Tailoring of Data Quality Objectives to Specific Monitoring Questions

Revital Katznelson, Ph.D., Senior Project Scientist, Woodward Clyde International-Americas, 500 12th Street, Suite 200, Oakland, CA 94607. (510) 874 - 3048.

Abstract

The use of simple field methods can be a cost-effective way to obtain very useful data for watershed management, although they are not always supported with traditional quality assurance and quality control (QA/QC) practices. Data quality objectives for field methods can be tailored to specific monitoring questions, provided the issues of tolerable error, monitoring design, choice of operators, and cost are addressed. Monitoring of water quality in urban creeks in California is used as an illustration of a conceptual framework for tailoring of data quality. The first steps involve formulation of study questions, selection of parameters, and developing sampling design. Next, the tolerable error is defined, methods are selected, sources of error are identified, and operators are trained. In the example provided, the acceptable magnitude of error that can be tolerated depends on the ecological significance of each water quality parameter at the range of values that is critical for the health of the organisms in the creek. As to the choice of operators, the key is training. Any non-professional monitors (staff or volunteers) can and should be trained to identify the sources of error and uncertainty and to minimize both.

1. Introduction

Watershed information, including water quality data, is essential for guiding watershed management decisions. Agencies allocate considerable resources for the acquisition of data of the best quality, using sophisticated instruments and analytical methods; this approach is based on a widespread belief that scientific measurements always have to be very accurate and precise, and that sophisticated instruments which generate more digits and more decimal places make scientific measurements even better. Intensive quality assurance/quality control (QA/QC) procedures are applied, and strict data quality objectives are developed for data collection using these methods.

As young science students, we have been taught to replicate our measurements extensively in order to provide formal estimates of error in our measurements and to provide statistically-robust data sets for hypothesis testing. This discipline is extremely valuable for any person practicing scientific work. As environmental scientists, we are obliged to use standard methods with rigorous quality assurance procedures to obtain data that will be defensible in court. This is extremely important if the data are to be used for determining if a hazardous-waste site should be cleaned, or if an effluent is violating discharge permits that specify water quality criteria protective of aquatic life.

But because resources are never unlimited, in situations where representative data are more useful than precise data we may opt to do 8 inexpensive field tests with 25 % error instead of 2 expensive laboratory analyses with 5 % error. When we monitor the quality of water in our watershed to characterize its conditions, to understand the processes going on in it, or to track changes that occur in it, we can use methods that involve far less effort and expense than the

state-of-the-art methodology and still obtain data that are good enough for what agencies want or need to know. The issues all boil down to: **What range of error can we tolerate?**

Based on the concept that tolerable error is a function of the question we want to answer and the importance (ecological, regulatory, or economic) of the tested parameter, this paper is an assortment of thoughts, ideas, learning experience, and suggestions, to be shared with persons involved in monitoring as "food for thought". A conceptual framework is described and specific examples, focused on field methods but including some laboratory methods, are provided. The intuitive, informal language that has been developed by the author in numerous training sessions is used in this paper.

2. Definitions

2.1 Accuracy and precision

The USEPA QA/QC guidance (USEPA 1996) provides excellent explanation for the concepts of accuracy and precision, using intuitive "bull's eye" examples. Essentially, accuracy is a measure of how close we are to the absolute true value, and can be assured or evaluated by analysis of standards from different sources. Standards are also spiked into the tested sample matrix to provide information on matrix effects, i.e., how far from the truth can our results be due to sample-specific interference. Precision is a measure of how reproducible our measurements are, and can be evaluated by analysis of replicates from the same sample and/or by repeated analyses of the same sample at different times.

2.2 Sensitivity, resolution, and detection limit

When we talk about the resolution of a method, we make a statement about the smallest increment that the method can discern with confidence. Detection limit is the lowest value that the method can report as significantly positive (usually an indication that there is 95% probability that the value is indeed positive). There is often a confusion between the two concepts (i.e., resolution and detection limit) owing to the common use of the term "sensitivity" to describe both. For example, people refer to a method as sensitive if it can detect low concentrations of an analyte in a sample, or if it can show minute differences in concentrations between two samples.

2.3 Error and uncertainty

Many workers have suggested rigorous definitions of error and uncertainty, and rigorous methodology to quantify both (e.g., the strategy developed by the Intergovernmental Task Force on Monitoring (ITFM 1995)). This formal derivation is essential in many fields, including ecological risk assessment. However, as operators we intuitively use a practical distinction between error and uncertainty: error we can do something about, uncertainty we have to live with. Another practical distinction: error has to do with the quality of our measurements, uncertainty has to do with what they represent. But in practice we can identify sources of both, both can be quantified (intuitively or formally), and measures can be taken to diminish both; the level of effort has to do with the amount of error we can tolerate.

For the purpose of this framework, the term "error" tells us how far our measurement could be from the truth for that specific sample, either as percentage of the value (e.g., plus or minus ten percent, or $\pm$ 10%) or as an increment of the value (e.g., $\pm$ 0.3 pH units), depending on the method. This term encompasses the concepts of accuracy, precision, and resolution. It does **not**

deal with variability, representativeness, comparability, and other attributes that are associated with natural variability, study design, and choice of methods.

2.4 Data quality and reliability

People often confuse data quality with data reliability. They associate level of accuracy and precision with quality: high precision and accuracy is considered high quality, and therefore reliable, data. In reality, data are reliable if the values fall within the range of error specified for them, and it is much more difficult to obtain reliable data at high precision and accuracy (narrow range of error) than at a wide range of error.

3. Conceptual framework for developing data quality objectives

Monitoring is performed for a variety of reasons, in a variety of settings, to provide answers to a variety of questions. Table 1 demonstrates an assortment of "Ammonia questions" that may be encountered by an environmental scientist. It is apparent from Table 1 that study design, data quality objectives, and methodology can and should be tailored to questions for cost-effective provision of data.

An approach to the process of tailoring data quality objectives to specific questions may include the following steps which are described using a specific example below:

1. Formulating study questions
2. Selecting parameters and developing sampling design
3. Defining tolerable error for selected ranges of values
4. Determining the feasibility and cost-effectiveness of available methods
5. Exploring the sources of error and uncertainty associated with each method
6. Training operators to minimize error and uncertainty and to achieve data quality objectives

The first step may involve a complex process of formulating watershed management questions with inputs from stakeholders to define the data needed, or it could be a simple question such as "can fish survive in the creek?". The following steps assume that the latter is the question.

In the second step, we would examine what we know about the ecological requirements of fish in creeks, list the factors and parameters that are ecologically significant for fish survival, and focus on those that we think may pose problems, for example, dissolved oxygen (DO) depletion. The sampling design may accommodate both routine monitoring of DO (e.g., every two weeks at 9-11 am at three fixed stations in different creek segments) and worst-case scenario (e.g., DO measurements at dawn in the remaining stagnant pools at the end of summer). Sampling for biochemical oxygen demand (BOD) will help evaluate the potential for oxygen depletion. This step is iterative: the study design is periodically refined based on review of the findings and parameters are added or deleted.

The third step would examine the range of DO values that is critical for the health of the fish in the creek, and determine the amount of error that can be tolerated. For example, we can tolerate an error of plus or minus 1 mg/l DO around a value of 8 mg/l (because values in the range of 7 to 9 mg/l are "healthy"), but not around 3 mg/l (because the difference between 2 and 4 mg/l means life or death to many organisms.

The fourth step involves selection of method to measure DO. Table 2 lists the three major methodologies for DO measurements and describes the principles and challenges associated with each of them. Table 3 provides more specific information on cost and attainable data quality for available methods. For our fish in creek example, the field kit utilizing the modified Winkler method with direct titration can provide data of adequate quality.

In the fifth step, sources of error and uncertainty are explored. Table 4 provides a broad "checklist" for the various methods. For our example, error and uncertainty associated with the use of the modified Winkler field kits for DO measurement have to be examined. Experiments would show that contact with air during sampling will introduce oxygen into the sample; this will probably have negligible effects on the results if we are in the high range, i.e., close to oxygen saturation, but can introduce an error of 100% (elevate the measured concentration from 1 mg/l to 2 mg/l) at low DO values. Another source of potential error is the dispensing of a fixed sample into the titration vial: should the 20-ml volume line lie below the meniscus or above it? This difference of almost 2 ml accounts for 10% of the volume. It must be noted that evaluating the magnitude of error contributed by each source requires experimentation.

The sixth step is where the understanding of the sources of error and the magnitude of error they may introduce can be translated into action. Operators trained in the use of the modified Winkler field kits for DO measurement need to be made aware of the sources of error and develop a "feeling" of how confident they are with the data they report. They have to watch for bubbles in the sample bottle and start all over again if they see any. They have to be consistent (e.g., always have the bottom of the meniscus merge with the 20-ml line) and confirm each reading by having two people "read" the output. They have to repeat the titration if they think they overshot the endpoint. They have to duplicate measurements on a regular basis (and always repeat the sampling, fixing and titration if the result "doesn't make sense"). They need to test the performance of the kit (e.g., measure DO in saturated clean water at a given temperature) at regular intervals to account for gradual deterioration of the reagents with time, and anytime a reagent has been replaced.

4. Example: Creek monitoring in California

4.1 The questions, the parameters, and the study design

The study question in that case was: Are conditions in the creek suitable for year-round support of fish populations? The parameters selected were dissolved oxygen, temperature, pH, electrical conductivity, and turbidity. The initial monitoring/sampling design called for routine monitoring of these parameters twice each month during the wet season (November through May) and into the dry season. As the study progressed, it became apparent that flows during dry weather conditions were the critical factor, particularly after the winter flows had subsided and water remained only in a few upstream segments of the creek. Flow measurements had to be added to the list of parameters and the question was re-phrased : "Are condition in the creek during the worst time of year and worst time of day still supportive of fish survival?". Monitoring that describes the worst case scenario will provide an answer.

Experience shows that in order to describe the worst case scenario for a typical creek in the San Francisco Bay Area in California, water quality parameters that change during the 24 hour period (in response to solar radiation) need to be measured at the most critical time of day. Temperature and pH (also DO for supersaturation effects) will be most extreme during the early

afternoon, at about 14:00 or 15:00 summer time, any day of the week, during August - September. Temperature fluctuation can best be evaluated using an automatic data logger (hobo) that can be deployed in the creek for up to two months. The lowest DO values (for evaluation of depletion) are likely to be measured at dawn or early morning, any day of the week, during August-September when flow is minimal. Turbidity and electrical conductivity may not show diurnal cycles but do fluctuate in urban creeks. Atypical dry weather values of these parameters may be encountered during the weekends (more yard, garden, curbside activities in residential watersheds, and more human and dog access to the creek). Turbidity questions relevant to fisheries are: how often does the creek become turbid, and for how long. Drastic changes in conductivity, which could indicate illicit discharges to the creek, may cause osmotic stress. However, the values for all five water quality parameters are strongly dependent on flows, and flow should be evaluated any time measurements are made.

4.2 Tolerable error

Preliminary suggestions for tolerable error were set on the basis of ecological/physiological significance, as follows:

- Dissolved oxygen: a) an error of plus or minus 1 mg/l in the ranges of 0-3 mg/l and 8-10 mg/l (because values in the range of 0-4 mg/l are inadequate and values in the range of 7 to 11 mg/l are "physiologically comfortable"). b) an error of plus or minus 0.4 mg/l in the range of 5-7 mg/l (this is the critical zone for warm water fisheries and cold water fisheries).
- Temperature: an error of plus or minus 0.5oC
- pH: plus or minus 0.5 pH units in the ranges 1-6, 7-8, and 9-14 (two uncomfortable zones and one comfortable zone), and plus or minus 0.3 in the ranges of 6-7 and 8-9 (two zones of transition between comfortable and uncomfortable).
- Electrical conductivity: error up to 30% of the measurement
- Turbidity: error up to 50% of the measurement (because the fish couldn't care less if it is 50 JTU or 80 JTU, all they want to know is when this turbidity is going to go away).
- Flow: error of up to 100% of the estimate in the range of 1 to 30 gallons/minute, or up to 2 cubic feet per second (cfs). The relevant factor during dry weather is the detention time of water in an average pool, or how often the entire pool volume is replaced; information on turbulence immediately upstream of the pool is relevant as well.

4.3 Training of operators

Almost any person with minimal skills, be it an agency staff person, a volunteer, or a high school student, can be trained to use pH probes, conductivity meters, thermometers, turbidity kits, and even the modified Winkler field kits for DO measurement. But training has to go beyond the manufacturers instructions contained in the kit itself. The most important elements of training are about including the operators (those who are actually collecting the data in the field or in the lab) in the monitoring effort in a way that shares the understanding of the

objectives, promotes personal responsibility for the reliability and usefulness of the data they generate, and creates a sense of ownership and participation of/in an important process.

Training has to be conducted in phases that match the learning curve and the internalization of concepts and "feeling". We can start with imparting awareness and intuitive understanding of error and uncertainty. One way to do this is to ask the trainees what value they can "stick their neck out for" with confidence, e.g., "if you report 4 mg/l, could it also be 3 or 5, or are you sure that it cannot be less then 3.5 or more than 4.5?". That may elicit their curiosity about their own performance and they will try to find out for themselves, examining the sources of error and intuitively using all the quality assurance procedures you can teach them. In this way, the operators define the data quality objectives (DQOs) for themselves, based on their experience, and prove that they can achieve them (and if they cannot achieve the DQOs tailored to the question, another method needs to be examined). And - if personal responsibility for data reliability is promoted - operators need to be assured that it is sometimes OK to leave a blank space in a data sheet ("if you cannot do it right, don't do it") because no data is much, much better than wrong data, and that their honest "explanation notes" are their way of communicating their experience with the data user.

Training has to remind the operators to use eyes, brain, and common sense in everything they do. It has to teach and encourage operators to keep neat records, to pay close attention and be consistent when dispensing volumes of liquids, to avoid contamination, to wait for instrument readings to stabilize, to confirm each reading by having two people "read" the output, to repeat measurement on a regular basis, to question if the result "makes sense", and to repeat measurements if it does not. It also has to reinforce the awareness of measurement "drift", as instruments move away from calibration and reagents change their reactivity over time, and encourage the operators to test the performance of kits and instruments at regular intervals to account for gradual deterioration of the reagents with time, and anytime a reagent has been replaced.

Another important concept related to training for field work is that everything that applies to good laboratory practices and to quality assurance is also valid in the field. This includes behavior/attitude issues ("Quick and dirty" does not mean that we can tolerate contamination, and sloppy work is not acceptable) as well as methodology (e.g., a run of the colorimetric salicilate field kit for ammonia with multiple test tubes, to accommodate a few samples, reagent blank, and a calibration curve of three standards). The concept that some samples may need to be re-analyzed, diluted in distilled water to fit a range of color intensity increments that the human eye can perceive, also needs to be taught (and the equipment to make dilution needs to be provided in the field kit).

Standard Operating Procedures (SOPs) may be used as a reminder but cannot replace training by a person. At a later phase, trainees can be introduced to the more "traditional" QA/QC programs and familiarize themselves with the formal elements of a QA/QC plan. However, the most elaborate QA/QC plan will not assure data quality and reliability if the operators had not been properly trained. It is highly recommended to construct monitoring programs in ways that allow direct contact between the data collectors and the QA/QC officers (rather than have them several ranks removed), and to allow periodical communication between the data collectors and the data users.

Acknowledgments

This article summarizes experience that has been acquired in numerous projects sponsored by The Hebrew University of Jerusalem, Israel, Mekorot Water Company, Israel, the Alameda Countywide Clean Water Program, Hayward, CA, and the City of Newark, CA. Preparation or the article was supported in part by the Alameda County Flood Control and Water Conservation District. The author wishes to thank Dr. Marion Lamb, Dr. Peter Mangarella, and Ms. Arleen Feng for their helpful review comments.

References

Abeliovitch, A, and Y. Azov (1976) Toxicity of ammonia to algae in sewage oxidation ponds. Applied Environ. Microbiol. 31:801-806

Intergovernmental Task Force on Monitoring Water Quality (ITFM). 1995. The strategy for improving water-quality monitoring in the United States: Final report of the Intergovernmental Task Force on Monitoring Water Quality. U.S. Geological Survey, Reston, Virginia.

Scheiner, D. (1976) Water Research 10:31-36.

U. S. Environmental Protection Agency (USEPA) 1996. The volunteer monitor's guide to quality assurance project plans. Office of Wetlands, Oceans and Watersheds, Washington, D.C. EPA 841-D-96-001.

Woodward-Clyde Consultants (WCC). 1997. Summer ecological study of Lakeshore Park, Newark. Report prepared for the City of Newark, CA, April 1997.

Paper source

Katznelson, R. (1998). Tailoring of Data Quality Objectives to Specific Monitoring Questions. in: Proceedings of the First National Monitoring Conference of the National Water Quality Monitoring Council, "Monitoring: Critical Foundations to Protecting Our Waters", July 7-9, 1998, Reno, NV.

Table 1: Questions Related to Ammonia in Various Environmental Settings

	Scenario or Setting	Question	Significance	Design	Method[1] and Cost[2]	Data Quality Objectives[1]	Supporting Information
1	Illicit discharges to urban creeks	Is ammonia discharged into the creek during dry weather flows?	Illicit connections, broken sanitary sewers malfunction of septic systems	anytime, anywhere	Nessler reagent, one-tube, 2 minutes test kit $0.08+$2 / sample	DL 1 mg/l, Error ± 1 mg/l	City sewerage maps
2	Ammonia detected in a creek	What is the source of ammonia?	Source ID	different watershed locations, same time	Colorimetric 2-reagents, 15 min, test kit (e.g., salicilate), 2 sample dilutions + blank + standard $0.5+$10 / sample	DL 0.2 mg/l Error ± 0.2 mg/l	Watershed activities and storm drain maps
3	Ammonia suspected/detected in fisheries habitats (creeks, rivers, ponds, lakes)	Is it toxic to fish?	Above 5 mg/l at pH above 8 may be a problem	pulses in rivers, vertical gradients in lakes	Colorimetric (e.g., salicilate), test kit $0.5+$10 / sample	DL 1 mg/l Error ± 0.5 mg/l	pH (± 0.3 units) Temperature (± 0.5^{o}C)
4	Intensive aquaculture tanks	What is the level of ammonia in the tank at any given point in time? Does ammonia accumulate to toxic levels?	Unionized ammonia above 0.2 mg/l may be a problem	continuous monitoring	Ion specific electrode, on line, with recorder. Cost varies.	DL 0.1 mg/l unionized NH_3 Error ± 0.1 mg/l NH_3	Fish density, feeding rate
5	Algal populations in aquatic systems	Is ammonia nitrogen available?	Low importance (Even if not available, nitrogen is usually not the limiting nutrient)	pulses, gradients, time-course	Laboratory method (e.g., indophenol[3]) with concentration step $50 / sample	DL 0.005 mg/l	Other N sources, P, presence of nitrogen fixer

[1] Total ammonia (free ammonia, NH_3, plus ammonium ion NH_4^+) is measured unless otherwise specified.
[2] Cost is approximate and includes reagents plus labor (or both combined).
[3] Scheiner 1976
[4] Abeliovitch and Azov 1976
[5] WCC 1997

Table 1 (cont.): Questions Related to Ammonia in Various Environmental Settings

	Scenario or Setting	Question	Significance	Design	Method[1] and Cost[2]	Data Quality Objectives[1]	Supporting Information
6	Urban or agricultural storm runoff	How much ammonia is contributed by a given watershed?	Loads assessments	flow-weighed storm composite	Colorimetric, (e.g., indophenol[3]), absorbance, calibration curve $30 / sample	DL 0.05 mg/l, Error ± 0.05 mg/l	Storm event data
7	Total ammonia in sediments	Sink or source	Nitrogen budget,	flux, gradients	Laboratory, distillation $50 / sample	DL1 mg/kg, Error ± 1 mg/kg	TOC
8	Ammonia in sediment pore water	What proportion of the ammonia is free, i.e., extractable in elutriate?	Potential toxicity resulting from resuspension or dredge material disposal	different locations, times	colorimetric, (e.g., salicilate), extraction with H2O for free, with KCl for total (freshwater sediments[5]) $1+$20 / sample	DL 0.5 mg/l in extract, Error ± 0.5 mg/l	Sediment and lake volume
9	Nitrification in sewage treatment plants	What is the rate of ammonia removal?	Effectiveness of nitrogen transformation processes	different locations	Colorimetric, (e.g., indophenol[3]), absorbance, calibration curve $30 / sample	DL 0.5 mg/l, Error ± 0.2 mg/l	loads, sludge age, other process info
10	Photosynthetic activity in oxidation ponds operating under high organic loads	Are the ammonia concentrations high enough to inhibit photosynthesis?	30-40 mg/l ammonia may inhibit Chlorella photosynthetic (oxygen-producing) activity[4]	different times of day, photic zone	Nessler test kit, sample dilutions $0.16+$10 / sample	DL 5 mg/l, Error ± 2 mg/l	chlorophyll a

[1] Total ammonia (free ammonia, NH_3, plus ammonium ion NH_4^+) is measured unless otherwise specified.
[2] Cost is approximate and includes reagents plus labor (or both combined).
[3] Scheiner 1976
[4] Abeliovitch and Azov 1976
[5] WCC 1997

Table 2: Available Methods for Monitoring of Dissolved Oxygen

Group	Measurement Principle	Challenge	Application
Polarographic	Electrodes measure the flux of oxygen across a membrane	Keep flushing sample liquid at the membrane surface to constantly replace the oxygen consumed by the electrode. (Rapid-Pulse and microelectrodes exempted)	Measurement of DO along gradients or transects where many samples are needed in a short time, measurements of kinetics of change in DO concentrations, continuous monitoring of DO (automatic data-logging Rapid-Pulse probes), micro-scale DO gradients on sediment surface (microelectrodes).
Colorimetric	Chemical reagents added in excess interact with oxygen to form a colored product (that absorbs light at a visible wavelength). Color is proportional to oxygen concentration.	Collect samples and introduce reagents without contact with air	Screening for anoxic conditions, rough and rapid measurements by non-professional operators, etc.
Titrimetric	Chemical reagents in excess interact with oxygen to form a product, and another chemical (the "titrant") is used quantitatively to "neutralize" that product. The amount of titrant needed is proportional to oxygen concentration.	Collect samples and introduce reagents without contact with air	Routine monitoring of DO in creeks, biochemical oxygen demand (BOD) measurements, etc. Laboratory applications: DO electrode calibration, BOD, etc. Samples can be collected and fixed in the field, and titrated later in the lab using high-precision burettes.

Table 3: Properties of Oxygen Measurement Devices

Principle	Device	Error	Instrument/ kit cost	Cost per sample	Work
Polarographic	DO meter+electrode	±5%	$800	$0.10	prep/calib 1 h measure 0.5-3 min.
	Rapid-Pulse probe, for Sonde	±5%	~$10,000 for entire Sonde	$0.10	prep/calib 2 h download 1 hr
	Specialty, e.g., microelectrode	varies	$1,000-$5,000	NA	prep/calib 2 h
Colorimetric	Reagent ampoules and comparator (e.g., "CHEMets")	±2 mg/l	$20	$0.50	measure 2 min.
	Reagent ampoules and colorimeter or spectrophotometer (e.g., "Vacu-Vials")	±1 mg/l in the 0-10 range, or ±0.2 mg/l in the 0-2 range.	$610	$0.50	measure 2 min.
Titrimetric	BOD bottle, reagents for fixing DO, vial, indicator solution, titrant solution, and syringe for titration	±0.4 mg/l	$40	$0.20	measure 5 min.

Table 4: Sources of Error and Uncertainty in Oxygen Measurements

Sources of error and uncertainty	Polaro-graphic	Colorimetric Visual comparison	Colorimetric Absorbance Measurement	Titrimetric
Electrode not assembled properly (e.g., air bubble trapped under membrane)	X			
Electrolyte too weak	X			
Instrument not calibrated correctly	X		X	
Membrane not under equilibrium at recording time	X			
Sample has been in contact with atmospheric air during collection		X	X	X
Constituents in the sample interfere with chemical reagents		XX	XX	XX
Pigments in sample interfere with color absorbence measurements		XX	XX	
Particles in sample interfere with color absorbence measurements		XX	XX	
Color intensity keeps changing as a function of time and temperature		X	X	
Human eye cannot distinguish small color increments; human eyes are not subjective		XX		
Dispensing of volumes is not accurate				X
Titration endpoint is not clear-cut, blue color of indicator reappears after a while.				XX
Titration is performed too fast or too slow				X
Sample bottle and/or other kit utensils are contaminated with titrant, reagents, and/or other interfering substances		X	X	X
Reagents and/or titrant are not reacting as specified		X	X	X

X - error/uncertainty can be diminished or controlled by operator (better training, more attention, more patience, fresh reagents)
XX - error/uncertainty due to nature of sample or operator and cannot be reduced.