pennsylvania
DEPARTMENT OF ENVIRONMENTAL
PROTECTION

BUREAU OF CLEAN WATER

WATER USE ASSESSMENT DECISION-MAKING BASED ON PHYSICOCHEMICAL AND BACTERIOLOGICAL SAMPLING

DECEMBER 2013

Prepared by:

Brian Chalfant
PA Department of Environmental Protection
Bureau of Clean Water
11th Floor: Rachel Carson State Office Building
Harrisburg, PA 17105

Acknowledgments
Thanks to Roy Irwin, Eric Money, Doug McLaughlin, and Karen Blocksom for sharing their thoughts, and to Rod Kime and Gary Walters for insightful reviews.

TABLE OF CONTENTS

INTENTION AND AIMS

This document discusses the technical, contextual, and conceptual aspects of procedures applied by the Pennsylvania Department of Environmental Protection (PADEP) to inform water use support decisions based on physicochemical and bacteriological sampling.

This document contains relatively little discussion of the planning and execution phases of physicochemical and bacteriological water quality sampling projects, such as outlining study objectives, choosing sampling plan designs, and setting data quality objectives.

PADEP strongly recommends anyone planning to collect physicochemical or bacteriological sampling data for consideration in the water use assessment process fully familiarize themselves with this document as well as the following guidance published by the United States Environmental Protection Agency (U.S. EPA) and PADEP before initiating sampling:

- Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan (U.S. EPA 2002a)

- Guidance on Systemic Planning Using the Data Quality Objectives Process (U.S. EPA 2006)

- Consolidated Assessment and Listing Methodology – Toward a Compendium of Best Practices (U.S. EPA 2002b)

- Designing Your Monitoring Program – A Technical Handbook for Community-Based Monitoring in Pennsylvania (PADEP 2001)

BACKGROUND

Federal law – specifically section 303(d) of the Clean Water Act – requires that states identify water bodies not meeting water quality standards. Pennsylvania regulations – specifically, Chapters 16, 92a, 93, and 96 in Title 25 of the Pennsylvania Code – present water quality standards and associated implementation requirements applicable to water bodies in Pennsylvania. Water quality criteria are implemented taking into account the magnitude (concentration), frequency, and duration.

SAMPLING AND INFERENCE

Within the regulatory framework outlined above, PADEP must determine if water bodies meet water quality standards. Often, these determinations require evaluating if water bodies meet water quality standards "at least 99% of the time." A number of interrelated considerations – outlined in this document – must be addressed when assessing if water bodies meet water quality standards "at least 99% of the time" based on physicochemical and bacteriological samples.

Sampling Space-Time

Water quality conditions – like many perceptible phenomena – change in space and in time. Interpreting water quality sample results often requires considering how far in space and for how long in time observed conditions can reasonably be considered representative of unobserved conditions that were not sampled.

This inferential process of using discrete, spatiotemporally-limited observations (i.e., samples) to estimate a larger set of unobserved, continuously dynamic conditions can introduce uncertainty into the use support decision process. Such uncertainty is called sampling error. Uncertainty can also enter the use support decision process through variability attributable to analytical measurement techniques, or measurement error. Sampling error and measurement error subject decisions based on sampling to decision error by introducing the potential for inaccuracy and imprecision into the observational process via sampling and analytic quantification. The aim of much of the rest of this document is to describe how the inherent variation and uncertainty introduced by sampling (i.e., sampling error) is addressed by PADEP in the use assessment decision process for physicochemical and bacteriological water quality sampling data. Variability attributable to analytical and laboratory techniques and equipment (i.e., measurement error) is discussed elsewhere (PADEP 2010). Both forms of error can be minimized by applying quality assurance procedures during sample collection, processing, and analysis.

Here’s a simple scenario to illustrate sampling error. Imagine someone dutifully collects a one-liter water quality sample from a set location in a creek each month for two years. That’s 24 one-liter samples. For the sake of discussion, let’s imagine that 600 billion liters of water flow past that sampling location each year. Many considerations arise when we evaluate how representative those 24 liters of sampled water are of all 1.2 trillion liters of water that flowed past that location in those two years. Imagine that none of the 24 samples show violation of a relevant criterion. When we look to that sampling data to inform our decision if the creek is meeting standards “at least 99% of the time,” we must ask ourselves how representative we think the *observed* concentrations are of the *unobserved* concentrations during the monitoring period. We may not observe violations if they occur at times when – or in areas where – samples were not collected. How we deal with these sampling error considerations is the primary focus of much of the rest of this document.

Sampling Error Implications

Unless we continuously observe – or census – every quantum of water in a stream, and as long as we rely on limited observations derived from sampling, we have to acknowledge the possibility of sampling error.

Ideally, use assessment decisions for surface waters in Pennsylvania based on physicochemical and bacteriological data will be informed by sampling conducted frequently enough to definitively characterize the conditions for each parameter of concern over a long enough time frame to account for variations in concentration attributable to changes in all relevant factors. This may be possible for some constituents in some locations through deployment of automated, continuous in-stream monitoring devices or through extremely intensive monitoring efforts. However, many water quality sampling efforts require human beings to visit sites with chemical test kits or hand-held probes to measure water quality conditions or to collect grab samples of the water for later analysis at laboratory facilities. Such monitoring efforts require personnel, funding, and site accessibility among other considerations. As a result, chemical and bacteriological water quality sampling often provides limited windows into the dynamic continuum of water quality conditions at any given location at any given time.

Continuous in-stream monitoring devices can measure conditions on a relatively frequent basis (e.g., every 15 minutes, every hour). Monitoring water quality conditions at such high frequency minimizes how far sample results have to be extrapolated into unobserved time, thereby minimizing the potential for sampling error. With less temporally-dense monitoring approaches (e.g., periodic grab samples collected by a person and analyzed in a laboratory) the amount of temporal extrapolation will likely need to be extended further in time, and the potential for sampling error may increase. Some continuous in-stream monitoring devices can be deployed in remote locations and set up to report observations via telemetry or through occasional retrievals and downloads. While these devices can provide extremely detailed, temporally-dense observational records, funding and staffing considerations limit the number of locations at which these devices can practically be deployed and maintained. Furthermore, many such devices can only measure a few water quality constituents (e.g., dissolved oxygen, temperature, conductivity, pH) for which water quality standards exist.

In the absence of temporally-dense observations, if we understand and have information on enough relevant variables (e.g., stream flow, precipitation, water temperature) as related to the constituent of interest, we may be able to confidently infer or extrapolate unobserved conditions from observed conditions based on empirical understanding of variability, and thereby reduce uncertainty attributable to sampling error. A wide variety of interrelated factors can contribute to spatial and temporal variation in the concentrations of water quality constituents of concern, including but certainly not limited to: precipitation rates, durations, and locations; thawing of ice and snow; stream flow; geologic and soil characteristics; annual and diel cycles of solar input; atmospheric conditions (e.g., cloud cover); discharges from permitted facilities; chemical spills; watershed drainage patterns; watershed land use; and hydrologic alterations (e.g., dams). Different water quality constituents often vary in unique ways relating to these and other factors. For example, dissolved oxygen concentrations in streams often exhibit strong annual and daily patterns attributable to interrelated patterns of solar flux, stream temperature, and photorespiratory activity of green plants. Meanwhile, total dissolved solids concentrations often vary much less with diel and annual patterns of solar flux, and more often vary primarily with stream flow and related patterns of surface runoff, geologies, and groundwater flow patterns. Knowledge and understanding of such patterns can strengthen inferences about unobserved conditions. U.S. EPA (2005) recommends,

> *"… states should decide how far out in time to extrapolate from the time at which a particular single grab was collected. EPA recommends that such decisions be based on contextual information regarding conditions when and where the grab was taken. For example, such information might include: 1) precipitation, 2) streamflow, 3) location of point source discharges in relation to the monitoring site, 4) land use patterns in the vicinity, 5) expected patterns of pollutant loading from the different kinds of sources present in the watershed, 6) occurrence of a chemical spill or other unusual event, and 7) historic patterns of pollutant concentrations in the monitoring segment and/or waterbodies similar to it."*

In some situations, it may be possible to extrapolate patterns observed at nearby, physiographically-similar, or hydrologically-similar locations to unobserved locations for certain constituents. For example, data from past monitoring show that concentrations of total dissolved solids in many lotic systems in Pennsylvania often exhibit an inverse power type response to stream flow, so it may be possible to predict – or inferentially estimate – total dissolved solids concentrations with some confidence if we know stream flow and have reason to believe the often-observed relationship between stream flow and total dissolved solids holds for the particular situation at hand.

Even if we confidently observed or inferred the condition of every possible quantum of flow at a particular location and even if we quantified the concentration of the constituent of interest with immaculate accuracy and precision, the phrase “at least 99% of the time” allows for some temporary, rare violation of criteria, and introduces another set of considerations to the use assessment process aside from sampling error.

99% of Time

Along with considerations of uncertainty introduced by spatiotemporally limited sampling of continuously-dynamic conditions, the phrase “at least 99% of the time” introduces to the use support decision process the additional consideration that some temporally rare criteria violations are acceptable. Some criteria explicitly state they must be met “at all times,” and some criteria define averaging periods and minimal sampling requirements, but many criteria do not explicitly specify associated sampling requirements or temporal aspects.

The vast majority of physicochemical and bacteriological water quality criteria – as expressed in Pennsylvania regulation – are written as numeric values: concentrations of various constituents not to be transgressed. These specific numeric concentrations comprise the magnitude components of criteria. For standards that must be met “at least 99% of the time,” PADEP must consider not only the magnitude components of standards expressed as concentrations, but also temporal components, namely the frequency with which violations of those concentrations occur and the duration of how long violations last.

The phrase “at least 99% of the time” addresses the temporal aspects (i.e., frequency and duration) of criteria for which these considerations are not otherwise specified in Chapter 93. However, the phrase “at least 99% of the time” needs further consideration and definition as it does not specify a time period to which it applies (e.g., 99% of one year; 99% of one month; 99% of one day, 99% of one minute). Contrast this with the comparatively exact specificity of the Water Contact use criterion for fecal coliform. Therefore, interpreting the “at least 99% of the time” phrase requires context-specific considerations that take into account the particular standard(s) being evaluated as well as site-specific evaluation of expected patterns of variability in the constituents of interest. The criteria for toxic substances are a bit different with regard to temporal aspects because toxic criteria are based on specified exposure durations and because U.S. EPA recommends certain acceptable exceedance frequencies for some toxic criteria, as discussed below.

Considering these temporal aspects of water quality criteria, the underlying concept in the phrase “at least 99% of the time” is straightforward: there is some acceptable – albeit relatively low – frequency at which, and duration for which water quality criteria concentrations presented in Chapter 93 can be exceeded with the water body still being considered as “meet standards.”

Note that sub-section 96.3(c), in reference to narrative criteria, stipulates that *general water quality criteria* contained in section 93.6 shall be achieved in surface waters "***at all times at design conditions***."

Sampling Design

Sampling error is influenced by how, when, where, why, and by whom samples are collected. As such, these considerations play a critical role in the use support decision process. While sampling plan design is not the focus of this document, some considerations on this topic are discussed in this section because sampling plan design largely determines what analytical procedures can tenably be used to assess the sampling data. The following discussion of sampling plan design is not intended to be exhaustive or complete by any means, rather to highlight some issues that are particularly relevant to making water use assessment decisions in the regulatory context of Pennsylvania. As stated above, PADEP strongly suggests that anyone planning to collect physicochemical or bacteriological sampling data for consideration in the water use assessment process familiarize themselves with the guidance documents listed in the introduction of this document for more thorough discussions of sampling plan design, data quality objectives, and other study planning considerations.

Thoughtful study design and execution are critical to assuring water quality sampling efforts provide the information necessary to address the study questions. U.S. EPA (2006) details step-by-step considerations of study design and data assessment. U.S. EPA (2002a) provides further specific details on designing a sampling plan. While all these sampling plan design considerations are not repeated here, PADEP feels it is important to address a few interrelated implications of sampling plan design in light of Pennsylvania's "at least 99% of the time" regulatory language. In a particularly germane excerpt, U.S. EPA (2002a) recommends,

> *"…sampling design development should also take into account existing regulations and requirements (for example, state, municipal) if they apply."*

In Pennsylvania, the phrase "at least 99% of the time" is a critical regulatory consideration in designing a sampling plan. Regardless of how the "at least 99% of the time" phrase is quantified in specific applications, the phrase implies that criteria exceedances are only acceptable for a very small proportion of time. As such, any study or investigation aiming to assess criteria must aim to observe an extreme end – or ends – of water quality constituent frequency distributions. The phrase "at least 99% of the time" implies we are concerned with assessing the 99th (or, conversely, the 1st) percentile of a given frequency distribution.

Due to interrelated considerations of decision error rates, sample sizes, and extreme percentiles of frequency distributions, it will very often be impractical to employ a probability-based sample design to assess against the "at least 99% of the time" stipulation without collecting large numbers of samples, at least at any reasonable decision error rates. It will often be the most resource-effective approach – especially when accounting for monitoring costs – to focus monitoring at times when violations are most likely to occur based on understanding of the factors affecting the constituent of interest in the particular monitoring situation being assessed. In the rest of this document, PADEP refers to these times when criteria violations are most likely to occur as **critical sampling periods**. Sampling focused on these critical sampling periods will necessarily be based on human understandings of the variables at play. In the terminology used by U.S. EPA (2002a), these critical sampling periods can be thought of as temporal "hot spots," and sampling targeted to observe these "hot spots" based on understandings of context-specific variations is referred to as "judgment-based sampling" (as contrasted with probability-based sampling). Since this targeted, judgment-based sampling is not suited to rigorous quantitative statistical analyses, assessment processes based on such sampling will necessarily draw less on quantitative statistical tools than will assessment processes based on probability-based sampling designs.

Of the various sampling plan designs discussed by U.S. EPA (2002a), PADEP believes the so-called "judgment-based" sampling design is the most suited method to assess extreme, infrequent ends of distributions stipulated by the phrase "at least 99% of the time." Other sampling plan designs (e.g., simple random sampling, systematic sampling) presented by U.S. EPA (2002a) are unlikely to provide accurate, precise estimates of such extreme ends of distributions while maintaining reasonable decision error rates without requiring large numbers of samples. For example, systematic sampling (i.e., sampling at regular temporal intervals) may be useful for certain applications (e.g., determining temporal trends) and can be attractive in terms of scheduling personnel and logistics, but is unlikely to directly observe infrequent, extreme events (i.e., heavy storm flows or conditions that occur 1% of the time or less) unless many samples are collected at relatively short intervals. Such systematic sampling will usually require a very large number of samples to accurately and precisely estimate extremely infrequent conditions.

Regarding systematic sampling, U.S. EPA (2002a) notes,

> *"… if the scale of the pattern or feature of interest is smaller than the spacing between sampling locations [or times], then the systematic pattern of sampling is not an efficient design unless the spacing between sampling locations [or times] is reduced or some other procedure such as composite sampling is introduced into the design.*
>
> *Systematic sampling would be inappropriate if a known pattern of contamination coincides with the regularity of the grid design. Such a coincidence would result in an overestimation or underestimation of a particular trait in the target population of interest."*
>
> *"Systematic/grid sampling may not be as efficient as other designs if prior information is available about the population. Such prior information could be*

> *used as a basis for stratification or identifying areas of higher likelihood of finding population properties of interest.”*
>
> *“… if nothing is known about the spatial characteristics of the target population, grid sampling is efficient in finding patterns or locating rare events unless the patterns or events occur on a much finer scale than the grid spacing. If there is a known pattern or spatial or temporal characteristic of interest, grid sampling may have advantages over other sampling designs depending on what is known of the target population and what questions are being addressed by sampling.”*

For example, a systematic sampling plan for dissolved oxygen where samples are collected the 15th day of every month at noon would be likely to sample the highest dissolved oxygen concentrations because photosynthetic activity usually peaks around midday.

Regarding judgment-based sampling designs, U.S. EPA (2002a) states that,

> *“In judgmental sampling, the selection of sampling units (i.e., the number and location and/or timing of collecting samples) is based on knowledge of the feature or condition under investigation and on professional judgment. Judgmental sampling is distinguished from probability-based sampling in that inferences are based on professional judgment, not statistical scientific theory. Therefore, conclusions about the target population are limited and depend entirely on the validity and accuracy of professional judgment; probabilistic statements about parameters are not possible. As described in subsequent chapters, expert judgment may also be used in conjunction with other sampling designs to produce effective sampling for defensible decisions.”*

As noted by U.S. EPA (2002a), many commonly-used statistical analysis methods assume either implicitly or explicitly that data were obtained using a probability-based – often simple random – sampling design. Probability-based sampling designs allow for application of statistical tools, which allow for quantification and control of decision error rates. In short, probability-based sampling designs offer the benefit of being amenable to statistical results interpretation, but may require a lot of sampling. Judgment-based sampling designs may not be conducive to standard inferential statistical analyses – primarily due to sample selection bias – but offer the benefit of more resource-efficient sampling (i.e., needing fewer observations to achieve a given level of precision) by incorporating existing understandings of the site and systems being sampled. U.S. EPA (2002a) notes that,

> *“Judgmental sampling is useful when there is reliable historical and physical knowledge about a relatively small feature or condition ….”*

> *"… whether to employ a judgmental or statistical (probability-based) sampling design is the main sampling design decision …."*
>
> *"An important distinction between the two types of designs is that statistical sampling designs are usually needed when the level of confidence needs to be quantified, and judgmental sampling designs are often needed to meet schedule and budgetary constraints."*
>
> *"Data obtained from convenience or judgment sampling cannot be used to make formal statistical inferences unless one is willing to assume that they have the same desirable properties as probability samples, an assumption that usually cannot be justified …."*
>
> *"Although statistical methods for developing the data collection design … are strongly encouraged, not every problem can be resolved with probability-based sampling designs. On such studies … the planning team is encouraged to seek expert advice on how to develop a non-statistical data collection design and how to evaluate the results of the data collection."*

When designing a sampling plan, U.S. EPA (2002a) recommends considering tradeoffs among considerations of desired data quality (e.g., characteristics of the constituents of interest, applicable analytical approaches, decision error estimates) and practical constraints (e.g., budgets, personnel, time, site accessibility) for a given constituent in a given situation.

U.S. EPA (2002a) also suggests that some sampling plan designs – like stratified random sampling – draw on aspects of both probability-based designs and judgment-based designs. Stratified random sampling can be used to more efficiently focus sampling resources to critical sampling periods at a given location based on existing understanding about variability of the constituents of concern at the particular study location (i.e., where and when criteria violations are most likely to occur). For example, an understanding – or, to use a U.S. EPA term, "conceptual model" – of dissolved oxygen concentrations could be used to define three temporal strata based on likely concentration ranges: likely low-level (pre-dawn, summer), likely mid-range (autumn and spring mornings and evenings), and likely high-level (mid-day, winter). Such a stratified approach may also incorporate spatial aspects with backwater, less-turbulent areas being more likely to have lower concentrations of dissolved oxygen than faster-flowing, more-turbulent areas mid-channel. In stratified random sampling designs, each member of the target population has a known – although perhaps unequal – probability of selection into the sample. Therefore, techniques of statistical inference can be applied to data resulting from stratified random sampling designs. Regarding stratified random sampling, U.S. EPA (2002a) notes,

> *"When stratification is based on correlation with an auxiliary variable which is adequately correlated with the variable of interest, stratification can produce estimates with increased precision compared with simple random sampling*

> *or, equivalently, achieve the same precision with fewer observations. For increased precision, the auxiliary variable used to define the strata should be highly correlated with the outcomes being measured. The amount of increase in precision over simple random sampling depends on the strength of the correlation between the auxiliary variable and the outcome variable being measured."*
>
> *"Stratified sampling needs reliable prior knowledge of the population in order to effectively define the strata and allocate the sample sizes. The gains in the precision, or the reductions in cost, depend on the quality of the information used to set up the stratified sampling design. Any possible increases in precision are particularly dependent on strength of the correlation of the auxiliary, stratification variable with the variable being observed in the study."*

U.S. EPA (2002a) also acknowledges that no sampling plan design is completely objective, noting (emphasis original),

> *"Implementation of a judgmental sampling design should not be confused with the application of professional judgment (or the use of professional knowledge of the study site or process). Professional judgment should always be used to develop an efficient sampling design, whether that design is judgmental or probability-based. In particular, when stratifying a population or site, exercising good professional judgment is essential so that the sampling design established for each stratum is efficient and meaningful."*

MAKING DECISIONS

In the past, PADEP adopted an approach that stipulated minimum data requirements for chemical and bacteriological use assessment datasets that were applied across all criteria. These requirements stipulated a minimum number of samples (i.e., 8), sampling frequency (i.e., at least quarterly) and duration (i.e., at least one year) needed to assess sampling data against any criteria. While this approach attempted to direct sampling so that a variety of conditions would be observed (e.g., different flow conditions, different times of year), this approach did not address the idea of critical sampling periods, discussed above. Regarding data quantity, U.S. EPA (2005) states,

> *"EPA encourages the collection of adequate data to make well-grounded attainment determinations. EPA has not established, required, nor encouraged the establishment of rigid minimum sample set size requirements in the WQS attainment status determination process. EPA is particularly concerned with application of such thresholds state-wide, without regard to key factors like the manner in which applicable WQC are expressed, variability in segment-specific conditions, and fluctuations in rates of pollutant loading. Rather if employed, target sample set sizes should not be applied in an assessment methodology as absolute exclusionary rules, and even the smallest data sets should be evaluated and, in appropriate circumstances,*

> *used. While it may be appropriate to identify target sample sizes as a methodology is developed, states should not exclude from further consideration data sets that do so solely because they do not meet a target sample size. A methodology may provide for an initial sample size screen, but should also provide for a further assessment of sample sets that do not meet the target sample size. (EPA suggests that states avoid setting target sample set sizes higher than the amount of data available at most sites.)"*

Presently

Presently, PADEP recommends context- and site-specific approaches to evaluate various criteria, accounting for the fact that the water quality criteria in Chapter 93 are presented in different ways, and because some parameters vary in different ways with changing natural conditions (e.g., diel and annual cycles of solar radiation, changes in stream flow) and may exhibit variable responses to these factors at different locations. PADEP believes it inappropriate to develop data requirement guidelines applicable to all criteria across the board since different monitoring efforts may utilize different means and may have different goals, and because different constituents, criteria, and situations call for different monitoring approaches. The present approach is consistent with recommendations from U.S. EPA (2005) that,

> *"Any target sample set size thresholds must be consistent with the state's EPA-approved water quality standards. Hence, when making a determination based on comparison of ambient data and other information to a numeric WQC expressed as an "average" concentration over a specified period of time, a statement of a desired number of samples may be appropriate. Still, the methodology should provide decision rules for concluding nonattainment in cases where the target data quantity expectations are not met, but the available data and information indicate a reasonable likelihood of a WQC exceedance (e.g., available samples with major digressions from the criterion concentration, corroborating evidence from independent lines of evidence such as biosurveys or incidence of waterborne disease, indications that conditions in the waterbody and loadings of the pollutant into the waterbody have remained fairly stable over the period in question)."*

All relevant data will be considered in PADEP's use support assessment process regardless of sample size, but – because waterbody assessments are made on a continual basis in an effort to document current conditions – more recent data take precedence over older data, especially in situations where conditions have recently changed (e.g., installation of pollution remediation projects, alteration of permit limits in the watershed, changing land use patterns, discontinuation of combined sewer overflows). In some instances, older and newer data may be considered in concert to document temporal trends.

PADEP makes every effort to verify the accuracy of all data used in the use support decision process. PADEP strongly encourages anyone submitting data to familiarize themselves with PADEP Bureau of Laboratories quality assurance and quality control procedures (PADEP 2010) regarding record keeping, methods documentation, sampling techniques, selection of analytic laboratories, chain of custody concerns and so forth. PADEP will not drop extreme values (AKA outliers) from a dataset unless there is reason to believe the extreme value is invalid. For example, a dissolved oxygen concentration of 100 mg/L is physically impossible at tropospheric temperatures and pressures – it is likely that such a record is a typographical error meant to really be 1 mg/L or 10 mg/L. Similarly, in a water temperature dataset submitted in degrees Celsius where one value is recorded at 72, it is highly unlikely this is a valid reading and may be recorded in degrees Fahrenheit. PADEP does not want to discount any data or information from consideration outright, so no strict guidelines are set forth with regard to what sampling designs are acceptable because different sampling approaches may be necessary to answer different questions in different situations depending on the particular constituent and water body in question.

PADEP strongly recommends that any physicochemical or bacteriological water quality sampling datasets intended for consideration in the use assessment decision process be collected using a “judgment-based sampling” design – as discussed above – with sampling targeted to critical sampling periods when water quality violations (or human exposure in the case of bacteriological samples) are most likely to occur based on knowledge of the conditions affecting the constituent(s) of interest. Some considerations are common to sampling design decisions for many constituents.

> **Diel cycles of solar radiation** – The rising and setting of the sun drives daily cycles of photosynthesis and respiration across much of the earth’s surface, including in surface waters. During peak influx of solar radiation, photosynthetic activity tends to peak, resulting in higher in-stream dissolved oxygen concentrations and pH levels. Incident sunlight also increases in-stream temperatures, which affects levels of dissolved oxygen in the water and photosynthetic rates. Other constituents (e.g., TDS, alkalinity) may also exhibit some diel cycling.
>
> **Annual cycles of solar radiation** – Northern hemisphere locations, such as Pennsylvania, receive their most intense and prolonged sunlight in the summer months of June and July, with less intense and shorter exposure to sunlight in winter months of December and January. Stream systems reflect these cycles in annual cycles of water temperature and dissolved oxygen. Fluctuations in pH are often less dramatic in winter months as well, likely due to reduced photosynthetic activity with the colder temperatures. Some nutrient parameters may also exhibit variation with annual seasons.
>
> **Annual cycles of precipitation, evapotranspiration and stream flow** – Across Pennsylvania, stream flow patterns vary annually with the lowest stream flows typically observed from July through September, gradually increasing through autumn and winter with peak flows often observed January through April and tailing off again May through June to summer base

flows, although hurricane and tropical storm remnants occasionally dump heavy rain on Pennsylvania in early autumn. These stream flow patterns reflect annual cycles of precipitation, snow and ice melt, and evapotranspiration. Such patterns vary locally with a variety of factors (e.g., soil types, hill slopes) and fluctuate year to year. Some constituents of concern often exhibit predictable patterns of response with precipitation and stream flow. For example, TDS concentrations tend to decrease consistently with increasing stream flow; in many situations this has to do with surface runoff containing lower dissolved ion concentrations than groundwater inputs to a stream, which tend to be higher in dissolved minerals. Likewise, some constituents, like various forms of phosphorus, usually enter stream systems attached to particulate matter washed in during periods of high surface runoff. Such patterns may also vary with surrounding land use, like fertilizer application, impervious surfaces and so on.

Conservative and non-conservative substances – Some water quality constituents – like sulfate – are considered conservative in that their concentrations are not directly affected by biological processes. These conservative substances do not decay, are not selectively incorporated by living organisms, and concentrations are affected mostly by sedimentation and dilution. Non-conservative substances – such as phosphorous – are removed from the water column by biological processes.

Decision Framework

PADEP will implement the following framework when evaluating physicochemical and/or bacteriological water quality monitoring data in the use assessment decision process. The details of this appraisal process may vary from application to application based on the unique characteristics and contexts of each situation. However, PADEP will follow this process as much as possible in order to maintain consistency in the use support decision process and so that interested stakeholders can clearly see how PADEP evaluates physicochemical and bacteriological sample results. This process will be documented for each use support decision. The decision framework aims to document and communicate each step of the decision process in a clear, consistent manner addressing the study designs, data quality, data analysis, assumptions, uncertainties, and consequences associated with each use assessment decision. PADEP attempts to be as concise as possible within this framework while not compromising adequate discussion of critical issues influencing the decisions.

(1) **Describe monitoring effort.** Describe the waterbody and the watershed, including basin size, land uses, geologies, and other characteristics. Discuss any germane history and context pertaining to the monitoring effort. To the extent possible, describe the motivations and intentions of the monitoring effort, including the individuals and organizations involved as well as the intended use of the information collected. Clearly state study goals. Describe and map monitoring locations. Include any photographs.

(2) **Check data quality.** Evaluate any study plans and objectives, including sampling plan design details such as record keeping, data management, training, sampling techniques, and analytical methods. Check data for typos and other anomalies. Document non-detects and censored data.

(3) **Gather information on likely sources of variation.** At a minimum, this information will typically include characterization – and quantification where possible – of tributary locations, upstream discharges, geologies, and land uses. Potential sources of this information include stream gages, climatological records, and discharge monitoring reports. Include maps, figures, and diagrams as needed. Discuss relevant physical, chemical, and biological processes and other potential sources of variation for the constituent(s) of concern. Address context-specific considerations (e.g., dams).

(4) **Explore data.** Perform various graphical analyses (e.g., histograms, probability distribution functions, boxplots, time-series plots, scatterplots with likely sources of variation, LOWESS) to visually explore and illustrate data characteristics. Document summary statistics (e.g., minimum, maximum, mean, median, standard deviation).

(5) **Evaluate data representativeness.** Evaluate how representative samples are of unmonitored conditions, mindful of the sampling plan design (e.g., sample collection frequency, locations, timing, targeting) and the likely sources of variation with special attention to any critical sampling times and locations. Consider if the system is likely to be spatially well-mixed at monitoring location(s) and how quickly conditions are likely to change in time.

(6) **Describe the relevant standards.** Identify which criteria are being evaluated and the uses to which they apply. Describe how the constituents of concern impact the protected use (i.e., exposure pathways, detrimental effects). Review the associated regulatory language including any relevant criterion rationale documentation.

(7) **Apply appropriate analytical procedures.** Select and apply appropriate analytical techniques, mindful of the sampling plan design, monitoring objectives, and the relevant criteria, constituents, and context. State and verify any assumptions associated with each analytical technique. Evaluate decision error rates, if applicable. For hypothesis tests, evaluate null hypothesis choice. Discuss the frequency, duration, and magnitude of any criteria violations.

(8) **Consider other sources of relevant use support information.** Additional sources of information may include: previous or concurrent monitoring efforts; data from water supply intakes; biological surveys; and discharge monitoring reports.

(9) **Evaluate all relevant lines of evidence.** Bring together the previous steps into a narrative that addresses contextual data interpretations, possible counter arguments, alternative decision choices, and decision consequences, including evaluation of decision error consequences. Explicitly address any policy ramifications if applicable.

(10) **Decide.** Decide what to do with the dataset and waterbody in question. At a minimum, each decision will include placing the waterbody in one of the Integrated Report categories.

Clarifications

When compared to physicochemical or bacteriological sampling datasets used to make listing decisions, datasets used to inform **delisting decisions** (i.e., decisions to remove a waterbody from Category 5 of the 303(d) list) must: (1) have been collected more recently; (2) have been collected as frequent or more frequently; and (3) contain more samples.

For criteria written as time-average criteria, PADEP encourages multiple sampling events within the specified time periods. For example, the total iron aquatic life use criterion is written as a 30-day average, so PADEP encourages sampling multiple times in a given 30-day period to compare existing conditions to the standard. Like any sampling, this sampling should represent the most likely violation times and spaces as discussed above. As a general guideline, PADEP encourages at least three sampling events in the time period expressed in various time-averaged criteria.

Section 93.8 presents two types of water quality criteria for toxic substances designed to protect human health: threshold effect criteria and cancer risk criteria. Regarding threshold effect criteria, section 16.32 states:

> (a) A threshold effect is defined as an adverse impact that occurs in the exposed individual only after a physiological reserve is depleted. For these effects there exists a dose below which no adverse response will occur. Threshold toxic effects include most systemic effects and developmental toxicity, including teratogenicity. Developmental toxicity includes all adverse effects in developing offspring resulting from prenatal exposure to a causative agent.

And regarding cancer risk, or non-threshold effect, criteria, section 16.33 states:

> (a) A nonthreshold effect is defined as an adverse impact, including cancer, for which no exposure greater than zero assures protection to the exposed individual. Thus, in contrast to the threshold concept discussed in § 16.32 (relating to threshold level toxic effects), the nonthreshold approach to toxics control is based upon the premise that there is no safe concentration of the toxic.

Currently, U.S. EPA policy establishes that the duration for human health criteria for carcinogens should be derived assuming lifetime exposure, taken to be a 70-year time period (U.S. EPA 2007).

All criteria for toxic substances – aquatic life acute criteria, aquatic life chronic criteria, human health threshold criteria, and human health non-threshold criteria alike – are expressed as maximum values.

For toxic pollutant assessment decisions, PADEP follows the guidelines set forth by USEPA (i.e., CMC pollutants violating criteria more than once in a given 3-year period constitute an impairment; CCC pollutants with either a 30-day mean concentration violating a criterion more than once in a 3-year period or a 4-day mean concentration exceeding twice the CCC in any 4-week period constitute an impairment). The 3-year period stipulated for toxic pollutants starts at the time monitoring begins and not when the first exceedance occurs, unless they coincide. For example, if a single violation is observed at the end of the year for a routinely monitored station, another violation must not be observed at that site for the next two years for the water to still be considered attaining its use for that constituent. Similarly, if the first sample at a previously unmonitored site shows a toxic criteria violation, another violation must not be observed at that site for the next 3 years for the water to still be considered attaining its use for that constituent. The 4-week period stipulation is applied in the same manner.

Guidance from U.S. EPA defines the exposure period associated with CMC as one hour and with CCC as four days. Federal regulations specify that the acute criteria – or CMC – for a pollutant must not be exceeded more than once in a given 3-year period for any water body. For CCC, U.S. EPA guidance states that: (1) The 30-day mean concentration of a pollutant in a body of water must not exceed the CCC for that pollutant more than once in a 3-year period; and (2) no 4 consecutive samples collected on different days during a 4-week period can exceed twice the CCC. For example, 10 samples are collected within a 4-week period, any 4 consecutive samples are averaged and compared against the CCC for Benzene (130 µg/L); if the average of those samples exceed 260 µg/L (twice the CCC of Benzene), then the criterion is violated and the water is impaired.

For conventional (i.e., not toxic) pollutants, PADEP considers grab samples to be representative of a day unless convincing evidence exists to suggest otherwise (e.g., a documented spill, influence of a known biological process, supporting high-frequency monitoring data). Under this assumption, four days with conventional pollutant criterion violations in a year constitute use impairment for criteria expressed as maxima or minima because the standards are violated more than 1% of the time (i.e., 4 days / 365 days ≈ 1.1%, which means criteria are being met less than 99% of the time). For conventional pollutants expressed as 30-day or monthly averages, any one month or 30-day period showing a criterion violation will be considered an impairment based on the reasoning that the water is not meeting standards 99% of the time (i.e., 1 month / 12 months ≈ 8.3%).

For bacteriological monitoring, the critical sampling period should capture the time frame when the public is likely to be engaging in primary water contact recreational activities (i.e. May 1 through September 30). Sampling should not take place during those times when physical conditions such as stream discharge render primary contact recreation hazardous to the public. Typically, elevated fecal indicator levels are observed during periods of high discharge. However, people are unlikely to be participating in primary water contact activities during those conditions. Another consideration is that collecting samples during high flow conditions can be dangerous to the collector. Therefore PADEP recommends that monitoring bacteriological parameters for the purpose of Recreational Use Assessment occurs during average or low flow conditions when primary contact recreation is likely to occur in a particular stream. This can be determined by using best professional judgment of average flow conditions for a particular stream. A general guidance is that sampling should not take place immediately following 0.25 inch of rain or more.

REFERENCES

Department of Environmental Protection. 2001. Designing Your Monitoring Program – a Technical Handbook for Community-based Monitoring in Pennsylvania. Harrisburg, Pennsylvania.

Department of Environmental Protection. 2003. Water Quality Antidegradation Implementation Guidance. 391-0300-002. Harrisburg, Pennsylvania.

Department of Environmental Protection. 2010. Quality Assurance Manual for the Pennsylvania Department of Environmental Protection Bureau of Laboratories. Harrisburg, Pennsylvania.

United States Environmental Protection Agency. 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. PB85-227049.

United States Environmental Protection Agency. 2002a. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan. EPA QA/G-5S.

United States Environmental Protection Agency. 2002b. Consolidated Assessment and Listing Methodology – Toward a Compendium of Best Practices. First edition.

United States Environmental Protection Agency. 2005. Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act.

United States Environmental Protection Agency. 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4.

United States Environmental Protection Agency. 2007. Water Quality Standards Handbook. Second edition. EPA-823-B-94-005.