Comments on the Draft Assessment and Listing Methodology for Integrated Water Quality Monitoring and Assessment Reporting

Contribution No. 2013009

Submitted By:
Stroud Water Research Center
970 Spencer Road
Avondale, PA 19311

November 25, 2013

Below please find a series of comments, questions, and observations on the various sections of the draft Assessment and Listing Methodology open for comment from the Pennsylvania Department of Environmental Protection (draft dated February 2013). While our comments are numerous, please know that we have no major objections to the intent or content in these chapters. PADEP has made a significant effort to document the monitoring and decision methods, and the revisions and additions provided make these methods clearer now than they every have been. That said, there appear to be opportunities to provide clearer documentation of the core content that is behind these methods (and presumably in PADEP internal reports). In addition, we think there is a need to make sure the methodological details are sufficient to support an effort by the public (or private organizations such as consulting companies) to replicate the methods described in support of the general mission to understand and protect the waters of the Commonwealth of Pennsylvania through monitoring. This has the potential to increase the confidence that PADEP may have in monitoring data received from outside individuals and organizations, which in turn can make those data a positive contribution to the assessment and listing efforts of PADEP.

Assessment and Listing Methodology…

Pg 2, Para. 1 – "variability of the indicator" – this seems like a good spot to address accuracy versus precision, and how you try to balance that issue.

Pg 2, Para. 2 – this is the 1st introduction of the "99% rule". We have a feeling the public is going to become more confused about this issue, when it applies, how the choices for violation frequency are determined, and the science behind it. It shows up in some of the other sections more extensively, and is becoming more of an issue as continuous data are collected.

Pg. 2, Para. 3 – We would replace diurnal with diel throughout all of these documents. Diel refers to only 24-hour periods or patterns while diurnal can be that or refer to just the daytime (versus nighttime or nocturnal).

Pg 3-4, items i to x – There is a lot of information being discussed here that the public rarely sees. Is there a way to insure that all of this information is conveyed to the public? It would build a stronger connection between data and decisions.

Pg 5, Para. 2 – it is important that the old SSWAP protocol is not lost as these data are almost all we have from the late 1990s and early 2000s. While the old SSWAP protocol is not included as an active protocol, it should be included as metadata somewhere so that 25 years from now we know how those data points (which will surely be all electronic) were generated. We did not see any reference to a method or protocol archive, which becomes increasingly important as protocols evolve.

Pg 5, Para. 6 – while the ICE protocol is not being changed here, please note the reference to a "new probabilistic monitoring approach" to be developed in 2013 and future years. No other information is given.

Pg. 6, Para. 1 – Unless we are missing something, identification levels are frequently not 100% to genus in a sample. We see many streams that have at least 40% chironomid midges and oligochaete worms, which we believe are left at family or higher in the current protocol.

Pg. 7, Para. 2 – There are now two macroinvertebrate protocols – one for Nov to May and another for Jun to Sep. Jun to Sep is not preferred, but an option for some sites.

Pg. 9, Para. 1 – We were not aware that nitrate/nitrite was the principal screening tool for potable water. It is not clear how that relates to pathogens.

Riffle/Run Freestone Streams

This protocol outlines the macroinvertebrate sampling protocol for freestone streams and wadeable rivers.

Pg. 1, Para. 3. We do not see any reference to visual magnification to assist in sorting and identifications. Standardized assistance with magnification reduces differences between individuals and seasons. Otherwise, labs that process with magnification will produce different data from those that do not. If magnification is not allowed, the methods intentionally miss smaller, difficult to find and identify individuals, and that needs to be clarified. That does not mean the method without magnification generates bad data, but it does affect the comparability of data among individuals (good vs poor eyesight) and among projects.

Pg. 10. It would be nice to differentiate among CWF, TSF, and WWF when defining impairment. Many CWF are very good streams, and to allow the IBI to degrade to a 50 before receiving attention seems to not be very protective. HQ and EV streams also have a 10-point decline as part of their trend analysis and antidegradation evaluation (Pg. 15). A similar consideration might be appropriate for CWF.

Pg. 10 and 11 – we would like to see some documentation on how stream orders were characterized given the differences in stream resolution depending on the map scale and source. If you count permanent streams on the ground in a watershed you know, a 25 square mile watershed would be far larger than 3rd order in Pennsylvania. However, it is a reasonable outcome from lower resolution maps.

Pg. 11. EPA has been working for some time to bring very small streams and springbrooks (including those that may dry in places or for part of the year) into the stream monitoring discussion. These are very important and common aquatic habitats across all landscapes (rural and urban). Has PADEP begun to collect parallel information to be ready to address that issue when it comes to Pennsylvania?

Pg. 12, Para. 2 – We were not comfortable with the idea that seasonal impairment is acceptable, especially if it repeats itself. Obviously ecological function in the example is lost for part of the year, and with it the ability to support the designated use. If there is a reason to believe the low IBI was accurate for one period of the year, then the observation is meaningful and should not be ignored.

Pg. 14, Para. 1 – We suggest caution around the exceptions listed here. It requires detailed knowledge of the ecology of a site and the specifics of the collection conditions. Those insights are not always available (e.g., species are missing due to emergence). We agree that Nov to May sampling should be encouraged to avoid some of these complications.

Pg. 14, Para 4 – is it possible to establish a reference condition that reflects the gradient of limestone influence? It is probably a challenge because of the limited number of reference sites, but it would be a more accurate representation of the unimpaired condition.

Pg. 16 – This table needs a legend that describes the contents.

Pg. 18, Para. 3 – rather than resample, could you not just remove *Prosimulium*, or XX% of the *Prosimulium* if you believe it is impacting the IBI. Why throw out the data when the rest of the sample is probably good. The composite should have enough material to get 200 individuals without *Prosimulium*.

Continuous Instream Monitoring

We did not have a chance to review this in detail.

Periphyton

This protocol outlines the field collection procedures for:

- Streambed chlorophyll, cellular carbon and nitrogen concentrations, and cellular phosphorus concentrations
- Visual estimates of periphyton cover (and types) across the streambed (utilizing a method developed in New Zealand)
- And various physical and chemical measurements and observations
    - Conductivity/pH/Temperature/Dissolved Oxygen with a meter (should reference the water chemistry protocol that includes GLPs for operating meters)
    - Measuring discharge and noting hydrologic conditions and recent history -
    - Measuring stream size
    - Observing substrate type %
    - Measuring canopy closure/density

In general the protocol is well written, intuitive, and capitalizes on well-established protocols.

- We would like to see the details about the required training and documentation of training that PADEP employees are expected to have. DEP should (and probably does) maintain documents on "certified" field staff for any given method
- The protocol does not reference any other protocol and should, especially for the water chemistry meters (water chemistry protocol) and the physical channel measurements (habitat assessment protocol)
- Would be nice if the canopy method could consider a photographic –based method. As this may be more efficient for field staff (recognizing that image recognition software and training would be necessary part of the office work)
- The observation of substrate type was not described adequately and there was limited reference to appropriate methods to introduce the concept of inorganic/mineral substrate versus organic substrate, let alone introducing the concept of grain size…this should probably reference the physical habitat assessment protocol

Streambed Sediment

This method utilized the EPA rapid assessment protocol – we have no comment or concern with what is presented in this document

Surface Water Collection

Surface Water Collection Protocol –

- The protocol is written in a general manner to describe the collection process, strategies for sampling under various hydrological conditions, risks of (and methods to avoid) contamination, and a QA/QC protocol to determine variability. The methods for quantification of specific chemical analytes are described in other protocols that must be specific to the analysis. However, typical choices of type of sample container, types of filters, volumes required, and variation in holding times and sample treatment/preservation are adequately described. At this point in time, we have only raise three questions or points to address:
    - Is there a procedure to train employees/staff to be "certified" to collect water chemistry samples and to be assigned a collector ID#? Seems that standardization among DEP staff should be the highest priority for all methods presented herein. I imagine there is probably some documentation of staff training and some certification for any given employee for each of these protocols?
    - The document does not address requirements for determining when water levels are appropriate for sampling "baseflow" conditions. There are discussion points regarding storm sampling and a strategy for collecting enough samples throughout a hydrological year for annual loading estimates (including stating an "unbiased" sampling of all hydrological conditions), but no specifics about what constitutes baseflow conditions (e.g., 10 days after any rain event that resulted in a storm hydrograph? Or any other discharge statistic?).
    - For submission of chemical samples to any DEP lab, does the lab have authority/ permission/ protocol to reject any sample (or sample sets) submitted that does not have the FULL metadata requested in the form?

Semi-Quantitative Fish Sampling

This review of the semi-quantitative fish sampling protocol considers if the standards are clearly identified, if the methods are likely to provide a reasonable level of accuracy and precision, and if precautions are in place to ensure fish are treated humanely.

The rationale for developing a fish sampling protocol is to extend biomonitoring to areas where macroinvertebrates are difficult to sample, specifically larger wadeable streams and rivers from fall through spring. Monitoring fish is appropriate because some fish species have recreational, commercial or conservation importance. Fish may provide a similar indication to macroinvertebrates of the health of the ecosystem, but it appears that the fish monitoring will be used primarily as a supplement to macroinvertebrate monitoring.

The “Semi-Quantitiative fish sampling protocol” indicates that a fish index of biotic integrity (IBI) is in development for the Ohio River basin and the combined Potomac and Susquehanna River basins. A fish IBI for each basin is a logical choice because the IBI is routinely used to convert biological information into a score of ecosystem health. It is also appropriate to develop different standards for each basin because of differences in some of the fish species (including species introduced to a basin). The actual standards were not presented, so it was not possible to review them. This revision of the methodology is an opportunity to coordinate with neighboring states when developing the IBI standards to ensure that standards are uniform or comparable within a river basin.

It is appropriate to develop a standardized sampling protocol to generate data that will be comparable among sites and over time. However, a thorough review of the protocol is not possible because without the standards it was not possible to determine the accuracy and precision of the IBI metric that could be generated with the proposed sampling protocol. Instead, I considered another of the original IBI metrics of Karr et al. 1986, which includes standards for species richness and composition, trophic composition (%) and fish abundance and condition. Karr et al. 1986 indicated that the sampling methods to generate a fish IBI should be adequate to generate a snapshot of the entire fish community including small and large individuals as well as species that are not recreationally or commercially important. The use of electrofishing is likely to be appropriate for the systems targeted (wadeable streams) as well as the sampling effort in terms of number and size of electrofishing units and number and experience of personnel. Electrofishing has the ability to capture both small and large individuals and can be used in a variety of habitats. The reach is of sufficient length to capture a variety of habitat types and thus be representative of the rest of the stream.

While we recognize the need to reduce costs and maximize efficiency, we are concerned that single pass will provide a sample that may not be accurate (biased towards certain species that favor the habitat sampled) or precise (differences in sampling efficiency among species and size classes will affect the percentage of individuals caught). Lack of block nets adds an additional source of error because an unknown number of fish may enter or leave the sample reach. The risk is that low accuracy and precision will increase the amount of time required for a change in species composition or IBI to be detected, and the change in stream status between attaining and impaired will be delayed.

Mark-recapture or multiple pass depletion sampling could provide a more robust measure of the stream fish community, although these methods also have their flaws. Mark-recapture misses small fish that cannot be marked and preventing migration is impractical in most streams and rivers. Mark-recapture also requires the initial capture and marking of a large portion of the population (>50%) to reduce error bars to acceptable levels. Depletion sampling also has a number of assumptions that are likely to be routinely violated, such as consistent fishing effort and sampling efficiency among passes, no change in fish behavior, and no migration. This more intensive effort may be justified in some circumstances.

The protocol does not indicate if the physical condition of the fish will be recorded. The Karr et al (1986) IBI protocol used percentage of fish with lesions or parasites as one of the standards. The data sheets also indicate that species will be enumerated, but there is no space to record length or weight or other observations of individual fish. Although length and weight may not be necessary for the IBI, these data can be used to calculate biomass, and infer population dynamics and production rates.

The sampling protocol indicates that the intention is to return most of the fish alive to the stream, but it is not clear if handling will be sufficient to reduce stress and minimize incidental damage or mortality. The crew chief should have training in electrofishing theory and safety, such as is provided by the US

Fish and Wildlife Service. Effort should also be taken to maintain low densities in holding tanks, minimize warming during hot days and prevent direct exposure to the sun, maintain adequate dissolved oxygen and reduce buildup of waste products. Coldwater streams should not be sampled when the air temperature exceeds a given threshold (e.g., 75°F). Accidental morality should be recorded so that methods that do not work can be identified and improved upon.

When a fish is retained, a portion should be preserved in ethanol instead of formalin so that it's tissue will be useable for DNA analysis, which can also be part of the identification processes.

The protocol should also define a taxonomic specialist versus a taxonomic expert, and indicate the training/experience required and how their proficiency is measured.

Chemistry – Bacteriological Evaluations

We wondered how a short-term chemical violation of a regulatory threshold relates to being classified as not sustaining a designated use? For example, what is the view on accidental spills or infrequent discharges that violate a number of chemical thresholds and possibly result in biological changes (e.g., dead adult fish) that might be measurable for more than a year? We are sure this has been sorted out, taking into account the various types of data, but some clear examples would be helpful to the public.

Pg. 4, Para. 1-7 – It would be nice to know if and how regulatory standards may change as more data become available from continuous monitoring, and if continuous monitoring provides indirect insights into environmental stressors that cannot be easily monitored continuously (i.e., rely on grab samples). Two key elements in stressor response are intensity and duration. Frequency is a third element, but one that has not been examined extensively (e.g., variable but repeated stress events). Continuous monitoring is going to provide data that were previously unavailable at many sites, making the chemical data from a site a combination of continuous data and far less frequent observations.

Pgs. 7-9 – We are pleased that PADEP has attempted to bring this this section on sampling error to the regulatory discussion. However, we are not sure the public fully understands the difference between accuracy and precision, and this section may not help. It is possible the use of the term "Error" here may confuse the public because they view an error is a mistake (an inaccurate representation of true conditions), not just sampling variability that confounds confidence in conclusions. We are not sure they see that incomplete data could result in an inaccurate representation of true conditions, but it does not have to and frequently may not. That said, a "temporary, rare violation of criteria" may be ecologically significant even if it does not exceed the 1% threshold. There is a lot here on statistical significance, but not much about ecological significance.

We think this issue highlights an important question about all of these sections – who is the primary audience? It is primarily PADEP employees trying to complete the monitoring (i.e., as SOPs), the public trying to understand the outcome of PADEP monitoring, or the regulated community trying to implement the protocols in their own work? The challenge of communicating effectively to a diverse audience at the same time is no small task because of the diverse backgrounds. Given the number of sampling approaches to addressing the different stressors of interest, we believe that most audiences outside of PADEP will be looking for more information. For example, a table that highlights recommended critical sampling periods, sampling frequency, and acceptable exceedance frequency would help all see the challenges for PADEP to collect and interpret data collected annually, monthly,

weekly, daily, and/or hourly. It is possible that this table would not be easily generated and very complicated because of all of the exceptions within and among parameters.

Pg. 11, Para. 3 – Here and in other places, we recommend adding a page number to the direct quotes from EPA publications. That would help others see the context of any of the support statements, and find any information that was presented in support of the statement.

Pg. 19, items 5, 6, and 9 – We assume that some of the data needed to evaluate representativeness, or are considered relevant include reference sites. I would recommend mentioning that time series at the site, or comparable time series at a nearby reference site are important sources of relevant data.