

Chapter 6.0 Benthic Macroinvertebrates

By Brent R. Johnson¹, James B. Stribling, Joseph E. Flotemersch and Michael J. Paul

This chapter...

- reviews existing large river macroinvertebrate sampling methods
- recommends a bank-oriented multi-habitat approach

Macroinvertebrates are...

- important components of large river food webs
- proven indicators of biological condition
- responsive to a wide range of stressors

6.1 Introduction

Benthic macroinvertebrates include aquatic insects, crustaceans, annelids, mollusks, nematodes, planarians, bryozoans, cnidarians (*Hydra*), and nemerteans. They inhabit sediments or live on bottom substrates of aquatic ecosystems. At least some representatives of this assemblage can be found in virtually every freshwater environment on Earth.

Macroinvertebrates, specifically, are invertebrates retained by a mesh size of 500 μm (Hauer and Resh 1996). While early developmental stages may pass through a mesh of this size, 500-595 μm is generally considered suitable for biomonitoring purposes (e.g., Klemm et al. 1990, Barbour et al. 1999, Lazorchak et al. 2000). Smaller mesh sizes are required for ecological studies that focus on life histories and secondary production, and those that include meiofauna. Macroinvertebrates play a critical role in the transfer of energy from basal resources (e.g., algae, detritus and associated microbes) to vertebrate consumers in aquatic food webs, and they serve as the primary food resource for many commercially and economically important fish species.

Benthic macroinvertebrate are the most common faunal assemblage used in bioassessments of wadeable streams and rivers (e.g., Rosenberg and Resh 1993, Barbour et al. 1999, USEPA 2002, Carter and Resh 2001). After careful sampling using standardized field collection methods, laboratory species identification and enumeration, evaluation of structural and functional attributes of the assemblage are used to evaluate biological condition. The following factors have contributed to their becoming so widely used in biomonitoring programs (modified from Barbour et al. 1999):

- Macroinvertebrates are ubiquitous and abundant in most streams and rivers, including headwater streams where fish may be absent.
- Macroinvertebrates are relatively sedentary in the aquatic environment so they are good indicators of local condition.
- Many taxa are long-lived (1 year or more) and, thus, integrate short-term disturbances and reflect long-term site condition.

¹ US Environmental Protection Agency, National Exposure Research Laboratory, 26 W. Martin Luther King Blvd., MS 642, Cincinnati, OH 45268

- Macroinvertebrates are diverse in their habitat requirements, feeding modes and tolerance to pollutants and other stressors (e.g., low dissolved oxygen, temperature changes and sedimentation). They, therefore, provide valuable information about ecosystem health and source(s) of impairment.
- In most cases, sampling macroinvertebrate assemblages is relatively easy, requiring few people and inexpensive gear.

Despite their widespread use in streams, benthic macroinvertebrates have rarely been incorporated into formal bioassessments of large rivers. There is a general belief that macroinvertebrate assemblages become less diverse and more tolerant in large rivers (i.e., that the replacement of sensitive stoneflies and other “coldwater” taxa is a common occurrence). The unstable fine sediments typical of many large river bottoms generally support fewer taxa than smaller streams and rivers that have larger substrate sizes (Allan 1995). Due to the long history of benthic sampling in smaller streams, most of the common quantitative and qualitative methods for sampling macroinvertebrate assemblages require easy access to substrates.

Macroinvertebrate sampling in large rivers presents programs with several difficulties common to all assemblage surveys relating to spatial scale and sampling logistics:

- The diversity of habitat types in large rivers (e.g., back channels, inlets, floodplain wetlands) makes it difficult to obtain a standardized and representative sample.
- Balancing the appropriate reach length with time and cost constraints for macroinvertebrate assessment is more difficult as repeating habitat units are spaced farther apart and meander wavelength increases.
- Identifying reference conditions for large rivers is difficult due to the large areas of intensive human land use.
- Identifying specific stressors or causes of impairment, as required by the CWA §303(d), is more difficult in large rivers because of the cumulative impact of multiple stressors that result from disturbances within large drainage areas.
- Large river macroinvertebrate sampling is more costly and hazardous than on wadeable streams because it typically requires use of a boat on navigable waterways that are often subject to commercial traffic.

Despite these obstacles, many researchers have sampled large river macroinvertebrate assemblages for inventory and monitoring purposes or for targeted sampling around point sources of pollution. More recently, efforts have increased to standardize large and great river macroinvertebrate assessment programs (Lazorchak et al. 2000, Merritt et al. 2005, Angradi 2006). There is a lack of assessment information that characterizes the condition of large rivers and the need for these bioassessment programs has risen with this recognition. Table 6-1 provides a brief summary of five of these large river bioassessment programs. Michigan DEQ’s macroinvertebrate bioassessment program is also highlighted in this chapter.

TABLE 6-1. A comparison of large rivers program macroinvertebrate sampling approaches.

Program	Protocol Summary	Citation
USEPA EMAP- Surface Waters	An acceptable sampling point is identified in an area away from the river margin and less than or equal to 1 m depth. Two kick net samples are taken at each of 11 transects and composited. Samples are placed in a bucket, detritus is removed without removing the macroinvertebrates. Samples are placed in plastic jars and filled with 95% ethanol to preserve the sample.	Lazorchak et al. 2000
USGS NAWQA Program	The types of instream habitats are recorded and semi-quantitative samples are taken to determine relative abundance when it is possible. Semi-quantitative samples are taken from the richest targeted habitat (RTH). Typically, this is riffle habitat or woody snags. A 0.25 m ² area is sampled using a slack sampler (500-µm mesh) in riffles. Two snags are sampled by disturbing snags upstream as a sampler for woody snag sampling. Area of the snags sampled is estimated for that habitat. Qualitative samples: Proportional multi-habitat samples are taken along the study reach. Samples are taken with a D-frame kick net and visual collections and some grab collections are made. Water depth and substrate type are recorded. Large debris is removed along with large crayfish, hellgrammites and mussels. The sample is placed in a standardized bottle with a 10% buffered formalin solution.	Moulton et al. 2002
Ohio Environmental Protection Agency (OEPA)	<p>Quantitative methods: A modified Hester-Dendy (H-D) multiple-plate artificial substrate sampler, with eight plates and 12 spacers, is placed in the river and tied to a concrete construction block. In rivers more than four feet deep, a floater is attached to keep it within four feet of the surface. Whenever possible, the samplers are placed in runs. A sample consists of three multiple-plate samplers. Samples are retrieved by cutting them from the block and placing them in one-quart plastic containers while still under water. Formalin is added to make a 10% solution. Qualitative samples are collected at the same time for organisms in the natural substrate.</p> <p>Qualitative methods: Each station is sampled at least once between June 15 and September 30. If possible, a riffle, run, pool, and margin are sampled at each site. Organisms are collected using a triangle ring frame 30-µm mesh dip net and field picked with forceps for at least 30 minutes until no new taxa can be identified. The organisms are preserved in 70% ethanol.</p> <p>In both methods, a station description sheet is filled out and the length of time spent sampling is recorded.</p>	Ohio EPA 1989
Kentucky Division of Water (KDOW)	The 20-jab method is used augmented by dredge samples, a wood sample, and rock picking along a 300-meter reach of the river. The sample is placed in a 600-µm mesh washing bucket where the macroinvertebrates are removed and placed in 70% ethanol. When possible, 15 large rocks and 6 m of wood are picked and washed.	Kentucky DOW 2002
Michigan Department of Environmental Quality (MIDEQ)	The individual habitat types are counted. Habitats must be within the littoral area and large enough to collect a 15-second sample. A 15-second sample is taken for every habitat type with a D-frame net, with a mesh size of 500 µm. The net is emptied into a bucket or pan filled with water. Detritus is removed before placing the sample in a 500-µm sieve to remove excess water. The sample is placed in 95% ethanol.	Merritt et al. 2005

PROGRAM HIGHLIGHT

Qualitative Biological and Habitat Survey Protocols for Michigan's Non-Wadeable Rivers Submitted to the Michigan Department of Environmental Quality (Michigan DEQ) (Merritt et al. 2005)

The Michigan DEQ is responsible for water quality monitoring in the state. As part of their Strategic Environmental Quality Monitoring Program, they have conducted or are conducting biological and habitat surveys across the state to assess more than 80% of their stream and river miles. The specific goals of their program are to:

1. determine whether waters of the state are attaining standards for aquatic life,
2. assess the biological condition of the waters of the state,
3. determine the extent to which sedimentation in surface waters is impacting indigenous aquatic life,
4. determine whether the biological condition of surface waters is changing with time,
5. assess the effectiveness of best management practices (BMPs) and other restoration efforts in protecting and restoring biological integrity and physical habitat,
6. evaluate the overall effectiveness of DEQ programs in protecting the biological integrity of surface waters,
7. identify waters that are high quality or not meeting standards, and
8. identify the waters of the state that are impacted by nuisance aquatic plants, algae, and bacterial slimes.

The Michigan DEQ has an existing rapid assessment protocol for wadeable streams, but it is not applicable for their non-wadeable rivers. They contracted with Michigan State University scientists to develop a non-wadeable method for assessing macroinvertebrate and habitat condition.

Michigan DEQ Macroinvertebrate Sampling Methods

The Michigan DEQ macroinvertebrate method was developed using data from 45 locations on 13 non-wadeable rivers from across the state. The approach requires sampling between June and September during stable discharge and is designed to take approximately 0.5 days for a two-person crew. The sampling unit is a 2000-m reach split into 11 equally spaced transects. Along each transect, two littoral (20-m long X 10-m wide) plots are established. One plot, chosen by a coin flip, is sampled at each transect. If large woody debris (LWD) is present along eight of the 11 transects, then only LWD is sampled. If not, then all available habitats are sampled in each plot (fine particulate organic matter (FPOM), sand, gravel, cobble, LWD, and macrophytes). Each available habitat is sampled for 15 seconds using a D-frame dip net with 500- μ m mesh. If flow is insufficient, nets are swept through the habitats. For cobble, a cobble of at least 15-cm in width is placed in a bucket and brushed with a toilet brush. Similarly, LWD is brushed either above the kick net or the kick net is swept through the water. The net is swept through macrophytes for 15 seconds to dislodge organisms. Each sample is placed in a white enamel pan with water and the nets are cleaned. The pan material is sieved (500 μ m) to remove excess water and placed into a bucket with 95% ethanol. Individual transect samples are composited into one bucket. A plankton splitter is used to divide the composite sample into quarters. All the individuals in the quarter sample are counted and identified to family level. The macroinvertebrate data are used to calculate 13 individual metrics combined into an overall multimetric score for each site. The individual metrics are Plecoptera richness, EPT richness, Diptera richness, percent dominance, percent Diptera, total richness, functional feeding group diversity, and the ratio of (#scrapers + #collector-filterers)/(#collector-gatherers + #shredders). Individual metrics are scored differently depending on whether the multihabitat or LWD sampling methods are used, and different metrics are weighed differently based on how much among-site variability they explained. Final scores are broken into four classes: 0-15 (poor), 16-30 (fair), 31-45 (good) and 46-60 (excellent). For detailed descriptions of the metric development, please contact Michigan DEQ.

This chapter provides a review of several different active and passive methods for benthic macroinvertebrates in large rivers. It also gives recommendations for a protocol (Flotemersch and Blocksom 2004, Flotemersch et al. 2006) borne from some of these methods. If field sampling methods other than those recommended here are more suitable for a particular program, they should be thoroughly tested to ensure that they return data of sufficient quality and provide the capacity to address their intended and stated purposes.

6.2 Field Sampling Methods

Numerous studies have demonstrated that dramatic differences can exist among large river benthic sampling methods (Anderson and Mason 1968, Rabeni and Gibbs 1978, Slack et al. 1986, Diamond et al. 1994, Humphries et al. 1998, Leland and Fend 1998, Hoffman 2003, Poulton et al. 2003, Blocksom and Flotemersch 2005). Benthic grab/dredge samples or the use of artificial substrates have historically been the most common collection methods for large river macroinvertebrates and they remain common choices for many researchers. More recently, however, active sampling methods, such as kick net or D-net sampling along the shoreline and scraping large woody debris (LWD), have become more common in an effort to assess a river reach and to sample the most productive (per unit area) habitats for macroinvertebrates. Flow regime and substrate stability are major factors influencing distribution of large river macroinvertebrates. The location of benthic sampling within the channel can greatly influence results (e.g., high-velocity main channel vs low-velocity shoreline areas; fine sediments vs vegetation or larger mineral substrates). Most sampling methods, however, are only appropriate for, or artificially represent, one substrate type or area. A combination of methods and sample locations may prove best for assessment, but the choice of these methods should depend upon specific management questions and available resources. Numerous authors have provided comprehensive reviews of benthic macroinvertebrate sampling methods (Rosenberg and Resh 1982, Flannagan and Rosenberg 1982, Klemm et al. 1990, Merritt et al. 1996). The following sections provide a brief review of sampling methods as they relate to large river sampling.

6.2.1 Passive Methods

Passive methods include artificial substrate samplers defined by Klemm et al. (1990) as “devices made of natural or artificial materials of various composition and configuration that are placed in the water for a predetermined period of exposure and depth for colonization.” Artificial substrate samplers can be used to obtain qualitative and quantitative macroinvertebrate samples and they have been recommended for use in deep or turbid waters and in areas with muddy, sandy, or otherwise unstable bottoms (Taylor and Kovats 1995). Exposure periods are typically four to six weeks to allow for colonization of biofilm and subsequent macroinvertebrate fauna and samplers are usually deployed at 1- to 3-m depths. Deployment depth is chosen so that receding or rising waters during the exposure period will not leave samplers dry or too deep to retrieve and so the samplers will be in the photic zone. Typically, 4 or 5 Hester-Dendy’s (H-D’s) or 3 rock baskets are placed per sampling reach and the data are composited from all samplers retrieved. Placing multiple samples per reach and compositing data also helps buffer the effects of loss or vandalism. Upon retrieval, samplers are slowly lifted to the water surface. If possible, a net is placed downstream or around the sampler to collect any organisms that fall off or leave the samplers during removal. The samplers are placed in a bucket and the substrates are scraped or brushed into the bucket. The bucket contents are then sieved and preserved for laboratory processing. Alternatively, some choose to return the complete sampler to the laboratory for processing. Some advantages and disadvantages to using artificial substrate samplers are summarized in Table 6-2.

6.2.1.1 Rock Basket Samplers

Rock baskets are passive samplers that typically consist of plastic or wire baskets (e.g., square or cylindrical barbecue grilling baskets) filled with native rock or gravel. Baskets are typically tied to a rope that is fastened on the shore and then dropped into the river. Standard-sized quarry rocks can be used in baskets to help standardize surface areas and facilitate density calculations.

Rock basket samplers can have the advantage of providing a natural substrate with irregular surfaces and interstitial spaces that mimic those of the natural environment. However, rock baskets have the disadvantage of being slightly less standardized and quantitative than H-D type samplers. Rock baskets (similar to Figure 6-1) have been successfully used in Ohio (Anderson and Mason 1968, Mason et al. 1973), Maine (Rabeni and Gibbs 1978), Pennsylvania (Hoffman 2003) and along the Missouri River (Poulton et al. 2003). Rock-filled trays are similar to baskets and have been used to sample smaller streams (e.g., Townsend and Hildrew 1976, Clements 1991), but they are not as effective in large rivers due to their instability in fast currents.

TABLE 6-2. Advantages and disadvantages of artificial substrate samplers.

ADVANTAGES/DISADVANTAGES
Numerous researchers have described artificial substrate samplers and their relative advantages and disadvantages (Rosenberg and Resh 1982, Flannagan and Rosenberg 1982, Klemm et al. 1990, Merritt et al. 1996). Some of these are given below.
<i>Advantages</i>
1) Allow quantitative collection of benthic macroinvertebrates from sites that cannot be effectively sampled using other conventional benthic sampling methods.
2) Can be used effectively in shallow or deep water, making them useful for sampling throughout the large river mosaic.
3) Easy to use and usually require less time and effort in the field than active methods. The ease of deployment and retrieval helps reduce sampling variability associated with the operator.
4) Generally accumulate very little debris during incubations making sample processing more efficient.
5) Can be especially effective in reflecting water quality as a result of the standardized habitat they provide.
<i>Disadvantages</i>
1) Require two trips to the sample site (for deployment and retrieval) that can add time, cost and other logistical constraints.
2) Measure colonization potential rather than the resident assemblage.
3) Loss of individuals when retrieving the sampler can bias results.
4) Can effectively indicate water quality, but not sediment or other habitat quality.
5) Exact placement of individual sampler units can skew results (e.g., high vs low velocity).
6) Damage or loss of artificial substrates can occur due to vandalism, high flows, shifting channels or they may be left dry during drought conditions.



FIGURE 6-1. Rock-filled wire basket used as introduced substrate.

6.2.1.2 Multiplate Samplers

The most common type of artificial substrate samplers are variations of the H-D multiplate sampler (Hester and Dendy 1962). Many monitoring programs use these samplers for assessment of both point and non-point sources of pollution in large rivers. Configurations may vary greatly in size, shape, and number of plates used, but all consist of round or square plates (typically made of Masonite board or porcelain) with spacers placed in between and bolted together to form stacks (Figure 6-2). Spacing between plates is typically varied to provide different refuge sizes and flow regimes within the stacks. Stacks are tied together and attached horizontally to a brick or cinder block and placed on the river bed (Figure 6-2). Alternatively, stacks may be positioned vertically by screwing the bolts into the anchor blocks. These samplers have been successfully used on many large rivers, notably as part of standard programs in Florida, Wisconsin, and Ohio.

6.2.1.3 Other Passive Methods

Although rock baskets and H-Ds are by far the most common artificial substrates used in benthic studies, a number of other passive samplers may be used. Beak trays are round metal trays with expanded mesh inserts for colonization (Beak et al. 1973). Upon retrieval, a lid is lowered by rope to cover the tray and the sampler is lifted from the water. Beak trays can be effective in collecting macroinvertebrates from unstable or sandy substrates, but they have been shown to collect fewer taxa and individuals than multiplate and rock basket samplers (Slack et al. 1986). Flannagan and Rosenberg (1982) described several other types of samplers of various size, shape, and composition that have been placed on the substrate or suspended in the water column for sampling benthic macroinvertebrates. These include mesh bags, boards, tiles, bricks, plastic sheets or ropes (vegetation mimics), and buried pots, baskets or trays filled with organic or inorganic materials. However, many of these devices are inadequate due to the depth, elevated turbidity, and high flows of many large rivers. Drift nets are another passive method that can be used to sample large river macroinvertebrates if flow is adequate (Lazorchak et al. 2000), but

studies have shown drift net data are highly variable compared to other methods if not deployed properly (Blocksom and Flotemersch 2005). Poor performance of drift nets can be attributed to low velocities, length of deployment periods, and deployment season. Macroinvertebrate drift densities peak at night (Resh and Rosenberg 1984), so evening deployment of drift nets would be required to maximize their effectiveness.

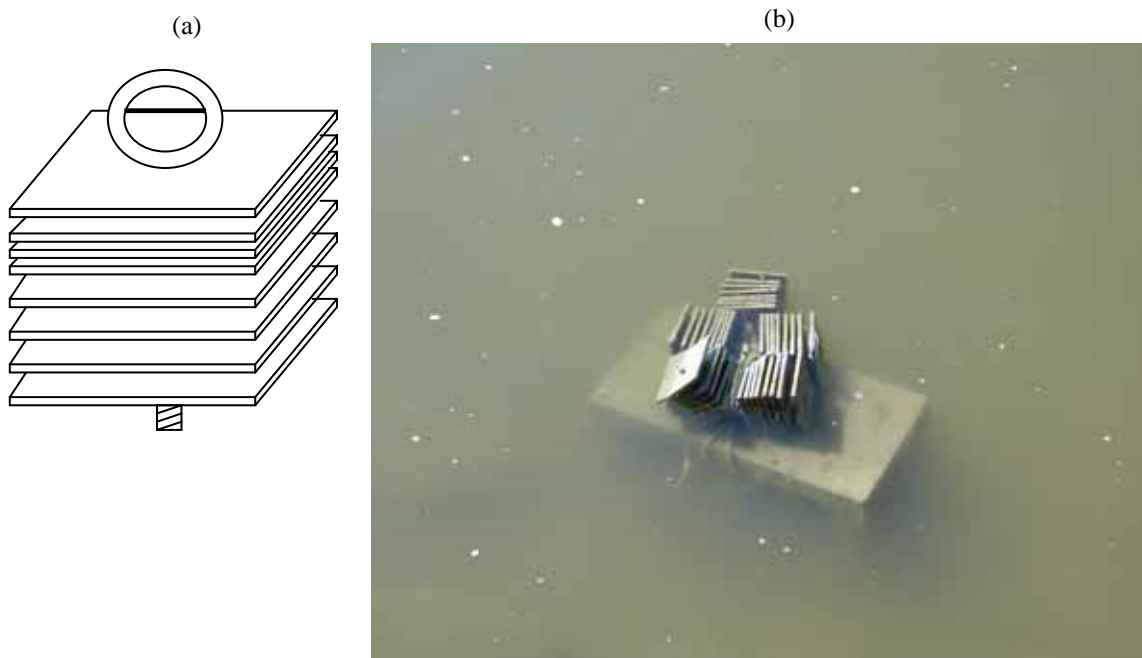


FIGURE 6-2. a) Modified Hester-Dendy multiplate artificial substrate sampler; b) Exposed Hester-Dendy sampler attached to cinder block anchor.

6.2.2 Active Methods

Active methods for sampling macroinvertebrates include a wide variety of sampling approaches that can be grouped into two categories: deep water and shallow water. Active methods are quantitative, semi-quantitative or qualitative and can be used alone or in combination. All active methods have the advantage of only requiring one trip to the sample site, thereby reducing travel cost and effort over passive methods. In addition, these methods focus on measuring or characterizing the existing macroinvertebrate assemblage at a site rather than colonization potential. Disadvantages include a generally high degree of sample variability and high sample debris accumulation that increases sample processing time.

6.2.2.1 Deep Water: Main Channel Sampling

Deep habitats of large rivers can be sampled from a boat using various dredge or bottom grab sampling devices described by Klemm et al. (1990) (e.g., Peterson, Ponar, Ekman, van Veen samples). These samplers are specifically designed for sampling less-stable substrates (e.g.,

sand, silt) usually found in depositional areas. Grab samplers are lowered to the bottom and penetrate the sediments under their own weight. Jaws of the samplers are forced shut by weights, levers, springs or cables to retrieve samples from a known surface area. Although these samplers are most commonly used in deep water, some can be adapted to shallow waters by rigging samplers on poles or by physically pushing samplers into the substrate. Bottom-grab samplers are available in several different designs, each with their own subtle advantages and disadvantages for specific habitats or substrate types (see Klemm et al. 1990 for a review) (Table 6-3).

TABLE 6-3. Advantages and disadvantages of bottom grab samplers.

ADVANTAGES/DISADVANTAGES
<p><i>Advantages:</i></p> <ol style="list-style-type: none"> 1) Requires only one site visit for sample collection, thus reducing overall cost and effort. 2) Results in a sample of the macroinvertebrate assemblage at the site. 3) Effective in sampling deepwater habitats not reachable by most conventional methods. 4) Effective for sampling organisms that burrow in soft sediments and are often the most abundant in large rivers (e.g., oligochaetes and burrowing mayflies). 5) Requires little training and can collect standardized, quantitative benthic samples. <p><i>Disadvantages:</i></p> <ol style="list-style-type: none"> 1) Usually operated “blind,” due to elevated turbidity common on large rivers, with little or no knowledge of specific substrate type that is being sampled (i.e., silt, sand or gravel). 2) Ineffective at sampling rocky or hard substrates. 3) Organisms often lost in “washout” as devices are lifted onto the boat and removed from water. 4) “Jaws” of many samplers can be easily blocked by debris. 5) Some dredges are heavy and cumbersome, occasionally requiring a mechanical winch. 6) Using these methods, reducing sampling variability by stratification is difficult due to the patchy distribution of organisms in sand and silt substrates. 7) Proper operation of many dredge samplers prevents them from being used in habitats with significant flow rates.

Deep waters of large river main channels can also be sampled by SCUBA divers. A diver-operated dome sampler contains a battery-operated pump that moves materials dislodged by a diver into a Nitex mesh sample bag (Gale and Thompson 1975). This quantitative method can be used to successfully sample a variety of deepwater habitats, including coarse substrates. Divers can also operate other devices for sampling benthos, including suction samplers, grab samplers, and corers; and can be used for placement and retrieval of artificial substrates (Gale and Thompson 1974, Klemm et al. 1990). A major advantage of using SCUBA divers is that the divers can see the habitats, making proportional or habitat-specific sampling of river bottoms

more feasible. However, cost, logistical and safety constraints usually render this method impractical for widespread and routine application.

Although more frequently applied in lakes (Muli and Mavuti 2001) and oceans, benthic trawls have also been used to sample the macrobenthos of deep large river main channels. Wright et al. (2000) used benthic trawls to survey the macroinvertebrate fauna of the Thames River. Similarly, benthic trawls have been used in estuarine sections of the Lower St. Johns River in Florida (Mason 1998) and in the Columbia River estuary (Jones et al. 1990). For additional information on trawl selectivity and efficiency, consult Stokesbury et al. (1999).

6.2.2.2 Shallow Water: Shoreline Sampling

Approaches for large river shoreline sampling are similar to well-developed methods for wadeable streams (Ohio EPA 1989, Barbour et al. 1999, Klemm et al. 2000, Flotemersch et al. 2001, Moulton et al. 2002, Merritt et al. 2005). They are often used in large rivers to help avoid logistical constraints encountered in deepwater sampling from a boat in the main channel (see Table 6.4 for a description of advantages and disadvantages). These methods often involve wading in shallow near-shore areas of larger rivers. Even though the wadeable shore zone only accounts for a small proportion of the entire river channel, it may be the most productive and diverse zone for benthic macroinvertebrates (Wetzel 2001). The shallows along main-channel margins have the greatest light penetration for benthic algae and aquatic macrophytes. Allochthonous organic matter also accumulates in the shallows as a result of direct riparian inputs and from backeddies and currents that deposit LWD and FPOM along the shore. The shoreline substrates of many large rivers tend to be dominated by LWD and other stable substrates, such as cobbles and boulders. As a result of their relatively high habitat complexity and productivity, large river shorelines are similar to the highly productive littoral zones of lentic ecosystems. This is particularly true of large, deep rivers where flow is heavily regulated.

Most sampling approaches used for wadeable streams can be used in the littoral areas of large rivers. Active sampling methods along the shoreline include a variety of qualitative, semi-quantitative, and quantitative techniques. When sampling larger substrate types that can be easily handled (e.g., rocks, woody debris/snags, macrophytes), macroinvertebrates may be removed by scrubbing the substrate with a soft brush or picking them individually with forceps. Conventional dip net-based methods include kicks, dips, jabs, or sweeps in one or more habitat types. D-frame or rectangular kick nets are commonly used at the wadeable margins and are most effective when flow is adequate to carry dislodged organisms into the net. Surber and Hess samplers (which quantitatively sample fixed areas) can also be used, but require greater flow velocity than do dip net methods. Although kick nets are most commonly used; grab samplers, corers, and suction samplers can also be used to sample fine sediments along the shoreline. Table 6-4 list some general advantages and disadvantages of active shoreline benthic sampling.

6.2.2.3 Snag Sampling

Sampling woody debris or “snags” (usually >10 cm in diameter) is another method that can be used either in the deep waters of the main channel, from a boat, or in shallow shoreline areas.

These substrates are natural and stable and have been recognized as some of the most productive macroinvertebrate habitats of large rivers, particularly in rivers dominated by unstable sandy bottoms (e.g., Benke et al. 1985, Benke 2001, Merritt et al. 2005). Snags are most frequently sampled by placing a dip net on the downstream side and gently scrubbing the snag surface with a soft brush, allowing the current to carry dislodged material into the net. Although a regular dip net is often used, Angradi (2006) describes a specialized “snag net” that resembles a D-frame net except that the frame is constructed so that the net fits over half the circumference of the snag. Snag bags have also been used to collect macroinvertebrates from woody debris (Growth et al. 1999). Snags have an advantage over artificial substrates because, in addition to providing stable habitats, they are natural substrates and the decomposing wood and associated biofilms serve as a food resource for macroinvertebrates. However, irregular size and shape often make it difficult to standardize the area sampled. The length of time the snag has been in the water, or the period of colonization, is also typically unknown. Yet it may be possible to use conditioned snag habitats for preliminary bioassessment, or “bioreconnaissance,” efforts on large rivers. Snag sampling is currently being incorporated into both large river and great river macroinvertebrate sampling protocols of the USEPA (Angradi 2006, Johnson et al. 2004) and the Michigan DEQ (Merritt et al. 2005).

TABLE 6-4. Advantages and disadvantages of shoreline benthic sampling.

ADVANTAGES/DISADVANTAGES
<p><i>Advantages:</i></p> <ol style="list-style-type: none"> 1) Requires only one site visit for sample collection, thus reducing overall cost and effort. 2) Assesses the macroinvertebrate assemblage found in the study reach. 3) Doesn't require a boat, therefore reducing cost and hazards associated with boat operation, if shoreline sample zone is wadeable and easily accessible. 4) Shallow shoreline habitats are often readily observable, making it possible to target specific habitats or to sample habitats proportionately. 5) Dip-net methods can be used to sample a variety of both stable (e.g. rocks, woody debris, macrophytes, cobble) and unstable (e.g., sand, silt, muck) habitats, enhancing sample representativeness. <p><i>Disadvantages:</i></p> <ol style="list-style-type: none"> 1) Samples can be variable due to diversity of habitat types and the patchy distribution of organisms, potentially requiring more replicate samples to reduce this variability. 2) Sorting macroinvertebrates from the debris of shoreline samples increases sample processing time and costs. 3) Difficult or impossible where there are steep drop-offs or sheer cliffs at rivers edge.

6.3 The Large River Bioassessment Protocol (LR-BP) for Benthic Macroinvertebrate Sampling

The LR-BP method is a hybrid of USEPA-EMAP (Lazorchak et al. 2000), USEPA-RBP (Barbour et al. 1999) and USGS-NAWQA (Moulton et al. 2002) sampling methods. The LR-BP uses transect sampling and can be applied in a systematic, unbiased manner for bioassessment. The LR-BP is a combination of semi-quantitative multi-habitat sampling methods applied in a systematic randomized fashion that has been studied for its performance characteristics and variability (Flotemersch et al. 2006) and was designed to be standardized, quantitative and user friendly. It incorporates proportional multi-habitat sampling and, therefore, should accurately reflect site condition. This method was shown to be responsive to a gradient of disturbance and can be used on a variety of large rivers (Flotemersch and Blocksom 2006).

The LR-BP specifies a reach length of 500 m because it: 1) has been shown to provide representative samples (Blocksom and Flotemersch 2006 [submitted]); 2) is manageable for investigators due to the entire reach usually being observable from a single point; and 3) works well for large river fish bioassessment when both banks are electrofished and, thus, provides comparable sampling reaches for both assemblages (1000 m total shoreline) (Flotemersch and Blocksom 2005). The target sample location (e.g., established by GPS coordinates for a probabilistic design) indicates the downstream end of the reach where sampling begins. At each site, there are a total of six transects. Transect A is located at the downstream end of the reach with the remaining five transects at 100 m, 200 m, 300 m, 400 m and 500 m (Figure 6-3). At each transect, a 10-m sample zone (5 m on each side of transect) on each bank defines where macroinvertebrates will be collected. The zone extends from the edge of water to the mid-point of the river or until depth exceeds 1 m (Figure 6-3), but sampling is largely bank-oriented except in shallow rivers. Six sweeps, each 0.5 m in length, are collected within the zone using a D-frame net (500- μ m mesh). Each sweep covers 0.15 m² of substrate (i.e., net width of 0.3 m and a 0.5 m length of pass); therefore, six sweeps will cover an area of 0.9 m². The six sweeps are proportionately allocated based on available habitat within the 10-m sample zone (e.g., snags, macrophytes, cobble). This method negates the need for separate collection nets in the field and helps standardize the area sampled. If water at a site is more than 1 m deep at the waters edge, the six sweeps should be collected from a boat if possible. Each transect has two zones (one on each bank) and samples from the entire reach are composited into a single sample. This results in each sample containing debris and organisms from 12 separate zones (total of ~12 m²) that represent the 500-m reach.

6.4 Field Preservation

In most macroinvertebrate sampling protocols, multiple steps are involved in processing samples in the field. Sample material is composited for the entire site, and then placed into a sieve bucket to drain excess water and allow washing of fine sediments. The number of samples comprising the composite sample will depend on the sampling method used at the site. Large objects (e.g., rocks, woody debris) are inspected, attached invertebrates are picked from them, and the objects are returned to the river. Each piece of substrate is then gently washed or scrubbed to remove attached organisms. Substrate pieces are removed from the bucket or sieve after cleaning.

After sieving, samples are typically transferred to a suitable container and preserved with ethanol (70% final concentration) or a 10% buffered formalin solution. Buffered formalin may be a better preservative for large river benthic samples as they typically contain a greater number of soft-bodied oligochaetes and leeches that are inadequately preserved by alcohol. Many investigators choose to first fix the sample in formalin and later transfer the sample to ethanol prior to laboratory processing (Klemm et al. 1990). In addition to externally labeling the sample container at the site, it is advisable to use an internal label. Additional details on field processing of macroinvertebrate samples are provided by Klemm et al. (1990).

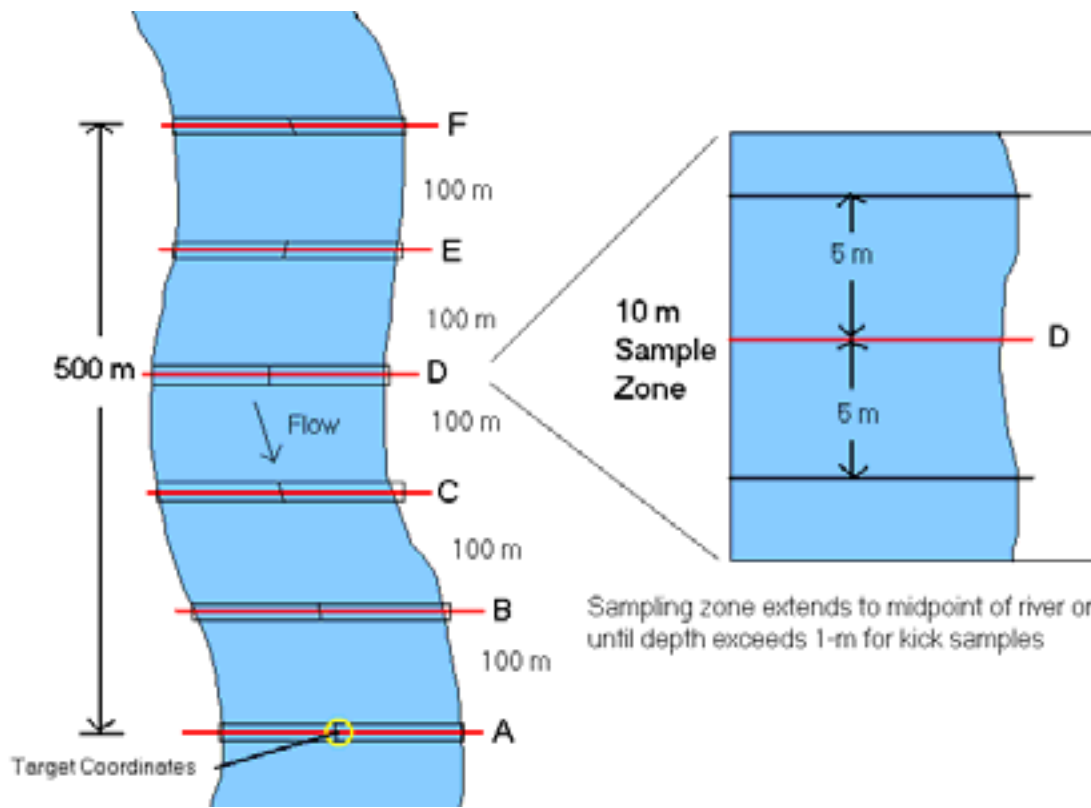


FIGURE 6-3. Example of the six transects and 6 sample zones for collection of benthic macroinvertebrates in large rivers using the LR-BP design.

6.5 Laboratory Processing

There are three components to laboratory processing of benthic macroinvertebrate samples: sorting/subsampling, taxonomic identifications and counts (i.e., enumeration). Several questions should be addressed prior to initiating laboratory processing.

- Will samples be sorted in their entirety, or will they be subsampled?
- If samples are to be subsampled, will the process be based on fixed volume or fixed count?
- If fixed count, what is the target (e.g., 100, 200, 300, 500 organisms)?

- Is there a target taxonomic level (e.g., genus), the lowest practical taxonomic level, or does it vary by group?
- What, if any, rules are there for counting?

6.5.1 *Sorting and Subsampling*

Although it is widely recognized that subsampling helps to manage the level of effort associated with bioassessment laboratory work (Carter and Resh 2001), the practice has been the subject of much debate (Courtemanch 1996, Barbour and Gerritsen 1996, Vinson and Hawkins 1996). If a fixed count method is used, power analyses can determine the most appropriate number of targeted organisms (Ferraro et al. 1989, Barbour and Gerritsen 1996). Fixed organism counts vary greatly among monitoring agencies (Carter and Resh 2001), with 100, 200, 300 and 500 counts being most often used (Plafkin et al. 1989, Barbour et al. 1999, Cao and Hawkins 2005). As part of the LR-BP development process, Flotemersch and Blockson (2005) provided an assessment of the effect subsample size had on metric performance from large river benthic samples. They concluded that a 500-organism count was best, based on examination of the relative increase in richness metric values (< 2%) between successive 100-organism counts. However, a 300-organism count was deemed sufficient for most study needs. Others have recommended higher fixed counts, including a minimum of 600 in wadeable streams (Cao and Hawkins 2005).

If organisms are missed during the sorting process, bias is introduced in the resulting data. Thus, the primary goal of sorting is to completely separate organisms from organic and inorganic material (e.g., detritus, sediment) in the sample. A secondary goal of sorting is to provide the taxonomist with a sample for which the majority of specimens are identifiable. Although it is not the decision of the sorter whether an organism is identifiable, straightforward rules can be applied that minimize specimen loss (Table 6-5). If a sorter is uncertain about whether an organism is countable, the specimen should be placed in the vial and not added to the rough count total.

TABLE 6-5. Example list of counting “rules”: what not to count.

Organisms that should *not* be counted include:

- a) Non-benthic organisms, such as free-swimming gyrid adults or surface-dwelling veliids (Insecta:Heteroptera)
 - b) Empty mollusk shells (Mollusca:Bivalvia)
 - c) Non-headed worm fragments (Oligochaeta)
 - d) Terrestrial insects (incidentals)
 - e) Copepoda
 - f) Exuviae (molted “skins”)
-

The sorting/subsampling process is based on randomly selecting portions of the sample detritus spread over a gridded Caton screen (Caton 1991, Barbour et al. 1999; Figure 6-4a, b). Prior to beginning the sorting/subsampling process, it is important that the sample be mixed thoroughly

and distributed evenly across the sorting tray to reduce the effect of organism clumping that may have occurred in the sample container. The grids are removed from the screen, placed in a sorting tray, and all organisms removed; the process is completed until the rough count by the sorter exceeds the target subsample size. This process should produce at least three containers per sample (all of which should be clearly labeled):

- Subsample to be given to taxonomist,
- Sort residue, to be checked for missed specimens, and
- Unsorted sample remains to be used for additional sorting, if necessary.

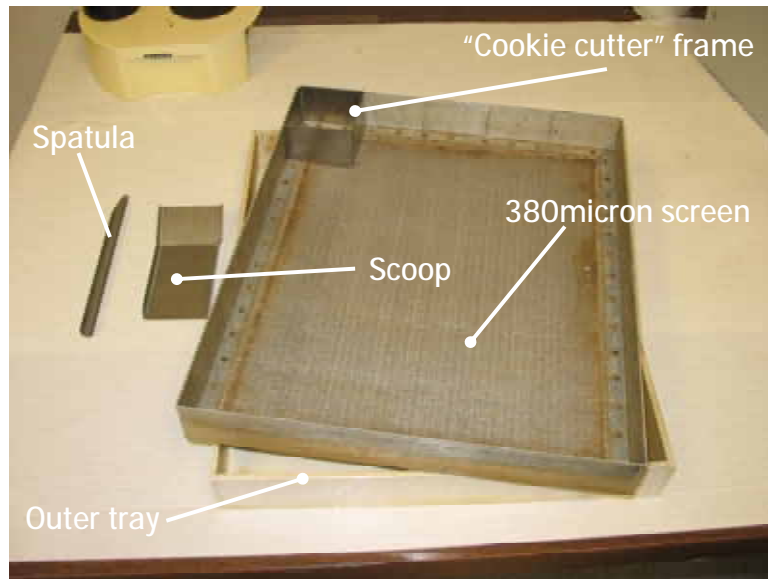


FIGURE 6-4a. Gridded screen (Caton 1991) used to facilitate subsampling.

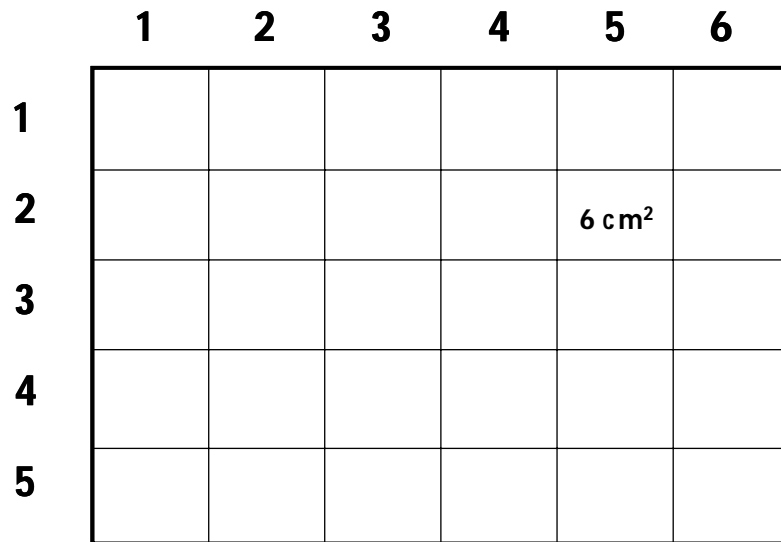


FIGURE 6-4b. Schematic diagram of the Caton gridded subsampling screen, consisting of 30 6-cm² grids.

6.5.2 Taxonomy and Enumeration

The next step of the laboratory process is identifying the organisms within the subsample. A major question associated with taxonomy is the hierarchical target levels required of the taxonomist, including order, family, genus, species or the lowest practical taxonomic level (LPTL). While family level is used effectively in some monitoring programs (Carter and Resh 2001), the taxonomic level primarily used in most routine monitoring programs is genus. However, even with genus as the target, many programs often treat selected groups, such as midges (Chironomidae) and worms (Oligochaeta), differently due to the need for slide-mounting. Slide mounting specimens in these two groups is usually necessary to attain genus level nomenclature, and sometimes even tribal. Because taxonomy is a major potential source of error in monitoring data sets (Stribling et al. 2003), it is critical to define taxonomic expectations and to treat all samples consistently, both by a single taxonomist and among multiple taxonomists. This, in part, requires specifying both hierarchical targets and counting rules.

An example list of taxonomic target levels is shown in Table 6-6. These target levels define the level of effort that should be applied to each specimen. If it is not possible to attain these levels for certain specimens due to, for example, the presence of early instars, damage, or poor slide mounts, the taxonomist provides a more coarse-level identification.

When a taxonomist receives samples for identification, depending upon the rigor of the sorting process (see Section 6.3.1), the samples may contain specimens that either cannot be identified, or should not be included in the sample (Table 6-6). The final screen of sample integrity is the responsibility of the taxonomist, who determines which specimens should remain unrecorded (for any of the reasons stated above). Beyond this, the principal responsibility of the taxonomist is to record and report the taxa in the sample and the number of individuals of each taxon.

Programs should use the most current and accepted keys and nomenclature. *An Introduction to the Aquatic Insects of North America* (Merritt and Cummins 1996) is useful for identifying the majority of aquatic insects in North America to genus level. By their very nature, most taxonomic keys are obsolete soon after publication; however, research taxonomists do not discontinue research once keys are available. Thus, it is often necessary to have access to and be familiar with ongoing research in different taxonomic groups. Other keys are also necessary for non-insect benthic macroinvertebrates that will be encountered, such as Oligochaeta, Mollusca, Acari, Crustacea, Platyhelminthes and others. Klemm et al. (1990) and Merritt and Cummins (1996) provide an exhaustive list of taxonomic literature for all major groups of freshwater benthic macroinvertebrates. Although it is not current for all taxa, the integrated taxonomic information system (ITIS; <http://www.itis.usda.gov/>) has served as a clearinghouse for accepted nomenclature, including validity, authorship and spelling.

TABLE 6-6. Example of taxonomic hierarchical targets used in benthic macroinvertebrate identifications.

TAXON	TARGET
PHYLUM ANNELIDA	
Class Branchiobdellida	Genus
Class Hirudinea	Genus
Class Oligochaeta	Genus
Class Polychaeta	Family
PHYLUM ARTHROPODA	
Class Arachnoidea	
Acari	Genus
Class Insecta	
Coleoptera	Genus
Diptera	<i>Identify all to genus except in the following cases:</i>
Chironomidae	Genus (tribe or subfamily, if specified)
Dolichopodidae	Family
Phoridae	Family
Scathophagidae	Family
Syrphidae	Family
Ephemeroptera	Genus
Heteroptera	Genus
Lepidoptera	Genus
Megaloptera	Genus
Odonata	Genus
Plecoptera	Genus
Trichoptera	Genus
Class Malacostraca	Genus
Amphipoda	Genus
Decapoda	Genus
Isopoda	Genus
Mysidacea	Genus
Class Ostracoda	Genus
PHYLUM COELENTERATA	
PHYLUM MOLLUSCA	
Class Bivalvia	Genus
Class Gastropoda	<i>Identify all to genus except in the following cases:</i>
Family Hydrobiidae	Family
PHYLUM NEMERTEA	
	Genus

6.6 Data Entry

Taxonomic nomenclature and counts are usually entered into the data management system directly from handwritten bench or field sheets. Depending upon the system used, there may be an autocomplete function that helps prevent misspellings. There are two methods for assuring

accuracy in data entry. One is the double entry of all data by two separate individuals, and then performing a direct match between databases. Where there are differences, it is determined which database is in error, and corrections are made. The second approach is to perform a 100% comparison of all data entered to handwritten data sheets. Comparisons should be performed by someone other than the primary data enterer. When errors are found, they are hand-edited for documentation, and corrections are made electronically. The rates of data entry errors are recorded and segregated by data type (e.g., fish, benthic macroinvertebrates, periphyton, header information, latitude and longitude, physical habitat, and water chemistry).

6.7 Data Reduction (Metric Calculation)

This section focuses on activities that convert raw data (taxa lists and counts) into numeric terms (metrics) to be used for subsequent analyses, (e.g., metric calculation). For example, Blocksom and Flotemersch (2005) tested 42 metrics relative to different sampling methods, mesh sizes, and habitat types (Table 6-7). Twenty-seven of the 41 metrics (66%) are taxonomically based. Those remaining require tolerance value and functional feeding group designations to calculate the metrics.

To ensure that database queries are correct and result in the intended metric values, a subset of values should be recalculated by hand. One metric is calculated for all samples, all metrics are calculated for one sample. When recalculated values differ from those values in the matrix, the reasons for the disagreement are determined and corrections are made. Reports on performance include the total number of reduced values as a percentage of the total, how many errors were found in the queries, and the corrective actions specifically documented.

6.8 Final Index and Site Assessment

Approximately 56 state or tribal agencies currently use macroinvertebrates in biomonitoring or bioassessment programs in the USA (USEPA 2002). Of these, more than 40 have developed an index of some type (multimetric or multivariate predictive) for use in site assessment. These indices are developed using reference sites. The final assessment for a site is usually determined based on a site score relative to the distribution of reference site scores. Approaches for scoring the reference distribution vary and depending on several factors (Barbour et al. 1999). The process for developing these indices is described in detail in Chapter 8.

TABLE 6-7. Benthic macroinvertebrate metrics evaluated by Blocksom and Flotemersch (2005) for responsiveness to measured disturbance gradients in large rivers.

Metric (by category)	Metric Description
Richness and diversity	
Number of taxa	The count of unique taxa in the sample. A standard level of identification (family, genus, species) must be defined for each taxonomic group
Number of Ephemeroptera, Plecoptera, Trichoptera (EPT) taxa	Number of taxa in the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies)
Number of Ephemeroptera taxa	Number of mayfly taxa
Number of Plecoptera taxa	Number of stonefly taxa
Number of Trichoptera taxa	Number of caddisfly taxa
Number of Ephemeroptera, Trichoptera, and Odonata (ETO) taxa	Number of taxa in the insect orders Ephemeroptera (mayflies), Trichoptera (caddisflies), and Odonata (dragonflies and damselflies)
Number of Odonata taxa	Number of dragonfly and damselfly taxa
Number of Chironomidae taxa	Number of midge taxa
Number of Hemiptera taxa	Number of “true” bug taxa
Number of Coleoptera taxa	Number of beetle taxa
Number of Mollusca + Crustacea taxa	Number of mollusk (snails and clams) and crustacean (e.g., amphipods, copepods, decapods) taxa
Shannon diversity	An index of richness and composition calculated as: $\Sigma -((n/N) * \text{Log}(n/N)) / \text{Log}(2)$; where n is the number of individuals in a taxon and N is the number of individuals in the sample, summed for all taxa in the sample. The index is commonly standardized on log of 2 (as shown here) or the natural log (log e)
Composition and evenness	
Non-insects (%)	Non-insect individuals in the sample as a percentage of all individuals
Oligochaetes and leeches (%)	Percentage of worm and leech individuals
EPT individuals (%)	Percentage of mayfly, stonefly, and caddisfly individuals
Taxa in EPT (%)	Mayfly, stonefly, and caddisfly taxa in the sample as a percentage of all taxa
Ephemeroptera individuals (%)	Percentage of mayfly individuals
Plecoptera individuals (%)	Percentage of stonefly individuals
Trichoptera individuals (%)	Percentage of caddisfly individuals
Chironomidae individuals (%)	Percentage of midge individuals
Taxa in Chironomidae (%)	Percentage of midge taxa
Hemiptera individuals (%)	Percentage of “true” bug individuals
Odonata individuals (%)	Percentage of dragonfly and damselfly individuals
Coleoptera individuals (%)	Percentage of beetle individuals
Elmidae individuals (%)	Percentage of riffle beetle individuals
Number of individuals per taxon	The average number of individuals per unique taxon
Dominant taxon (%)	Individuals in the most numerous unique taxon as a percentage of all individuals
Dominant five taxa (%)	Individuals in the five most numerous unique taxa as a percentage of all individuals

TABLE 6-7. Continued.

Metric (by category)	Metric Description
Pollution tolerance	In all of the pollution tolerance metrics, degrees of pollution tolerance must be defined per taxon. This may be done categorically (e.g., sensitive, facultative, tolerant) or on a more continuous scale, as in the Hilsenhoff scale from 0 to 10. In addition, the pollution to which the organisms are responding may be general habitat and water quality stresses or specific (e.g., metals, sediments).
Number of intolerant taxa	Count of unique taxa that are sensitive to stresses (e.g., Hilsenhoff values 0 – 3)
Taxa as intolerant (%)	Sensitive taxa in the sample as a percentage of all taxa
Intolerant individuals (%)	Sensitive individuals in the sample as a percentage of all individuals
Number of tolerant taxa	Count of unique taxa that are tolerant of stresses (e.g., Hilsenhoff values 7 – 10)
Taxa as tolerant (%)	Tolerant taxa in the sample as a percentage of all taxa
Tolerant individuals (%)	Tolerant individuals in the sample as a percentage of all individuals
Hilsenhoff Biotic Index	The average individual pollution tolerance value for the sample. Calculated as: $HBI = \sum (n) * (\text{tolerance value}) / N$; where n is the number of individuals in a taxon and N is the number of individuals in the sample that have known tolerance values; summed for all taxa in the sample. Modifications of the published index (Hilsenhoff 1987) may include assignment of tolerance values to previously unrated organisms or of groups of organisms at genus, family, or order taxonomic levels.
Functional feeding groups	
Number of collector-filterer taxa	Number of unique taxa that feed on particles filtered from the water column
Collector-filterer individuals (%)	Filtering individuals in the sample as a percentage of all individuals
Number of collector-gatherer taxa	Number of unique taxa that feed on particles encountered among the substrates and detritus
Collector-gatherer individuals (%)	Gathering individuals in the sample as a percentage of all individuals
Number of predator taxa	Number of unique taxa that feed on living animal organisms
Predator individuals (%)	Predatory individuals in the sample as a percentage of all individuals
Number of scraper taxa	Number of unique taxa that feed on algae and bacteria that are attached to the surfaces of hard substrates
Scraper individuals (%)	Scraping individuals in the sample as a percentage of all individuals

6.9 Performance Characteristics for Biological Assessments Using Benthic Macroinvertebrates

6.9.1 Field Sampling

Quantitative (QN) performance characteristics for field sampling are *precision* and *completeness* (Table 6-8). Repeat samples for purposes of calculating precision of field sampling are obtained by sampling two adjacent reaches, shown as 500 m in this example (Figure 6-5) and for which there are not dramatic differences in condition. This can be done by the same field team for intra-team precision, or by different teams for inter-team precision. For benthic macroinvertebrates, samples from the adjacent reaches (also called quality control [QC] or duplicate samples) must be laboratory-processed prior to data being available for precision calculations. Assuming acceptable laboratory error, these precision values are statements of the consistency with which the sampling protocols 1) characterized the biology of the river and 2) were applied by the field team, and thus, reflect a combination of natural variability and systematic error (see Chapter 3).

TABLE 6-8. Error partitioning framework for biological assessments and biological assessment protocols for benthic macroinvertebrates. There may be additional activities and performance characteristics, and they may be quantitative (QN), qualitative (QL) or not applicable (na).

Component Method or Activity	Performance Characteristics				
	Precision	Accuracy	Bias	Representativeness	Completeness
1. Field sampling	QN	na	QL	QL	QN
2. Laboratory sorting/subsampling	QN	na	QN	QL	QN
3. Taxonomy	QN	QL	QL	na	QN
4. Data entry	na	QN	na	na	QN
5. Data reduction (e. g., metric calculation)	na	QN	QN	na	na
6. Site assessment and interpretation	QN	QN	QL	QL	QN

The number of reaches for which repeat samples are taken varies, but a rule-of-thumb is 10% randomly selected from the total number of sampling reaches constituting a sampling effort (whether yearly, programmatic routine, or individual project). Metric and index values are used to calculate relative percent difference (RPD), root-mean square error (RMSE), and coefficient of variability (CV) (Table 3-2). Acceptance criteria for each of these would be established based on programmatic capabilities demonstrated via pilot studies, or through analysis of existing datasets produced using the same protocols. These criteria are not data quality thresholds beyond which data points should be considered for discarding. Rather, they are flags for potential

problems (errors) in sample collection or processing. They are used to help determine the source(s) of the problems and to help develop recommendations for corrective actions. (K. Blocksom U.S. Environmental Protection Agency, personal communication) characterized performance measures for the benthic macroinvertebrate LR-BP (Table 6-9) (field sampling precision and metric sensitivity) when sample reaches are categorized according to mean thalweg depth.

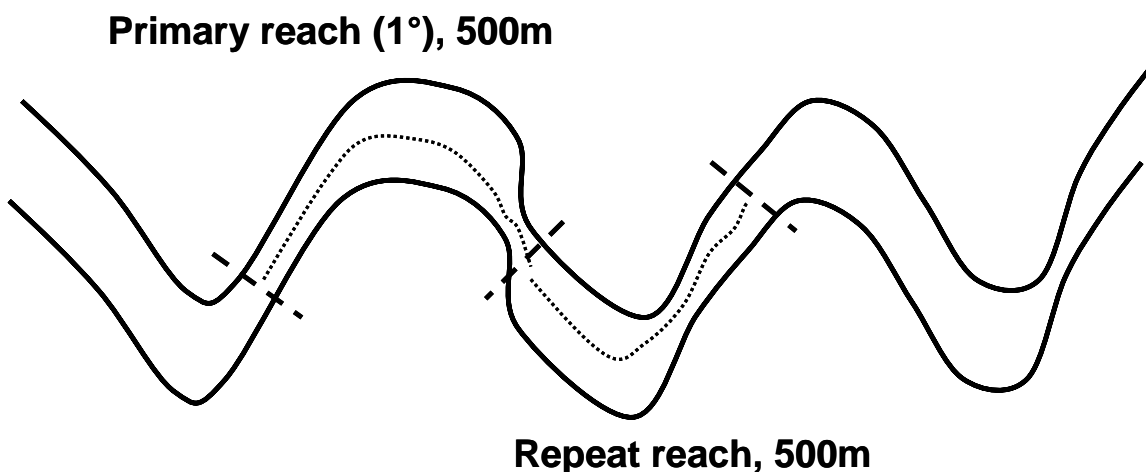


FIGURE 6-5. Adjacent reaches (primary and repeat) on a river channel.

TABLE 6-9. Precision and sensitivity of field sampling using the LR-BP for benthic macroinvertebrates (K. Blocksom, US Environmental Protection Agency, personal communication).

Metric	Mean*		Field Variance		Field CV (%)		Variance (field+lab)		DD (field+lab)†	
	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow
Total Taxa	43.7	56.4	17.3	6.4	9.5	4.5	56.4	51.7	14.7	14.1
EPOT Taxa	7.6	16.6	1.1	0.1	13.6	2.0	0.2	0.2	0.9	0.8
% Tolerant Individ.	50.7	32.5	10.4	25.2	6.4	15.4	47.9	80.6	13.6	17.6
% Chironomidae	49.0	33.0	73.6	25.7	17.5	15.4	158.2	88.1	24.6	18.4
% Dominant Taxon	34.0	19.8	62.3	18.5	23.2	21.8	137.0	72.7	22.9	16.7

*“Deep” and “Shallow” refer to different depth categories of sampling reaches

†Based on $\alpha=0.05$; $n=1$

Percent completeness (Tables 3-2, 6-8) is calculated to communicate the number of valid samples collected as a proportion of those that were originally planned. This value serves as one summary of overall data quality for a sampling effort and it demonstrates confidence in the final results.

Qualitative (QL) performance characteristics for field sampling are *bias* and *representativeness* (Table 6-8). Attempts to minimize the bias associated with the LR-BP for benthic macroinvertebrates include two components of the field method. First, it is not limited to one or a few habitat types (it is multihabitat and samples stable undercut banks, macrophyte beds, root wads/snags, gravel/sand/cobble). Second, allocation of the sampling effort is distributed throughout the entire 500-m sampling reach by use of six evenly-spaced transects, preventing the entire sample from being taken in a shortened portion of the reach. The LR-BP field sampling method is intended to depict the benthic macroinvertebrate assemblage physical habitat in the large river shore-zone (out to a depth of 1m).

Accuracy is considered “not applicable” to field sampling (Table 6-8), because efforts to define analytical truth would necessitate a sampling effort excessive beyond any practicality. That is, the analytical truth would be all benthic macroinvertebrates that exist in the river (shore zone to 1-m depth). There is no sampling approach that will collect all individual benthic macroinvertebrate organisms.

6.9.2 Laboratory Sorting/Subsampling

Precision, *bias*, and, in part, *completeness* are QN characteristics of performance for laboratory sorting and subsampling (Table 6-8). Precision of laboratory sorting is calculated by use of RPD with metrics and indices as the input variables (Table 3-2). If, for example, the targeted subsample size is 300 organisms, and that size subsample is drawn twice from a sorting tray without re-mixing or re-spreading, metrics can be calculated from the two separate subsamples. RPD would be an indication of how well the sample was mixed and spread in the tray; the “serial subsampling” and RPD calculations should be done on two timeframes. First, these calculations should be done, and the results documented and reported to demonstrate what the laboratory (or individual sorter) is capable of in application of the subsampling method. Second, they should be done periodically to demonstrate that the program routinely continues to meet that level of precision. Bias of the sorting process is evaluated by checking for specimens that may have been overlooked or otherwise missed by the primary sorter; checking of sort residue is performed by an independent sort checker. The number of specimens found by the checker as a proportion of the total number of originally found specimens is the percent sorting efficiency (PSE) (Table 3-2), and quantifies sorting bias. This exercise is performed on a randomly-selected subset of sort residues (generally 10% of total sample lot), the selection of which is stratified by individual sorters, by projects, or by programs. As a rule-of-thumb, an MQO could be “less than 10% of all samples checked will have a PSE \leq 90%”. Representativeness of the sorting/subsampling process is addressed as part of the standard operating procedure (SOP) that requires random selection of grid squares (Figure 6-4) with complete sorting, until the target number is reached within the final grid. Percent completeness for subsampling is calculated as the proportion of samples with the target subsample size (\pm 20%) in the rough sort. Considered as “not applicable”, estimates of *accuracy* are not necessary for characterizing sorting performance.

6.9.3 Taxonomy

Precision and *completeness* are QN performance characteristics that are used for taxonomy (Table 6-8). Precision of taxonomic identifications is calculated using percent taxonomic disagreement (PTD) and percent difference in enumeration (PDE) (Table 3-2), both of which rely on the raw data (list of taxa and number of individuals) from whole-sample re-identifications. The primary taxonomy is completed by the project taxonomist (T1); the re-identifications are performed by a secondary, or QC taxonomist (T2) as blind samples. The number of identifications in agreement between the two sets of results, as an inverse proportion of the total number of individuals, is precision of the taxonomic identifications, or “percent taxonomic disagreement (PTD)”. The percent difference in sample counts by each of the taxonomists (not the sorters) is “percent difference in enumeration (PDE)”. These two values are evaluated individually, and can be used to indicate the overall quality of the taxonomic data. They can also be used to help identify the source of a problem. The number of samples for which this analysis is performed will vary, but 10% of the total sample lot (project, program, year, or other) is an acceptable rule-of-thumb. Exceptions are that large programs (>~500 samples) may not need to do >50 samples; small programs (<~30 samples) will likely still need to do at least 3 samples. In actuality, the number of re-identified samples be program-specific and will be influenced by multiple factors, such as, how many taxonomists are doing the primary identification (there may be an interest in having 10% of the samples from each taxonomist re-identified), and how confident the ultimate data user is with the results. Mean PTD and PDE across all re-identified samples are estimates of taxonomic precision (consistency) for a dataset or a program. Percent taxonomic completeness (PTC; [Table 3-2]) quantifies the proportion of individuals in a sample that are identified to the specified target taxonomic level (lowest practical taxonomic level, species, genus, family, or other, including mixed levels). Results can be interpreted in a number of ways: the individuals in a sample are damaged or early instar, many are damaged with diagnostic characters missing (such as, gills, legs, antennae, etc.) or the taxonomist is inexperienced or unfamiliar with the particular taxon.

Accuracy and *bias* are QL performance characteristics for taxonomy (Table 6-8). Accuracy requires specification of an analytical truth. For taxonomy, it is 1) the museum-based type specimen (holotype, or other form of type specimen), 2) specimen(s) verified by recognized expert(s) in that particular taxon or 3) unique morphological characteristics specified in dichotomous identification keys. Determination of accuracy is considered “not applicable” for production taxonomy (most often used in routine monitoring programs) because that kind of taxonomy is focused on characterizing the sample; taxonomic accuracy, by definition, would be focused on individual specimens. Bias in taxonomy results from use of obsolete nomenclature and keys, imperfect understanding of morphological characteristics, inadequate optical equipment, and poor training. Neither of these performance characteristics is considered necessary for production taxonomy, in that they are largely covered by the estimates of precision and completeness. For example, although it is possible that two taxonomists would put an incorrect name on an organism, it is considered low probability that they would put the same incorrect name on that organism.

6.9.4 Data Entry

Efforts to understand the quality of data entry activity may seem trivial. However, the impact of errors can be substantial, and, if undiscovered and uncorrected, can become amplified through the assessment process. This QN performance characteristic quantifies the number of correctly-entered data values as a proportion of the total number of data values entered. The process involves having a QC person, distinct from the staff doing the primary data entry, check all data values (100%) against the original handwritten datasheets. With the datasheets as the analytical truth, the rate of errors is the *accuracy* of the data entry (Table 6-8). As errors are found, they are corrected electronically and the corrected value recorded. For their Wadeable Streams program, Mississippi DEQ found that the two data types with the highest error rates were the datasheet header information (e.g., stream name, latitude/longitude, date of site visit, names of field staff) and streambed particle size counts (Mississippi DEQ 2003). This allowed corrective actions to be focused where needed. All other performance characteristics are considered not applicable.

6.9.5 Data Reduction (Metric Calculation)

For most biological assessment programs, raw data are the list of taxa found at a site (in a sample) and the number of individuals recorded for each taxon. Preparation of those data for analysis requires conversion to metrics or other terms; metric calculation is a form of data reduction. When electronic spreadsheets or other data manipulation techniques are used, queries are often built to perform both complex and simple calculations. If queries are not performing as intended, or links to the raw data are incorrect, errors in metric values can occur. *Accuracy* of data reduction is a QN performance characteristic (Table 6-8) that helps ensure database/computer calculation routines are performing as intended. A subset of metric values is hand-calculated using only the taxonomic and enumeration data, which are then compared to those that result from the computer queries. A recommended approach involves calculating one metric for multiple samples (e.g., systematic, every third sample), as well as all metrics for at least one sample. If differences are found, each value should be checked for errors in the calculation process (hand calculator vs computer algorithm), and corrections made.

6.9.6 Site Assessment and Interpretation

QN performance characteristics for site assessment and interpretation are *precision*, *accuracy*, and *completeness* (Table 6-8). Site assessment precision is based on the narrative assessments from the associated index scores (good, fair, poor) from reach duplicates and quantifies the percentage of duplicate samples that are receiving the same narrative assessments. These comparisons are done for a randomly-selected 10% of the total sample lot. Table 6-10 shows that, for this dataset, 79% of the replicates returned assessments of the same category (23 out of 29); 17% were 1 category different (5 of 29); and 3% were 2 categories different (1 of 29). Accuracy is the proportion of samples for which the biological index correctly identifies sites as impaired; the calculation is discrimination efficiency (DE) (Table 3-2). DE is a value that is developed during the index development and calibration process. Percent completeness (%C) is the proportion of sites (of the total planned) for which valid final assessments were obtained.

QL performance characteristics for site assessment and interpretation are bias and representativeness (Table 6-8). The final assessment of a site can be biased if a small number of reference or stressor sites are used during the calibration process. Low numbers of stressor sites can potentially result in high discrimination efficiencies that are spurious. If interpretation of assessment results fails to take into consideration abnormal or extreme hydrologic or climatic events, or other non-natural catastrophic and localized events, results could be considered non-representative of ambient conditions.

TABLE 6-10. Assessment results shown for sample pairs taken from 29 sites, each pair representing two adjacent reaches (back to back). Assessment categories are 1-good, 2-fair, 3-poor and 4-very poor.

Site	Replicate 1		Replicate 2		Categorical Difference
	Narrative	Assessment Category	Narrative	Assessment Category	
A	Poor	3	Poor	3	0
B	Poor	3	Poor	3	0
C	Good	1	Good	1	0
D	Poor	3	Very Poor	4	1
E	Fair	2	Fair	2	0
F	Poor	3	Fair	2	1
G	Poor	3	Poor	3	0
H	Very Poor	4	Very Poor	4	0
I	Very Poor	4	Very Poor	4	0
J	Poor	3	Poor	3	0
K	Poor	3	Poor	3	0
L	Very Poor	4	Very Poor	4	0
M	Very Poor	4	Very Poor	4	0
N	Poor	3	Fair	2	1
O	Poor	3	Poor	3	0
P	Poor	3	Poor	3	0
Q	Poor	3	Very Poor	4	1
R	Poor	3	Poor	3	0
S	Fair	2	Very Poor	4	2
T	Fair	2	Fair	2	0
U	Good	1	Good	1	0
V	Poor	3	Fair	2	1
W	Fair	2	Fair	2	0
X	Poor	3	Poor	3	0
Y	Poor	3	Poor	3	0
Z	Very Poor	4	Very Poor	4	0
AA	Poor	3	Poor	3	0
BB	Fair	2	Fair	2	0
CC	Poor	1	Poor	1	0