DQM Information Paper 8.2.5
Sampling Design Principles

By Revital Katznelson, Ph.D

1.0 About this Information Paper

(This section is essentially common to DQM Information Papers from the 7, 8, and 9 series. If you have seen it already, please skip to Section 2 below). This Information Paper 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" technical information on sampling design principles and methodology. If you are a Trainer or a Technical Leader of any monitoring project, this may help you be aware of your options when you select a sampling design for your monitoring objective.

Section 2 of this IP presents a "menu" schematic showing a list of sampling design options and the relationship between them as applied to commonly used citizen monitoring Projects. Next, Section 3 provides a description of the different principles underlying each design option. Section 4 provides practical tips and advice on study designs based on our cumulative experience (The Clean Water Team and others). This section is meant to be updated as we learn more. Finally, the "Sources & Resources" section (Section 5) provides a list of related DQM guidance documents well as references and website leads into further information.

2.0 Sampling Design Options

When you design a monitoring study you go through a number of planning steps in which you formulate your study question, decide what characteristics to measure, decide where and when to collect samples and how many, and identify the methods that will work for you. The "Sampling Design Principle" is what you apply when you select locations and times to monitor, and there are several options you have to choose from. Figure 8.2.5-1 shows the dichotomy of sampling design principles based on practical considerations, and defines what each one means.

In an ideal world, every sampling activity is planned carefully and designed deliberately to depict the environment in what is perceived as the most realistic way that's relevant to the question asked. However, reality and experience tell us that sometimes sampling or other environmental data collection happens without deliberate planning, and this does not mean that the information is useless. In fact, many scientific discoveries and identification of hot-spots were made via anecdotal observations. The important thing is to communicate the fact that it happened that way.

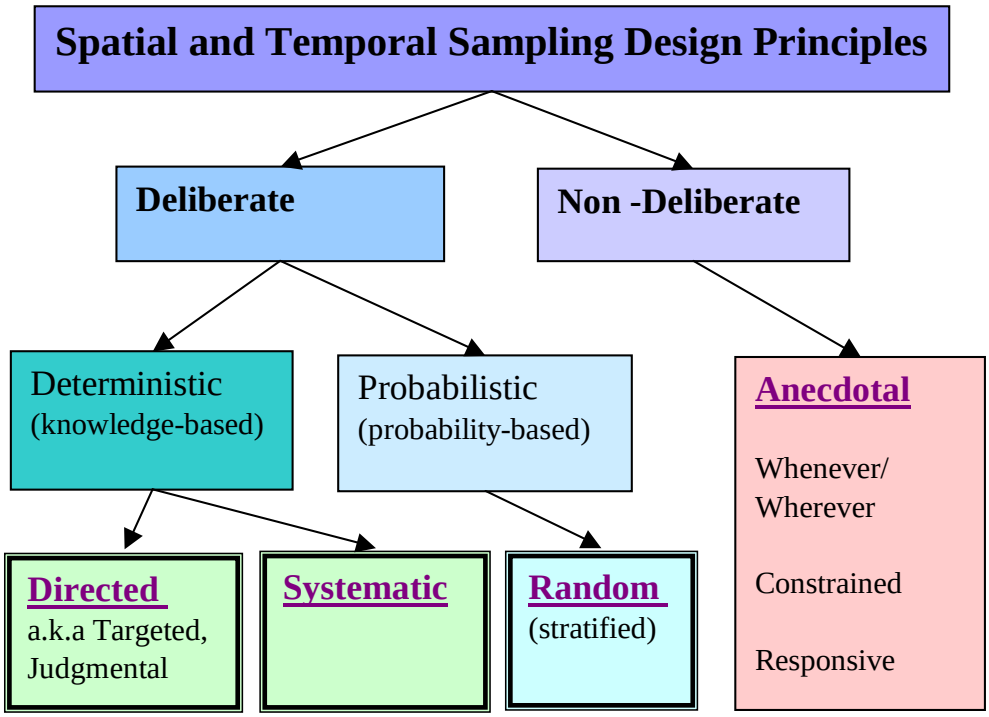

Figure 8.2.5-1: Lines of monitoring design principles

Definitions:

directed (to environment)	Deterministic approach, points selected deliberately based on knowledge of their attributes of interest as related to the environment monitored; also known as "targeted", "judgmental", "authoritative", "knowledge-based" etc.
systematic	Deterministic approach, points selected deliberately at fixed-intervals of area, length, or time
stratified **random**	Probabilistic approach, deliberate, points selected at random from a population stratified by specific attributes
non-deliberate (**anecdotal**)	Non-of-the-above, non-deliberate; points selected causally or whenever/wherever, or by given constraints, or opportunistically
deliberate to operations	Deterministic approach, points selected deliberately based on Project's operational requirements or constraints (rather than environmental representation)

Another common situation in which the sampling location is selected in a way that does not deliberately relate to its environmental representation may occur when there are constrains that dictate it (e.g., there is only one point with safe access along the entire stream reach), and Project planners do not have any choice in the matter. Monitoring in response to an event, e.g., spill, is also timed non-deliberately; however in some cases you will use the directed sampling design (see below) to choose the "where." The next section describes the principles and applications of deliberate sampling designs.

3.0 Principles and Applications of Various Sampling Designs

Sampling design is a set of rules and criteria that you apply when you select points (i.e., sample-units, or "chunks" of the environment) to analyze, measure, or observe. Because this is not about Science Fiction, i.e., you do not have access to the entire universe when doing environmental monitoring, your sampling design applies to a piece of the universe that is limited in space and time and delineates your "population" of possible points. Because you are not a multi-billionaire, you cannot do a census, i.e., analyze all the points in your population, either. You are left with selecting only a few out of all the choices in the universe.

Random sampling is when any point in your population has the same chance of being selected for sampling as any other point in the same population. There is no such thing as pure random sampling in environmental monitoring; usually you have to sort your entities (points) into groups of the same type in order to use the data you collect. For example, if you decide to collect a sample from the first thing you touch when you descend on Earth from the sky, you may end up with a jar of ice, a jar of soil, and a jar of feathers. Of course you can analyze for DDT in all of these environmental media, but you will not be able to compare this dataset with datasets generated for DDT in another combination of media (and in this line of reasoning, apples and oranges are much closer than you thought…).

What you can do is sort your entities into more homogeneous groups, e.g., water bodies (as opposed to landmass), freshwater (versus Seawater), lotic (flowing) system versus lentic systems (lakes and ponds), intermittent stream versus perennial streams, high-gradient stream reaches versus flat reaches, fruit versus vegetables, apples versus oranges. Or you can break it down into even smaller groups, e.g., Granny Smith apples versus Golden Delicious apples. Each of these groups represents a layer. Scientist use the term "stratified sampling design" ("stratum" means layer in Latin) when they refer to sampling a population that shares a certain feature, e.g., all the water bodies that are lakes.

Stratified random sampling designs are often used by National water monitoring programs to obtain knowledge about the general condition of waters of the United States, and in States in response to some of the Clean Water Act item 305(b) reporting requirements. They can also be used in smaller scales, for example, when you need to obtain an average number that represents an entire waterbody. There are variations in the way random sampling designs take into account variable probabilities, i.e., some points in the population are given more "weight" then other points, and you are encouraged to seek further guidance for these types of designs (e.g., USEPA 2002, QA/G-5S).

Directed sampling designs are very useful when you have a monitoring question that requires sampling at specific points in space and time, not just anywhere/anytime and not at random. Different guidance documents use different terms for the same thing: knowledge-based, judgmental, targeted, etc. There is one distinction made here that you might not see elsewhere, and it relates to "what do you direct your selection to?" The original definition relates to directing your choices to the conditions you are trying to capture in the environment, i.e., directed to environmental representation. For example, if you want to catch the lowest dissolved oxygen concentration in your stream you would deliberately visit it at dawn. This is captured in the "directed (to the environment)" category in the definitions under Figure 8.2.5-1.

However, selection of a certain time to conduct measurements is sometimes dictated by the logistics of field operations, rather than related to environmental conditions. Example: When you interpret results of lab analyses you often need to know what the conditions in the stream were like at the time the samples were collected. So, when you go out to collect water samples, and it may happen any time during the day (logistics dictate that you visit as many Stations as you can in one field day), you have to conduct field measurements at the same time. Thus your timing is deliberately directed to operations, because you do not address what you know about the diurnal variations on temperature and dissolved oxygen. This situation in included in the "deliberate to operations" category, and it is not a real sampling design principle.

Systematic sampling design is very useful when you select the time for routine monitoring. If you visit each of your stations at the same time of day on every trip, all data points will represent the same diurnal conditions and will provide information about seasonal trends. Furthermore, if you want to apply the systematic design to your routine monitoring in the same way that others are often using it, then do all your measurements between 9 AM and 11 AM. This will make your data comparable to their data sets because it represents the same time of day. You will not be capturing the worst case scenario, but that was not your intent anyway when you chose to do routine monitoring.

4.0 General Tips

Each of these principles applies to spatial scales and temporal scales, which are separated from each other and independent of each other. For example, you may apply the directed sampling design in selecting your season to look for algal blooms in a lake (i.e., select springtime), and may apply the random sampling design to select an array of sampling locations within the lake. However you will not be able to compare the average algal concentrations you get from random sampling with the concentrations you find, e.g., downwind at the surface, because you chose that location - using the directed sampling design - to capture the algal scum. In other words, comparability is possible only if the two datasets you are trying to compare were collected based on the same spatial and temporal design criteria.

The use of stratified random design to select a sampling location has been discussed in Section 3 above. On the temporal aspect, you might use the random design to deliberately show up at unexpected times (for example when you do your stream watch patrols).

The previous section also showed you how the directed, or targeted, sampling design principle is applied in timing the monitoring activity to capture the worst case scenario conditions. The same sampling design principle applies to station selection, for example when we monitor just downstream of an outfall with the intent of evaluating the impact of the discharge: it is deliberate, based on your knowledge, and directed to your intent.

Some study question require that more than one sampling design principle be applied at the spatial scale. For example, you use the directed sampling design to select a reach to do your benthic macroinvertebrates sampling for your bioassessments, then you find a population of "eligible" riffles in your reach, and select five of them using the stratified random sampling design. More than one design is also possible in the temporal scale, for example when you select the season in the directed sampling design, but within a given season you visit your Station systematically, e.g., every week on Saturday at 9 AM.

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. Sampling Design Principles, DQM IP-8.2.5. 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

SWAMP Field Methods Distance Learning Course, Common Element C (Representativeness)

U. S. Environmental Protection Agency (USEPA) 2002. Guidance on Choosing Sampling Design for Environmental Data Collection, EPA QA/G-5S. USEPA publication EPA/240/R-02/005, Office of Environmental Information, Washington DC. December.