Swanson Environmental Strategies

Thirty Years of Experience in Environmental Services

INTRODUCTION TO THE DATA QUALITY OBJECTIVES PROCESS

www.swansonenviro.ca

Swanson Environmental Strategies

What is the DQO Process?

The DQO Process is a systematic planning process for generating environmental data that will be sufficient for their intended use.

Swanson Environmental Strategies

What are DQOs?

DQOs are quantitative and qualitative criteria that:

- **Clarify study objectives**
  - **What must we accomplish?**
  - **What decisions will be supported by the data?**
- **Define appropriate types of data to collect**
  - **Chemistry, physical characteristics, biological parameters**
  - **How many samples? Where? How often?**
- **Specify the tolerable levels of potential decision errors**
  - **How sure do we need to be?**

Swanson Environmental Strategies

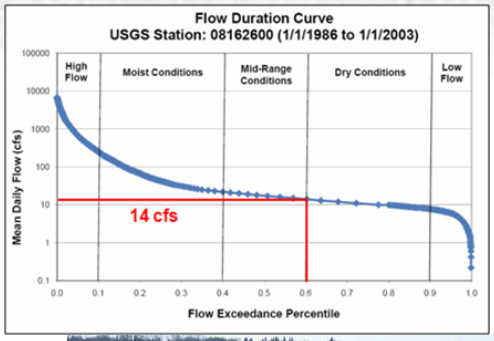
Advantages of the DQO Process

- **Planning Tool for Managing Decision Errors**
  - **What are the consequences of being wrong?**
  - **What is the acceptable likelihood of being wrong?**
- **Improves:**
  - **Planning Effectiveness**
  - **Design Efficiency**
  - **Defensibility of results/decisions**
- **Generates appropriate data**
  - **Type**
  - **Quality**
  - **Quantity**

Swanson Environmental Strategies

The DQO Process is Designed to Answer:

- **What do you need to know?**
- **Why do you need it?**
- **How will you use it?**
- **What is your tolerance for errors?**



www.swansonenviro.ca

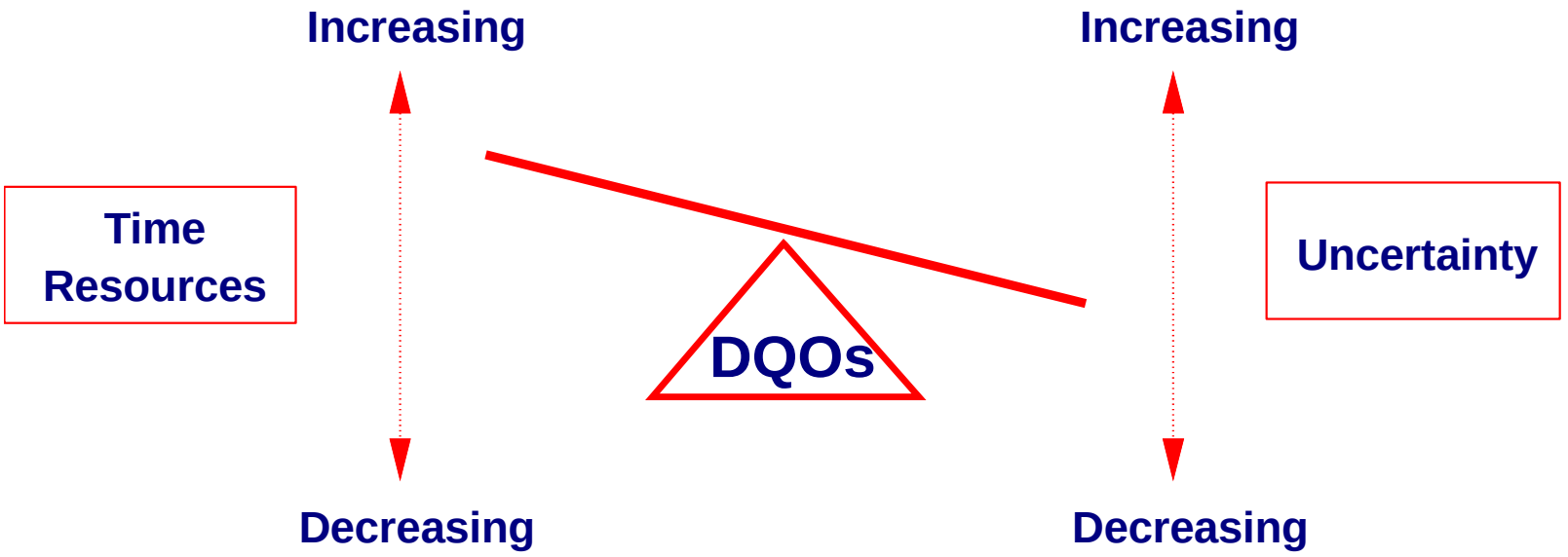
Swanson Environmental Strategies

Underlying Principles

1. All collected data have error.
2. Nobody can afford absolute certainty.
3. The DQO Process defines tolerable error rates.
4. Absent DQOs, decisions are uninformed.
5. Uninformed decisions tend to be conservative and expensive.

Swanson Environmental Strategies

DQOs Strike a Balance



Swanson Environmental Strategies

The DQO Process Encourages Efficient Planning

- **Clearly stated objectives**
- **A framework for organizing complex issues**
- **Limits on decision errors specified**
- **Efficient resource expenditure**

Swanson Environmental Strategies

THE SEVEN STEPS OF DQO

1. State the Problem

Define the problem, identify the planning team, examine budget, schedule

2. Identify the Goal of the Study

State how environmental data will be used in meeting objectives and solving the problem; identify study questions, define alternative outcomes

3. Identify Information Inputs

Identify data and information needed to answer study questions

4. Define the Boundaries of the Study

Specify the target population and characteristics of interest, define spatial and temporal limits, scale of inference

5. Develop the Analytic Approach

Define the parameter of interest, specify the type of inference, and develop the logic for drawing conclusions from findings

Decision making (hypothesis testing)

Estimation and other analytic approaches

6. Specify Performance or Acceptance Criteria

Specify probability limits for false rejection and false acceptance of hypotheses OR Develop performance criteria for new data or acceptance criteria for existing data

7. Develop the Plan for Obtaining Data

Select the resource-effective sampling and analysis plan that meets the performance criteria

Swanson Environmental Strategies

Step 1: Stating the Problem

- What is the problem?
- What resources are available?
- What time is available?
- What important social/political issues have an impact on the decision?

Swanson Environmental Strategies

Step 2: Identifying the Study Goal

- **Identify the principal study question(s).**
  - ***Exactly what must we know/understand in order to make confident decisions?***

- **Decide what needs to be measured.**
  - ***Associate a course of action with each possible answer***
  - ***What will we do if the data either exceed or do not exceed action levels?***

=

OR

Swanson Environmental Strategies

Step 3: Identifying Information Needs

- What do we already know? How reliable is that information?
- What do we need to measure?
- How well would different sampling and analysis methods perform?

Swanson Environmental Strategies

Step 4: Study Boundaries

- Over what geographic area will decisions apply?
  - E.g., will decisions be for individual operations or across the organization, specific streams, sub-watersheds, the entire watershed?
- Specify the time frame to which the study results apply and when sampling should occur.
  - How long a period should data be gathered before decisions are made? Should sampling be seasonal, monthly, annual?
- Identify practical constraints.

Swanson Environmental Strategies

Step 5: Analytic Approach

- **Will we be testing hypotheses?**
  - **(e.g. upstream vs downstream)**
- ***or*** **will we be comparing estimates?**
  - **such as mean area covered by calcite with derived action levels?]**

- **"Decision Problems" = hypothesis testing**
- **"Estimation Problems" = comparison of estimates with benchmarks**

Swanson Environmental Strategies

Decision or Estimation?

- **Setting tolerable decision errors for decision problems requires sufficient existing data for quantitative expression of uncertainty (standard error) followed by calculation of required sample size**

- **NOTE: if we don't have a lot of existing data and/or a good understanding of the situation we usually start with Estimation**
  - Estimates can be means, medians or percentiles

Swanson Environmental Strategies

Developing a Decision Rule/Analytical Approach

Develop an "if/then" statement that incorporates:

- **The population parameter of interest (e.g., mean, maximum, percentile)**
- **The scale of decision making (e.g., mainstem river, tributaries, off-channel ponds)**
- **The action-triggering value, if any**
- **The alternative actions**

Parameter	Example of Use
Mean	Comparison of mean to Action Level. NOTE: The arithmetic mean is greatly influenced by extremes in the distribution of data. Thus, for skewed distributions with long right tails, the geometric mean may be more relevant. Means are not useful if a large proportion of values are below the detection limit. Example: Mean seasonal nitrate concentration
Median	Better estimate of central tendency data that are highly skewed. Also may be preferred if there are many values that are less than the measurement detection limit. The median is not a good choice if more than 50% of the population is less than the detection limit because a true median does not exist in this case. Example: Median cadmium concentration in soil
Percentile	For cases where only a small portion of the population can be allowed to exceed the Action Level. Sometimes selected if the decision rule is for a stressor that has severe consequences. Also useful when a large part of the population contains values less than the detection limit. Often requires larger sample sizes than mean or median. Example: No greater than 25% of spawning habitat can be affected by sedimentation

Swanson Environmental Strategies

Principal Study Question

Estimation Problem

Decision Problem

Mean

Median

Percentile

Develop the If...then...statements

Estimate meets the desired confidence level and exceeds the action level

Statistical differences between reference and "mine-affected"

Mitigation Required

List of Alternative Actions

Mitigation Method Determined

Swanson Environmental Strategies

Step 6: Performance Criteria

Decision Problems

- Tolerable Decision Error for hypothesis testing
- Formal statistical tests (e.g. “upstream versus downstream”)
- Require preliminary data

Estimation Problems

- How confident you need to be in the estimate
- Compare the estimate to a benchmark ; e.g. the 90th upper confidence limit of natural background – if greater, then action required

Swanson Environmental Strategies

What is Tolerable Decision Error?

Tolerable decision error is the chance of either false acceptance or false rejection of the null hypothesis

Four Possible Outcomes of Hypothesis Testing		
Decision You Make by Applying the Statistical Hypothesis Test	**True Condition (Reality)**	
	There Really is *No* Difference	**There Really *is* a Difference**
Decide That the null hypothesis is true (e.g. conclude that there is no difference between upstream and downstream calcite)	Correct Decision	Type II Decision Error (False Acceptance). You have concluded there is no difference between upstream and downstream calcite when there actually is
Decide to reject the null hypothesis (e.g. conclude that there is a difference between upstream and downstream calcite)	Type I Decision Error (False Rejection). You have concluded there is a difference between upstream and downstream calcite when there actually is no difference.	Correct Decision

Swanson Environmental Strategies

Specifying Performance or Acceptance Criteria

- Determine the possible range of the parameter of interest
- Determine baseline condition
- Determine consequences of each decision error, which may include, for example:
  - Health risks
  - Ecological risks
  - Political risks
  - Social risks
  - Resource risks

Swanson Environmental Strategies

How is Tolerable Decision Error Defined?

- **The acceptable probability of either false acceptance or false rejection**
- **A common starting position is a 5% probability of false rejection and a 10% probability of false acceptance**
  - This assumes that the consequences of false rejection (e.g. $$$ spent on mitigation when it wasn't needed) are greater than the consequences of false acceptance (e.g. not mitigating when it should have been)
  - Regulators, industry, and other stakeholders may have different views on the severity of the consequences

Swanson Environmental Strategies

Tolerable Decision Error Determines Sampling Effort

- Achieving a 5% chance of a false rejection can require a LOT of sampling if your system is inherently highly variable
- You may not know how variable the system is yet if you haven't done enough preliminary sampling
- There will be trade-offs between how sure you want to be and how much effort and time it will take to be that sure

Swanson Environmental Strategies

Confidence in Estimates

- **Specify how sure you need to be in the estimate**
  - E.g. what standard error is acceptable for the mean seasonal nitrate concentration?
  - NOTE: the more uncertain the estimate, the less confident is the decision based on that estimate
    - How sure must you be that the mean nitrate coverage has exceeded the benchmark (and thus triggered mitigation action)?

Swanson Environmental Strategies

Required Confidence in Estimates Determines Sampling Effort

- **As with Tolerable Decision Error, the required confidence in estimates also drives the sampling effort**
- **In general, the "tighter" you need the estimate to be, the more sampling you need to do and over a longer period of time**
  - E.g. you decide that the upper 95th confidence limit must be no more than 50% higher than the mean total suspended solids concentration – this implies that you can't tolerate wider confidence limits because you require a reliable and "tight" understanding of the suspended solids situation

Swanson Environmental Strategies

Step 7 :Develop the Study Plan

- Develop general data collection design, such as:
  - Selection of representative stream segments
  - Timing and frequency of sampling
- Determine the sampling effort, number of samples, degree of replication (this will be driven by tolerable decision error and/or the acceptable uncertainty in estimates)
- Balance the sampling effort to meet decision error/uncertainty requirements with logistic constraints
- Be prepared to iterate – the first year is usually a scoping and learning exercise that can be used to refine our understanding of the inherent variability in the system we are monitoring

Swanson Environmental Strategies

Data Quality Assessment

- **Do the data meet the data quality objectives?**
  - Did we measure the right things at the right places at the right times?
  - Do we have enough data?
  - Do the data meet our acceptance or performance criteria?
- **Do the data support confident decisions?**
  - Is there sufficient evidence to draw conclusions?

Swanson Environmental Strategies

The DQO Process Provides Focus and Confidence

- **The data are matched with the question and have a specified level of tolerable error**

- **The only data collected are those required for the decision**