Appendix I – USEPA Data Quality Objectives Process

This page intentionally blank

USEPA Data Quality Objectives Process

USEPA (2000) describes a seven-step process (**Figure 1**) for establishing DQOs that integrates decision making with the more technical aspects of study design. These seven steps form the core of the process recommended for developing custom monitoring and assessment programs for individual WMAs. This process should be explicitly applied to ensure that all monitoring programs are designed in a consistent manner and meet a common set of design criteria. A requirement for implementing a custom monitoring design should be documentation that all 7 steps of the USEPA process have been completed. The descriptions of the 7 steps in the DQO process below are brief summaries of more detailed descriptions and discussions in USEPA (2000). This systematic description and evaluation of information needs will help guarantee that existing data (e.g., from past special studies or historical monitoring and assessment efforts) will be fully utilized before resources are committed to collection of new data.

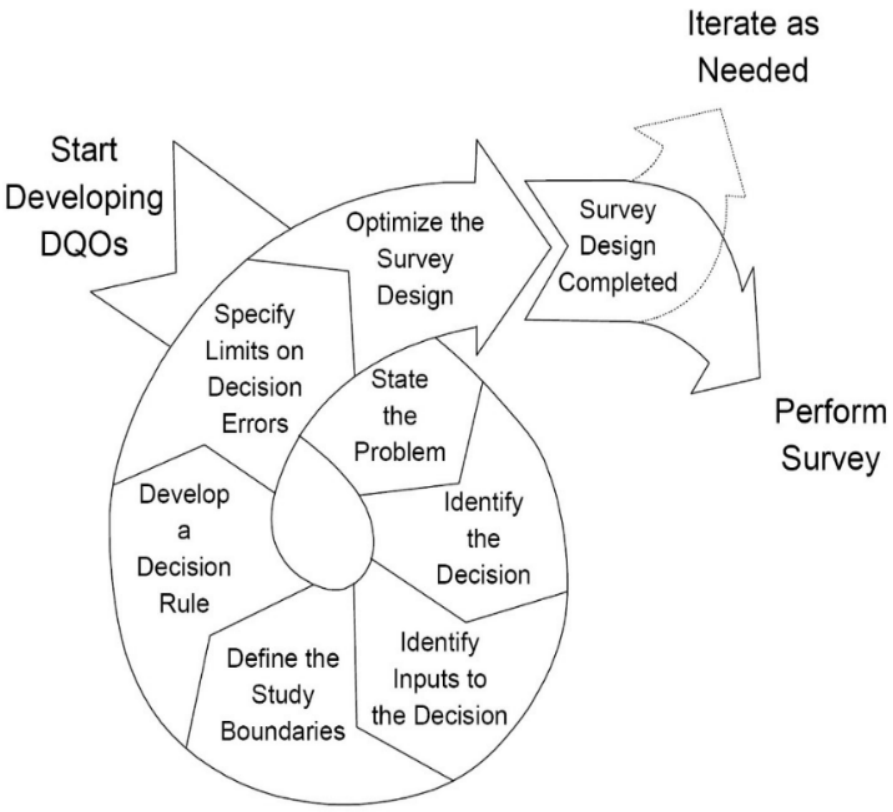

Figure 1. The 7 Steps in the USEPA Data Quality Objectives Process (USEPA 2000).

Step 1: State the Problem

The necessary first step in the study design process is developing an explicit definition of the problem that is broadly agreed on by all interested parties. This involves:

- Identifying the planning team members, including decision makers
- Describing the problem, including developing a conceptual model of the issue being investigated
- Determining resources available, including budget, personnel, and schedule

Both Bernstein (2010) and Regional Water Board (2012) emphasize that issues related to the assessment and management of environmental problems can be framed in terms of one or the other of the hierarchical questions (M1 – M4) illustrated in **Figure 3**. Agreement on such high-level questions is an essential starting point for monitoring and assessment design.

Conceptual modeling can then organize information and knowledge about the problem (**Figure** 2). Such graphical tools are valuable in facilitating discussion among parties with differing perspectives. They can represent varying levels of detail and also help identify the prioritization of data gaps and assist in proper resource allocation. Conceptual models can provide a logical structure for more quantitative modeling. For example, the darker green nodes in the conceptual model in **Figure 2.b** are those that can be more directly influenced by management actions.

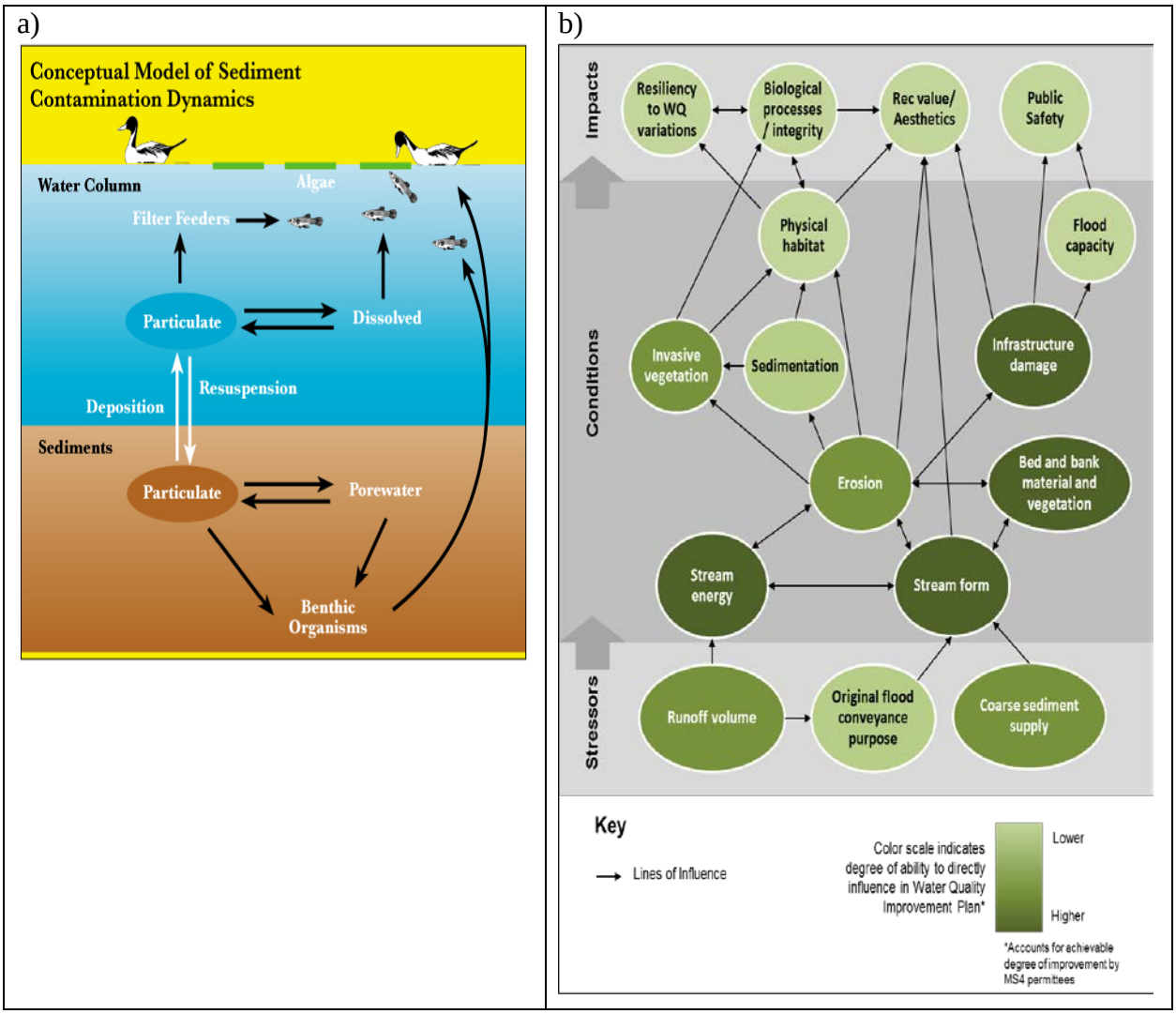

Figure 2. Example Graphical Conceptual Models Representing (a) Sediment Contamination Dynamics in an Estuarine System and (b) Stormwater Impacts on Stream and Watershed Processes in Wet Weather. (a: Courtesy of the Orange County Stormwater Program; b: Figure 7 from Orange County Public Works 2016)

Step 2: Identify the Decision

The second step in the study design process is specifically identifying the decision(s) facing managers, in as much detail as possible. This involves:

- Identifying the principal study question
- Defining alternative actions
- Developing a decision statement
- Organizing multiple decisions

Vague or poorly thought out study questions do not provide sufficient guidance for study design because they do not aid in determining which data are relevant and which are not. Principal study design questions can be stated simply, for example, "Does the concentration of contaminants in ground water exceed acceptable levels?" or, "Does a contaminant pose a human health risk?" or, "Are exceedance rates increasing or decreasing over time?". Such questions provide needed overall direction to the study design effort, but must be supplemented with additional information as described in subsequent steps of the process.

Where alternative management actions (e.g., a compliance determination, BMP implementation) are possible, a decision statement combines the principal management question with these alternatives. For example, "Determine whether levels of contaminant X require source control actions." Where complex issues involve multiple issues it is important to define their logical relationship (e.g., persistent exceedances are a necessary prerequisite for source control actions) so that monitoring and assessment efforts can be scheduled and sequenced intelligently. For example, proposed adjustments to the Bacteria TMDL call for first confirming that human sources of contamination are present before proceeding with site-specific, and more intensive, source identification studies. Such relationships can be organized and examined in conceptual models, in related decision trees, or with scenarios based on actual sample data or on simulated monitoring results.

Step 3: Identify the Inputs to the Decision

The third step in the study design process is identifying the inputs to the decision. This involves:

- Identifying the information needed
- Determining the sources of information
- Determining the basis for setting the Action Level
- Identifying sampling and [data] analysis methods that can meet the data requirements

An understanding of what information is needed should be based on the principal study question, the decision statement, and the problem conceptual model. At this stage of the process, needed information can be defined somewhat qualitatively (e.g., levels of contaminant X in the tissue of species X, trends in estimates of loads), with the understanding that more detailed and quantitative definitions will be developed in subsequent steps of the process. Once the types of needed information are defined, then their source(s) can be identified. It will be important to specify whether information must be gathered from field samples, estimated with models, or drawn from the existing literature or past studies, among others. Too often, a default assumption is that a new

study requires newly gathered data, with the result that existing data resources may be underutilized.

Sampling and data analysis methods should also be identified qualitatively in this step (e.g., comparison to standards, characterization of trends over time), although it will not be possible to develop complete specifications until the study boundaries, adequate levels of accuracy and precision, and an optimized study design have been defined. There is thus a necessary iterative element to the steps of the study design process, even though they are presented more simply as a set of linear steps from 1 to 7. Note that sampling and data analysis methods are referred to together in an effort to avoid the problem of waiting to define data analysis methods until after all data are gathered.

The Action Level is a key element of this step. It is the criterion for decision-making, for choosing one alternative over another. This can be a regulatory standard, a cost threshold, the relative difference between two outcomes, or other criterion. While Action Levels are often understood implicitly by decision makers, they are not always expressed explicitly or quantitatively. This is especially true where decisions have a strong values component, involve complex tradeoffs, or stem from situations with inputs that are inherently difficult to quantify. Setting Action Levels typically requires iterative interactions between managers and technical staff in order to identify which types of information (e.g., an increase/decrease of more than a certain amount in loads over a certain period of time) provide the best basis for decision-making. This will not always default to the most easily measured or quantified types of information. It is important to bear in mind the old saying that, "It is more important to have a rough answer to the right question than a very precise answer to the wrong question."

Step 4: Define the Boundaries of the Study

The fourth step in the study design process is defining the boundaries of the study. This involves:

- Defining the target population of interest
- Specifying the spatial boundaries that clarify what the data represent
- Determining the time frame for collecting data and making the decision
- Determining the practical constraints on collecting data
- Determining the smallest subpopulations, area, volume, or time for which separate decisions must be made

Step 4 extends the process of adding specific detail to the management questions and decisions by defining the specific set of contexts (e.g., space, time, logistics) that will help determine the specifics of the study design. This step requires extensive technical input to ensure that factors such as watershed processes, species distributions, seasonal and inter-annual cycles and patterns, or the capabilities of various sampling techniques are accounted for in the study design. Particular attention should be paid to defining the target population to ensure that it is directly related to management questions, the decision(s) to be made, and the decision statement and Action Level.

In addition, as details are added to the developing study design in this step, the conceptual model should be updated to ensure that added details fit within the problem description and that the conceptual model continues to represent the picture of how the problem is structured and how its

various elements interact. Similarly, preliminary decisions about preferred data analysis methods should be revisited as the technical detail in the study design increases.

Step 5: Develop a Decision Rule

The fifth step in the study design process is developing a decision rule. This involves:

- Specifying an appropriate population parameter (e.g., mean, median, percentile)
- Confirming the Action Level exceeds measurement detection limits
- Developing a decision rule (i.e., *if … then …* statement)

Just as Step 4 adds detail to the environmental and logistical aspects of the study design (e.g., time and space boundaries, logistics), Step 5 adds analogous detail to the specifics of the decision process. There is a wide range of potential parameters that can be calculated, even from the same original dataset. Each potential parameter will have unique statistical properties and implications for both the study design and for decision-making. Thus, means and medians, while both measures of central tendency, communicate different information about the underlying dataset, as do means and measures of maximum or minimum values. Similarly, raw data can be rescaled to fit on a consistent interval (e.g., 0 – 100), combined into multi-metric indices, or transformed to adjust their underlying distribution. Selecting the appropriate parameter for analysis and decision-making thus requires close interaction between technical staff and decision makers in order to ensure that the effects of different parameters on decisions are fully understood.

Confirming that the Action Level exceeds applicable detection limits is typically a straightforward process; however, developing the decision rule can be more challenging. The decision rule specifies how study results will be used in decision making. Not all decisions are as straightforward as comparing a value to a standard, and instead many require dealing with ambiguity or uncertainty, inherent limitations on understanding, or unresolved data gaps. The "if … then …" logical structure thus represents an ideal outcome that is not always possible to achieve. Discussions with managers framed around realistic scenarios can be a valuable and effective means of developing practical decision rules.

Step 6: Specify Tolerable Limits on Decision Errors

The sixth step in the study design process is specifying limits on decision errors. This involves:

- Determining the range of the parameter of interest
- Choosing a null hypothesis
- Examining consequences of making an incorrect decision
- Specifying a range of values where consequences are minor (gray region)
- Assigning probability values to points above and below the Action Level that reflect tolerable probability for potential decision errors

Step 6 represents an effort to address uncertainty and error and to incorporate the results into the study design. This step is often omitted from study design efforts for three main reasons. First, it requires familiarity with important concepts in statistics, probability, and uncertainty, such as Type I and Type II errors, or false positives and false negatives. Second, it requires clearly acknowledging that there is unavoidable uncertainty associated with decision making. Third, it

depends on evaluating the consequences of false positives and false negatives for decision making and policy outcomes and weighing a variety of often complex tradeoffs. Thus, Step 6 represents a crucial difference between the USEPA data quality objectives approach and most typical study design efforts. A number of iterations of Steps 6 and 7 (see next section) are often needed to arrive at an optimal balance among uncertainty, the consequences of false positives and false negatives, and the investment in sampling and analysis needed to reduce uncertainty.

The USEPA guidance document presents graphical and quantitative tools for performing the components of Step 6.

Step 7: Optimize the Design for Obtaining Data

The seventh step of the study design process is optimizing the sampling design. This involves:

- Reviewing the DQO outputs
- Developing data collection design alternatives
- Formulating mathematical expressions (i.e., statistical models) for each design
- Selecting the sample size that satisfies the DQOs
- Deciding on the most resource effective design
- Documenting details in the Quality Assurance Project Plan

The main purpose of Step 7 is to ensure that the study design will achieve the statistical and economic efficiency needed to accomplish the study's goals without either oversampling or under-sampling, both of which represent a waste of resources. This can be accomplished with optimization analyses that use the various sources of error (e.g., sampling error, temporal variability, spatial variability) to estimate the combination of sampling effort and replication that will accomplish the largest overall reduction in variability for the least amount of cost.

Alternative sampling designs can also be evaluated in terms of their statistical power, that is, the likelihood that they will detect a signal in the data that is actually present. **Figure 3** illustrates results of a statistical power analysis for one study design option evaluated by the San Diego County Stormwater Program. Analyses such as this illustrate where additional sampling effort produces little or no useful return on investment, as well as the amount of change it is feasible to detect within the logistical constraints of the study and the timeframe of the decision.

Both optimization and statistical power analyses depend on the development of explicit statistical models for alternative study designs that specify the sources and magnitudes of error or variability.

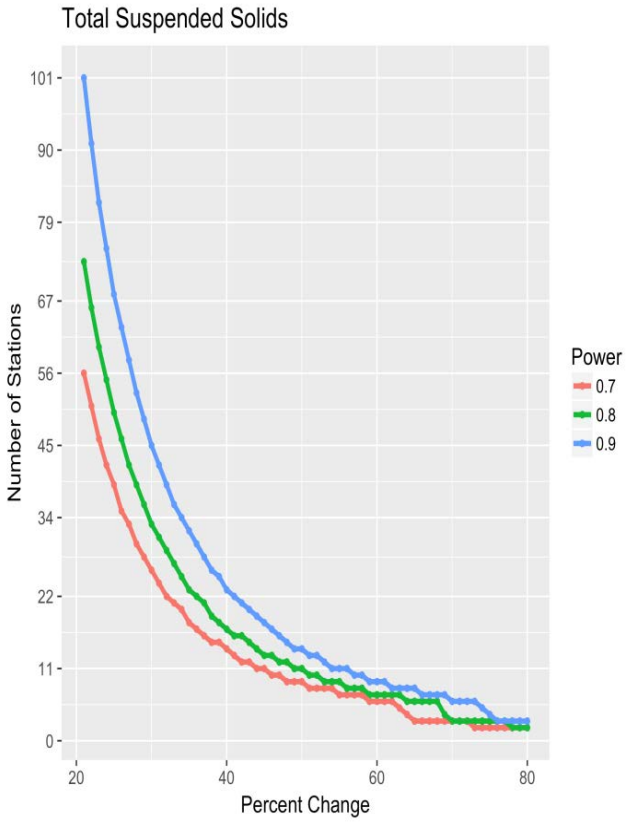

Figure 3. The Expected Sample Size Required to Detect a Specified Percent Change in Concentration for Total Suspended Solids with Power of 0.7, 0.8, and 0.9. Data and Figure Courtesy of the San Diego County Stormwater Program.