DQM Information Paper 8.2.4

Representativeness of Environmental Monitoring Data

By Revital Katznelson, Ph.D.

1.0 About this Information Paper

(This paragraph is essentially common to DQM Information Papers from the 8 and 9 series. If you have seen it already, please skip to the next paragraph below). This Information Paper is a new type of guidance. It has been created for our new integrated system of guidance and tools for water quality monitoring called "the Data Quality Management (DQM) System". DQM is implemented by the Clean Water Team (CWT) where needed to support the collection of **reliable** data of **known quality** in a fully documented, **scientifically defensible** manner. Most DQM materials are delivered in Parameter-Specific Folders, which provide both the traditional "protocol" materials and new, expanded guidance in three types of inter-related documents: Fact Sheet, Information Paper, and Standard Operation Procedures.

This Information Paper (IP), a part of the generic DQM contents materials, provides "big picture" information on data representativeness. If you are a Trainer or a Technical Leader of any monitoring project, this IP will prompt you to consider most of the important aspects of environmental representativeness as you design your study. You will also find references to other resource materials, including other DQM-IPs (e.g., on sampling design principles) and SOPs that provide detailed check-lists and instructions on how to collect representative data. Thus, this IP will enhance your ability to generate data that are comparable to other data sets and to communicate your sampling intent and design effectively. Note that this IP is focused only on environmental monitoring and does not relate to political, economic, or social representativeness, nor does it relate to 'representativeness' in the sense sample integrity (lack of sample contamination or deterioration).

Section 2 of this IP lists the different aspects of data representativeness. Section 3 provides an operation-oriented discussion of representativeness, with some tangible examples. Communication of representativeness is discussed in Section 4. Finally, the "Sources & Resources" section (Section 5) provides contact and website leads into further information.

2.0 Aspects of Representativeness

Box 8.2.4-1 shows a generic list of things to consider when designing a sampling effort that would generate a data set of known representativeness. There is one common principle underlying all these lists: You have to **separate the issues** in such ways that you – or anyone else using the data you generate - would not be comparing apples and oranges. For example, can you compare the worst-case-scenario temperature conditions (temporal) as measured in the best in-stream refuge area (spatial), to the temperature that was measured at random in the hypolimnion of a lake?. Box 8.2.4-1 provides some separation ideas.

Box 8.2.4-1: Aspects of Environmental Representativeness

Sampling Intent

Sampling intent captures the reason for sampling, or the essence of the study question. It affects the way a sample will be (or have been) collected in a profound way. Intent is applied on a spatial scale separately from the temporal scale. Examples include:

Spatial (Station Selection Intent):

- Overall Watershed Characterization
- Effectiveness of Management Practices
- Compliance Monitoring

Temporal (Timing selection Intent)

- Worst Case Scenario conditions
- Storm runoff monitoring
- Routine Monitoring @ same time of day, e.g., for trend analysis

Sampling Design Principles

Sampling Design is the approach taken for deciding where and when a sample will be (or has been) collected. Principles are based on the probabilistic approach or the deterministic approach, but sometimes sampling is done in a non-deliberate way. Options of Sampling Design Principles include:

- Random (stratified)
- Directed
- Systematic
- Non-deliberate (Anecdotal)

Location: Station type

Station Type conveys the type of drainage captured or the type of Waterbody sampled, and affects the content of environmental samples in a profound way. Examples include:

- Storm Drain Outfall
- River/Stream
- Lake
- Ocean

Weather Conditions

In California, the properties of a storm runoff sample are profoundly different from those of a sample collected during base flow. In addition, the interval between storm runoff events (as well as the amounts of rain received) affects the properties of a storm water sample. Relevant aspects include:

- Current flow conditions: wet (storm runoff) or dry (base flow)
- Antecedent dry period (days of no rain)
- Antecedent rain (totals for season and for the last 24 hours)

Inherent Variability

The environment is patchy, and it changes all the time. Oftentimes (especially during the low-flow periods of the California summer and fall), the exact spot or the exact hour in the day may affect the measured value or the content of an environmental sample in a profound way. Inherent Variability is apparent in two dimensions:

- Spatial (location in the stream channel)
- Temporal (time of day)

3.0 Collecting data of known representativeness

Representativeness does not happen; you have to make it happen. Very few people start their careers knowing exactly how to plan the sampling activities so that they will capture all the attributes of interest, or how to physically collect monitoring results that truly represent the environment. You build this knowledge over time, through your own experience and the experience of others. Sometimes the use of checklists helps.

Box 8.2.4-2 shows several sets of operational considerations related to different scales, principles, or waterbody types. This can be a starting list of items to think about when planning a monitoring effort. Essentially, **what each of your data point will eventually represent is totally dependent on the locations, timing, and compositing methods you have selected** [and hopefully documented!]. The critical question is: **what do you want your sample to represent**? For example, do you want it to represent the entire water body? The stream area just downstream of that big outfall? The thermocline? Unfortunately, the environment is patchy, and it changes all the time. In most aquatic systems, water temperature or nitrate concentrations are a constantly-moving target, and catching a representative spot at a representative moment is a challenge.

Because the environment changes all the time, you cannot choose a location and timing that will fit what you are shooting for if you do not know anything about how your characteristic of interest is actually distributed or how it changes over time. This is why it is always essential to have some characterization data prior to finalization of location and timing decisions. For example, if you are interested in collecting a sample that represents the surface water in your shallow lake during the summer, you need to know these facts: (a) the lake is stratified daily; (b) the wind starts mixing it at approximately 1 PM; (c) the wind action causes upwelling of hypolimnion water at the upwind end of the lake by 2 PM; and (d) this brings up water with profoundly different water quality. In other words, you simply need to know that the upwind end or the lake represents hypolimnion waters – not surface waters - on windy afternoons. Sometimes you can use easily-measured characteristic, such as temperature or specific conductivity, to see patterns or track changes, and make inferences on the distribution of associated characteristics (for the above example, lower temperature represents the hypolimnion, which also has lower levels of dissolved oxygen, less algae, and higher levels of ammonia).

So far we have discussed collection of one grab sample. You can always increase the representativeness of a sample by combining several sub-samples to create a composite sample, and Box 8.2.4-2 shows a number of compositing principles. However, a composite sample does not provide information on the inherent variability, and you will need to collect and analyze multiple samples to provide that information.

Box 8.2.4-2: Operational Considerations: Scales and Options

Location: Selection Considerations (Horizontal and vertical)

Streams and Rivers:

- Four spatial scales: Stream Network (Reach); Channels and confluences; Position in water column; Orientation of the sampling device Intake (e.g., isokinetic sampler)
- Tidal Influence
- Stratification and stagnation

Ponds and Lakes:

- Location of inlets and outlet
- Wind direction, wind intensity, fetch
- Stratification, Thermocline depth, photic zone

Conditions: Timing Selection Considerations

- Seasonal scale (e.g., summer flow or wet season flow conditions)
- Diurnal scale (e.g., peak temperature conditions)
- Event scale (e.g., storm runoff, irrigation runoff, discharge, etc.)

Sample Compositing options

- Random
- Systematic (e.g., grid)
- Distribution-pattern based
- Flow-weighed
- Time –weighed

Statistical strength considerations

- Types of statistical analyses and comparisons needed
- Number of independent Samples
- Ability to detect change
- Tolerable measurement error
- Tolerable rate of wrong decisions
- Level of confidence

The information about inherent variability may be essential for your future data analysis needs if you plan any hypothesis-testing, whether for comparisons of your data with water quality benchmarks, or for upstream/downstream (or before/after) comparisons, or for trend analysis. In other words, you need to think about the type of statistical comparisons you may want to make with the data, and for this you need to shoot for a certain statistical power. When you conduct what is known as “power analysis” you will need to know the inherent environmental variability, and plug this into the equation, together with the amount of error that you can tolerate and the level of confidence you need, to calculate the number of samples you will need to collect. There are several web sites that have power calculators on line. The USEPA Data

Quality Objectives process (USEPA 2000, QA/G-4) walks you through the aspects you need to consider about your decision and then directs you to the DQO calculator on-line.

You can use the contents of Boxes 8.2.4-1 and 8.2.4-2 to do your planning, and add other aspects and considerations that may be relevant to your specific project. You can also use the contents of Boxes 8.2.4-1 and 8.2.4-2 to communicate your plans to technical experts and get their input, and it is highly recommended because other eyes always see more good things to think about. The Clean Water Team has incorporated these concepts into an example of a Monitoring Plan for training purposes (see Section 5 below).

4.0 Communicating Representativeness

In the previous two sections you were offered two check lists to help you with targeting the monitoring activities to your needs. The same verbiage can be used to communicate your plans to a panel of technical experts, and to solicit their feedback on your plans, before you start your field activities. And – not surprisingly – the same verbiage can also be used to inform the user of your data what each of your data point represents (provided that you have actually followed the plan!). You can communicate these aspects in written reports, and you may also be able to use the same verbal categories – those you have selected earlier in the planning phase – to attach to your data as they travel into a database. You see, the information about the Intent and the Design of your Dataset, and the Station type, and all the other words you selected to describe what your data represent, are your "metadata". These metadata are essential to data interpretation and analysis.

5.0 Sources and Resources

This IP is an integral part of the Data Quality Management (DQM) System implemented by the Clean Water Team, the Citizen Monitoring Program of the California State Water Resources Control Board.

For an electronic copy, to find many more CWT guidance documents, or to find the contact information for the CWT Coordinators, visit our website at www.waterboards.ca.gov/nps/volunteer.html

If you wish to cite this IP in other texts you can use "CWT 2006" and reference it as follows: "Clean Water Team (CWT) 2006. Representativeness of Environmental Monitoring Data, DQM IP-8.2.4. in: The Clean Water Team Guidance Compendium for Watershed Monitoring and Assessment. Division of Water Quality, California State Water Resources Control Board (SWRCB), Sacramento, CA."

References and complementary Guidance materials

DQM-IP-8.2.3 - Good Data made tangible (in this compendium)
DQM-IP-8.2.5 - Sampling Design Principles (in this compendium)
Sycamore Creek Monitoring Project Plan (Example); see CWT toolbox.

SWAMP Field Methods Distance Learning Course, Common Element C (Representativeness)
SWAMP Advisor, an Expert System for generation of Quality Assurance Project Plans

U. S. Environmental Protection Agency (USEPA) 2000. Guidance for the Data Quality Objectives Process, EPA QA/G-4. USEPA publication EPA/600/R-96/055, Office of Environmental Information, Washington DC. August.