins, 2000).

As with any environmental monitoring protocol, BMI metrics are most robust when calibrated against known conditions from another watershed, or against a reference or control condition. If you have obtained certain EPT taxa richness values for a waterway, these values are best interpreted in the context of how what your expected richness is. These expected values could come from scientific investigations conducted in a similar or nearby waterway, or from a thorough understanding of the richness you would expect given the land cover, physiography, and in-stream habitat conditions. The more limited your knowledge of what the metric values should be, the more limited you will be in inferring impacts from the values. However, the information you have collected may become useful in future or more extensive investigations and should be reported in the watershed assessment so that they are not lost.

Coyote Creek Watershed Integrated Pilot Assessment (2003, EOA Inc. for SCBWMI)

In this study, BMI metrics were used as one way to characterize reach condition for 5 reaches in the watershed. Ten community metrics were calculated from a single collection in May, 1997, with identification to the species level.

The reference condition for comparison in this study was the highest metric value for each metric calculated (as opposed to a geographic reference). A benthic macroinvertebrate – index of biotic integrity (B-IBI) score was calculated for each reach and used in the discussion phase of the assessment.

The B-IBI scores were lumped into classes (e.g., medium) and used with other watershed information (e.g., landscape connectivity) to roughly rank reaches and project the potential changes in watershed condition in response to proposed management actions (e.g., flood control channels). Although the statistical power is not high in terms of data collected and reaches analyzed, there were easy to follow relationships between methods, results, and input to management planning.

http://www.valleywater.org/_wmi/Stewardship_plan/reference/PDF/Coyote%20Assmt%20Final%20%20Report.pdf)

Although there are quite a few research studies of BMIs in California, there are few standards for what might constitute a natural assemblage of BMIs in your watershed, or what assemblage might indicate a functioning watershed. In some places, the presence or absence of certain taxa or the dominance by particular taxa can strongly indicate that there are problems. This is particularly true of streams with riffles (fast moving water over rocks) where most sampling takes place. Slow-moving waterways with sandy or muddy bottoms have not been well-studied and for these there will be fewer references in the scientific literature.

One approach to use is the least-disturbed condition in a watershed or study area. This is the standard used by the EMAP program for their reference conditions. The strength of this approach is in having a reference condition within an area similar to or near to the reaches of concern or study. This makes it more likely that light, temperature, and other conditions will be similar between the reference and others sites. However, the reference can only act as such if other factors like stream gradient, elevation, and contributing watershed area are similar between the reference site/reach and the other reaches/sites. For example, a reference reach in a small creek/watershed at 3,000' elevation will probably not be an adequate reference for a river reach at 2,000' with a large contributing watershed area.

If you want to compare BMI metrics among sub-watersheds or reaches to understand the effects of disturbance (e.g., land conversion or fire), the best approach to take is to have reference conditions that are similar to the reaches of concern within the same watershed. In the context of a watershed assessment, this may mean selecting several reaches within least-disturbed sub-watersheds as the references, where disturbance is based on things like mapped human or natural disturbances, known sources of pollution, or potentially-impacting infrastructure (e.g., dams/diversions or roads within the riparian). You could then use BMI metrics to assess the effects of types of activities (e.g., all dams), specific activities (e.g., your town wastewater discharge), or just to find which reaches/waterways have degraded BMI community conditions.

Reporting BMI Metrics Within an Assessment

As is the case with the CSBP, we can't anticipate all possible conditions where BMI metrics will be used in a watershed assessment. However, here are some guidelines for their use within an assessment:

Number of reference vs. study sites/reaches
When making conclusions about disturbance for individual waterways or sub-watersheds within a larger watershed, it is important to compare your study sites (sites of concern) with a reference condition (if known) or appropriate reference site/reach. Valid conclusions can be made about condition and disturbance if there sufficient sites and reaches that are physically similar to the site/reach of concern, but lack the disturbance which you are concerned about. You will only want to compare sites with other sites of a similar type. Similarity will be based on things like watershed area, elevation, stream gradient, water temperature, upstream and adjacent land cover, and watershed geology. As discussed previously, reference sites can be within relatively undisturbed sub-watersheds (ideal), outside the watershed but within otherwise similar sub-watersheds, and the least-disturbed condition within your watershed. At least three reference sites within a type or class of sites is the best scenario because these can be compared with each other to get a statistical measure of natural variability. Study sites can also be compared to each other as well as to reference sites.

What constitutes a significant difference or similarity?
You will probably want to make statements about how similar your study sites (sites of concern) are to each other and/or to reference sites and conditions. You can calculate similarity based on comparisons between individual BMI community metrics for each site or groups of sites. You can also calculate similarity based on multiple metrics at the same time, which is a more robust approach. The calculations of similarity are conducted using statistical tests appropriate for the type of metric data used. For example, to compare percentages with each other, which would be the case for most of the community metrics in Table 1, you would use different tests than when comparing diversity or richness values.

In some cases the number of certain specific identified organisms will be low, meaning that the metrics that include those taxa may be hard to compare to the same metrics for other similar sites. For example, you might only find 5 stonefly larvae in your sample of 500 animals at site 1 and find 10 stonefly larvae at site 2. Site 2 has twice as many stonefly larvae, but that only means 5 more out of 500 total sampled BMIs. In cases like this, you may not be able to say much about individual taxa, but that is why the community metrics can be useful.

How many metrics to use

Usually, the more metrics you can use the better. If you have samples of 500 organisms per site, you will probably be able to calculate most or all of the metrics listed in Table 1. Even though you can calculate and potentially use metrics, as indicated above, you should only use metrics for which the natural variability isn't going to wipe out any observed differences in metric values among sites.

You can combine metrics to give you a composite index of conditions if you want. However, this value must be calculated carefully because, as pointed out by the San Pedro Creek investigators (see text box), many of the metrics are not independent of each other. The impact of this is to inflate the importance of extreme taxa values on the composite score.

G. Appropriate use of data/limitations

One reason that BMI investigations are not often used as part of watershed assessment is that to draw conclusions from the results, you need a lot of sampling and laborious taxonomic work. If this hasn't already been done by the Department of Fish and Game Aquatic Bioassessment Lab or other expert group, then the assessor will have to organize and implement this work. The implication here is that a small amount of BMI data may not be worth collecting because there will be the temptation to inappropriately draw conclusions from them. Similarly, even where there has been sufficient data collection, an expert knowledge of how BMIs respond to watershed and waterway conditions and statistical analysis are both required to use the data.

In the event that the assessment team doesn't feel comfortable assigning meaning to BMI community metric data, the data can still be described and incorporated into the database of existing information that should be developed as part of watershed assessment, monitoring, and adaptive management. This will make it easier for future efforts to be justified and build on existing information.

Benthic Macroinvertebrate Assemblages of San Pedro Creek, San Mateo County (Bioassessment Services and EOA Inc., 2002)

The San Pedro Creek Watershed Coalition (near Pacifica) has assessed condition in various ways, including by investigating BMI distributions. The investigators relied on volunteer samplers, trained by experts, and professional BMI taxonomists. Appropriate statistical approaches were used to study how related sites were to each other based on BMI metric values. Although the findings don't appear to have been incorporated into the Coalition's "Watershed Assessment and Enhancement Plan" for the Creek, the findings from the BMI study were congruent with expectations about condition and disturbance based on other types of information (e.g., from land-cover in GIS).

The report is entitled: and can be found at: http://www.pedrocreek.org/research.html

There are several main types of limitations to the use of BMI metric data in watershed assessment:

Spatial extent

Many BMI sampling efforts are limited in scope due to the amount of money required to identify hundreds of organisms per sample and site. For a 100,000 acre watershed spanning 1,000 meters in elevation and multiple land uses, you could need dozens of sites to adequately address questions about sources and severity of disturbance to aquatic biota. If you have funding for only 5 sites and you spread them throughout the watershed, then you will have inadequate information about the whole watershed.

Temporal extent

A potential solution is to place these sites in a single sub-watershed that is in some way representative of other sub-watersheds and contains the disturbance conditions you are most concerned about.

As is the case with spatial extent, many BMI monitoring projects are limited to a short time period. Often only one sampling will take place, while in the best cases back-to-back funding may provide 3 to 5 years of sampling and identification. Aquatic insect larvae are an important part of the BMI community. They emerge after several months to 2 years of development from the egg (depending on species and water temperature), spend days to months as an adult insect, and then reproduce and lay eggs in the same or different waterway. There are many natural cycles that can impact the process of egg-laying and development, which will in turn impact the BMI community sampled and assessed. If you can only sample for one year, each site will be subject to some unknown phase of natural cycles that could determine the occurrence of certain taxa. For example, the occurrence of midge larvae is sensitive to summer water temperature (Walker et al., 2003). These temperatures could vary with climatic oscillations (temperature and precipitation), water management, and up-watershed land cover (for shading). In a given year, the number of individual midge larvae and the number of midge taxa could be impacted more by natural conditions than disturbance and pollution.

Sampling intensity

A potential solution is to distribute limited sampling across several years and try to control for natural variability that may be controlled by climatic cycles and of natural influences.

Sampling intensity is based on the number of samples and number of sampling events. If you follow the 2003 CSBP, at a given site you will collect composite samples of between 300 and 900 BMIs (for CSBP Level II and III respectively) at one location during a sampling event. Natural variation means that the site at which you sample may or may not be representative of the reach or waterway. This can be statistically determined by sampling at least 3 locations and comparing the locations with each other. In addition, if your site is intended to represent a class or type of sites and reaches, then it may or may not do so, which can be statistically determined by sampling at ≥3 sites per class/type. Sampling intensity can determine whether or not you can conclude anything about conditions at the sampling site, the reach, the type of site/reach, and the conditions at the site relative to other sites (including reference sites). Adequate sampling intensity means that you will collect sufficient samples at sufficient number of sites to account for natural variation among sites and within BMI taxa at individual sites. Sufficiency can be statistically estimated using sample size analysis, which is related to power analysis (see CWAM Vol I, chapter 5).

H. State and Federal Programs Using This Basic Approach

EMAP

The US Environmental Protection Agency operates a program called the “Environmental Monitoring and Assessment Program” (EMAP, http://www.epa.gov/emap/) which uses multiple environmental parameters, including BMI metrics, to evaluate condition of the country’s waterways and watersheds. This is the most extensive program in the country, but information about individual watersheds is often limited to a few sites (depending on watershed size).

The EMAP samples perennial streams at over 50 reference and over 180 “probalistic design” sites. Probalistic design is a way of randomly selecting a limited set of sites that represent a larger set of sites and is described at: http://www.epa.gov/nheerl/arm/designpages/design&analysis.htm. EMAP reports in various scale units: 1) the West, 2) climatic units, and 3) aggregated ecoregions with the Sierra Nevada, California Cascades, North Coast ranges combined with Oregon mountains as the “Pacific Northwest mountains”. The reference conditions are defined as minimally disturbed or the least disturbed in the study area, or the best attainable conditions. The definition for the least-disturbed condition varies across California with climate/ecoregion.

There are multiple metric groups measured by EMAP: BMIs, aquatic vertebrates, water chemistry, physical habitat, fish tissue contamination, and the presence of alien taxa with multiple metrics within each group. Links are made between stressors and conditions and assessments made based on certain indicators that are grouped in slightly different ways than groups of metrics. The possible BMI indicators are screened for appropriateness and the indicator scores re-scaled to a 0-100 index. The individual metrics used vary by ecoregion. An index is calculated from the sum of the metrics and ranges from 0 to 100.

RIVPACS

One of the most influential ways that has been developed for thinking about using BMI metrics to indicate environmental condition is the River Invertebrate Prediction and Classification System (RIVPACS, http://dorset.ceh.ac.uk/River_Ecology/River_Communities/Rivpacs_2003/rivpacs_introduction.htm). The system relies on the ratio of observed to expected biota as an index of condition. The expected biotic community is derived from models that are site-specific and standardized for sites. It is a measure of the taxonomic completeness of the biological community observed at a particular site.

The value exported from the RIVPACS is “Observed/Expected” biota (O/E), which is based in a logical design that allows comparability among very different types of sites. It is based on statistical models for the probability of capture (“pc”) of any taxon in any place, which are calculated for the particular sites. The sum of the taxa-specifc “pc’s” is an estimate of the the number of taxa that should be observed at the site given standard sampling among sites. You do need to estimate the prediction error for site assessments so can compare sites to “reference” conditions.

One approach to comparing sites with each other is to first conduct cluster analysis to form regional classes based on environmental conditions, such as flow variability, temperature, precipitation, and latitude/longitude of the sites. You can then use fitting models (e.g., the “R” model system) to determine which environmental conditions and variables explain O/E values among sites within a class. In previous studies, watershed area, latitude/longitude, and temperature were the best predictor conditions to explain variability in O/E values.

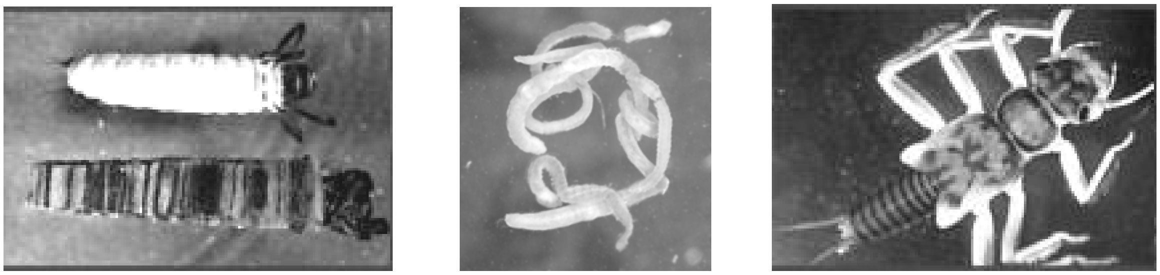
There are several caveats to the use of this approach: 1) taxonomy must be consistent among sites being compared, 2) the sampling methods must be similar, and 3) the reference site quality must be similar to the sites of concern.

CMAP

The California Department of Fish and Game has a program called California Monitoring and Assessment Program (CMAP, http://www.dfg.ca.gov/cabw/Field/cmap.html) that is similar in conceptual scope to the EMAP program, but has more sites.

CMAP also uses probabilistic design for selection of sampling locations. The sample reaches are based on segmented streams maps with unique identification numbers and known lengths. A “reverse bias” is used to de-emphasize smaller streams, because they are more prevalent than larger streams, so that larger waterways are well-represented. After initial site selection, the sites are filtered for major confounding factors like access across private property, consistent stream dryness, and the stream being encased in concrete (in urban areas). Regional and statewide surveys are conducted, with the regions being North, Central, and Southern Coast areas.

Environmental metrics sampled include attached benthic algae, riparian habitat characteristics, and physical habitat measurements. One of the main purposes of the CMAP program has been to develop regional indices of biotic integrity (IBIs). The scale at which these IBIs are valid varies with the sampling intensity within the region, watershed, and sub-watershed.



(photos from DFG-Aquatic Bioassessment Lab Web site)

Individual BMI community metrics or IBIs can be compared with values for relative risk or stressors. Examples are high chloride concentrations or fines/sands in the benthos compared with predominance of pollution-tolerant genera or changes in feeding groups and loss of pollution-tolerant genera.

Washington’s Stream Monitoring Program

The state of Washington’s Department of Ecology also has its own stream benthos monitoring program that is similar to the EMAP program in terms of goals and protocols (http://www.ecy.wa.gov/programs/eap/fw_benth/index.html). The DoE provides reports of their findings and methods at: http://www.ecy.wa.gov/biblio/bioassessment.html.

I. References

Aquatic Bioassessment Laboratory (2006). Standard operating procedures for collecting benthic macroinvertebrate samples and associated physical and chemical data for ambient bioassessments in California. California Department of Fish and Game Water Pollution Control Laboratory. Rancho Cordova, CA.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.D. Stribling (1999). Rapid Bioassessment Protocols for Use in Streams and Wadeable River: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Harrington, J. and M. Born (2003). Measuring the Health of California Streams and Rivers, Second Edition. Sustainable Land Stewardship International Institute.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, R.J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541-556.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: A method for assessing stream potentials. Environmental Management 10:629-635.

Walker, IR ; A.J. Levesque; R. Pienitz; J.P. Smol. 2003. Freshwater midges of the Yukon and adjacent Northwest Territories: a new tool for reconstructing Beringian paleoenvironments? Journal of the North American Benthological Society. Vol. 22, no. 2, pp. 323-337.

Wolman, M.G. 1954. A method of sampling coarse river-bed material: Transactions of the American Geophysical Union 35: 951-956.