QUALITY CONTROL FOR REGULATORS AND CONSULTANTS: LABORATORY METHODS

William Miller[1] and David Wenner[2]

AUTHORS: [1]Department of Crop and Soil Science and [2]Department of Geology, University of Georgia, Athens, GA 30602.
REFERENCE: Proceedings of the 2001 Georgia Water Resources Conference, held March 26-27, 2001, at the University of Georgia. Kathryn J. Hatcher, editor, Institute of Ecology, the University of Georgia, Athens, Georgia.

Abstract: Measurements made as part of environmental assessment and monitoring activities are subject to both random and systematic errors (bias) that can reduce data quality and influence sound project conclusions. Federal quality assurance standards are seldom applied to smaller state and private environmental projects. Many of the potential errors in such projects arise from poor quality control (QC) during sample preparation and analysis in the lab, and from failure of project managers to request and evaluate QC data. Basic sample set preparation can detect the presence of systematic error, and can be used to quantify the level of random error in a set of measurements. Recommendations are given for types of QC samples to include with data sets, and kinds of information to request from in-house or contract analytical laboratories.

INTRODUCTION

Environmental management, assessment and regulation depend to a large degree on reliable analytical data for environmental samples. Natural waters, waste streams, soils and sediments all must be analyzed for constituents of environmental interest (the *analtye*), either in the field or at off-site analytical facilities.

All analyses are subject to error; error results in uncertainty in the analytical value, relative to some hypothetical "true value" for a given analyte in a given sample. This true value in environmental samples is never known, and is largely unknowable. Thus, the analytical process must contain safeguards to ensure that errors are minimized so that the measured value for an analyte approaches as closely as possible this true value. Quality control (QC) in this paper refers to a range of laboratory procedures that aim at minimizing or eliminating errors, thereby resulting in analytical results which approach, as closely as possible, the "true" value. In environmental studies conducted or funded by Federal agencies such as EPA, extensive quality asssurance project plans (QAPP) must be filed describing sampling, sample handling, data management and reporting, as well as laboratory-oriented QC measures (USEPA, 1998). Most states also have QAPP guidelines in place, although this is not true of Georgia. In this paper we will focus on the *analytical* phase of overall QA plans, mostly with reference to non-Federal projects where QAPPs are not in place, and often where QA/QC is ill-defined.

THE ANALYTICAL PROCESS

The process of data collection in an experimental context includes a number of steps (Fig. 1). A complete QAPP specifies procedures and remedial action for each step in this process. For small-scale exploratory projects with limited objectives, full documentation of each step may not be cost-effective or warranted. Planning and sampling are usually handled by project personnel on-site, but often sample preparation, analysis, and data *quality* interpretation are done by off-site or in-house labs. Often such labs do not provide explicit QC checks without specific requests from field personnel. Project managers may all too often accept such data at face value, but without proper quality checks, error can seriously compromise the conclusions drawn from such data.

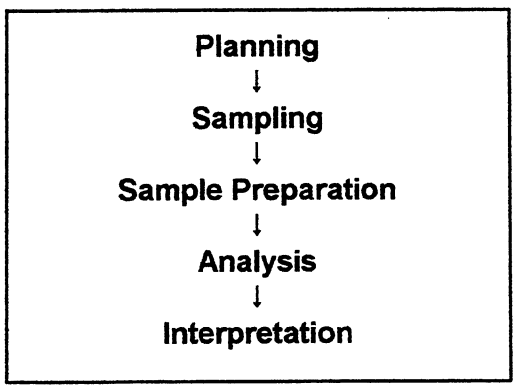

Figure 1. Component steps in the data collection process.

TYPES OF ERROR

Broadly speaking, error can be divided into two types: systematic error (bias) that affects each sample in a similar (erroneous) fashion, and random error that causes deviation of a measured value in a random fashion. This is illustrated in Fig. 2 using the analogy of the bulls eye as the "true" value, and individual "shots" as measured values. Random error results from sources such as contaminated glassware, mis-dilution of a sample, a random recording error, or electronic fluctuations in an analytical instrument. It affects each sample differently, either positively or negatively (relative to some "true" value), and and can easily be seen to cause an increase in the standard deviation (SD) of replicate measurements in Fig. 2. *Precision* refers to the level of random error in a data set, more precise data having a lower SD. On the other hand, systematic error (*bias*) skews all measured values away from the true value more-or-less equally. This may be due to poor instrument calibration, mis-dilution of *all* samples in a run, or a serious error in sample preparation (systematic loss of analyte, or uniform contamination). *Accuracy* refers to data that approaches the true value to within some specified guidelines. Note that random error cannot be eliminated from analysis, because we cannot hope to control all the random effects that create it; it can only be minimized, through good laboratory QC, and must be quantified so that we know something about the uncertainty in our data (i.e., the magnitude of the SD). Systematic error can largely be eliminated through good QC, but if not, must also be documented and quantified (Taylor, 1987).

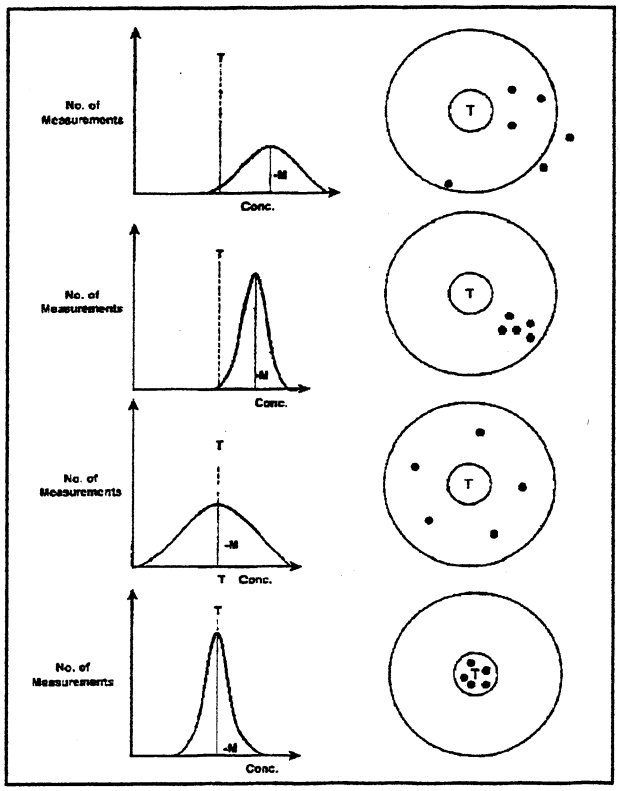

Figure 2. Precision and accuracy: precise data is tightly grouped with low standard deviation, while for accurate data the mean value approaches the true value (after Keith, 1991)

CONTROL OF ERROR

Quality control is a set of procedures attempting to minimize error during analysis, thus resulting in a measured value that approaches the true sample value for an analtye. In the lab, random and systematic error must be dealt with more-or-less individually.

Random Error

Replication is the only way to quantify random error. The multiple "shots" in Fig. 2 are replicates of the same "sample", resulting in a mean (M) and SD. Often a *coefficient of variation* (CV) is computed as

$$CV = [SD / M] \times 100$$

which expresses the SD as a percentage of M. Note that at least three replicates are needed to compute a SD. We would like this value to be low, say 5 or 10%, but in some types of replication it may be higher. *Field replicates*, sampled simultaneously side-by-side in the field, may show greater CV's due to natural sample variation (this is especially true for soils and sediments, much less so for waters). *Sample preparation replicates* will have less variation, but will have random errors in weighing, sample extraction or digestion, dilution, and other sources. *Analytical replicates* are multiple analyses of the same sample extract, and should show the least variation. Sample preparation replicates are the most useful in examining overall lab error without confounding the situation with natural variation, which can be evaluated separately.

It should be noted that water samples, inherently more homogeneous and requiring less sample prep (maybe filtering only), should have lower CV's than soils or sediments, where samples must be well-mixed, subsampled, and extracted or digested prior to actual analysis. It is also true that some types of analysis are inherently more variable: metals analysis via atomic absorption (AA) or inductively coupled plasma (ICP) methods are typically less

variable than analysis of organics by chromatographic methods. Also note that CV's always increase as analyte levels decline towards the method detection limit, since M becomes smaller in relation to SD. Goals for acceptable CVs of analytical reps, typically 10-15% at high analyte levels, may have to be increased as analyte levels approach detection limits (see below).

Splits or *duplicates* (two replicate measurements) are often performed, but show little other than something is wrong if the two values do not agree within some limit. *Triplicate* measurements at least give a SD that can be quantitatively evaluated (Day and Underwood, 1974).

Systematic Error

To determine accuracy a lab must have a sample similar in composition and analyte level to the one that is being analyzed, with a *known* value of that analyte. Thus a *% recovery* (%R) can be computed as

$$\%R = [M / T] \times 100$$

where M is the mean measured value and T the true value. In theory we never know the true value, but samples that have been extensively tested for given analytes using multiple methods are available. The National Institute of Standards and Testing (NIST) sells rock, soil, sediment, and water samples having certified values for given analytes; these materials, known as standard reference materials (SRM), can be processed through a lab's sample prep and analysis procedures to assess accuracy based on %R (NIST, 2001). While these materials are expensive, they are the best way to evaluate accuracy where complex sample preparation always raises the question of analyte loss or recovery (metals digests, organic extractions).

Individual labs often keep *control samples* that are continually re-analyzed to check on consistency (really, precision) of analysis over time. If, however, such samples are exchanged between different labs using various sample prep or analytical methods, and if a mean or consensus value for an analyte is agreed upon, then accuracy can be evaluated using such samples (Keith, 1991). A more rigorous version of this method is used by NIST to arrive at their certified values (NIST, 2001).

Spiking

Spiking refers to addition of a known level of analyte to a sample prior to analysis. A sample or extract is split, one part is spiked with analyte at roughly the level expected in the original sample, then both this spike sample and the original are analyzed. A %R can then be computed on this spiked amount, to determine how much of it was recovered by the analysis (see Csuros, 1997, or similar for examples of calculations). Spiking of analyte or a chemically similar compound (*surrogate spiking*) is often used in organic analysis as an indication of accuracy. Spiking may be done prior to sample preparation in order to test extraction or digestion recovery as well as analytical performance. For water samples spiking may be valuable in assessing accuracy in the lab or field, where complex sample preparation is not an issue (Wenner and Miller, 2001) For metals in soil materials, however, analytes are typically tightly bound and must be extracted, and therefore adding spikes of soluble metals does not give much information about recoveries of metals.

Detection Limits

Most laboratories report detection limits (DL) for analytes, but not in uniformly understandable ways. All DL's are based on variation in instrument readings of a blank or standard solution over time; this variation indicates if an analyte *signal* can be distinguished from the background *noise*.

Instrument detection limits (IDL) represent the lowest level of analyte that can be detected; it is computed using the SD of multiple (usually 7) blank (deionized water or solvent) determinations, re-computed to concentration units (parts per million [ppm] or billion [ppb]). The IDL only suggests the presence of analyte, but not its amount. *Method detection limits* (MDL) are computed using the SD of multiple standard determinations, and represent the smallest *quantifiable* level of analyte. If sample prep or dilution was conducted prior to analysis, then MDL should be adjusted to show this analyte amount *in the original sample*. Thus if 1 ppb is the MDL of an aqueous solution, but that solution represents an extract of a soil amounting to 100 times dilution (e.g., 1 g soil extracted with 100 mL solution), then the MDL on a soil basis is 1 ppb × 100 or 100 ppb (0.1 ppm).

Some environmental regulations are based on IDLs, but most regulatory levels are now above common MDLs for most analytes. Thus samples falling between IDL and MDL values are often simply reported as "BDL" (below detection limit), where the DL = MDL. Statistically BDLs are difficult to handle: the value is not zero, and some data managers will enter MDL/2 or MDL/10 as the sample value. This topic is currently hotly debated (see USEPA, 1998).

RECOMMENDATIONS

All project managers, whether state employees or private companies, are ultimately responsible for the data

quality in their reports. Only data free from serious errors should be used to make environmental decisions affecting the health of the public and the environment. All data sets requires estimates of accuracy and precision. Some analytical labs will provide documentation of QC procedures used in their labs *in general*, or actual QC results obtained during the analysis of an actual submitted sample set. Others will volunteer little such information, as QC clearly increases costs in a laboratory.

The following recommendations should be considered by those submitting samples to in-house or commercial analytical laboratories for environmental analyses:

1) *Ask what QC data is routinely reported:* Typically DL and information about instrument calibration is routinely given on data reports, but some labs may offer SRM analyses, spiking, or replicate analysis at no or low cost, depending on the size of the sample set. For small (<20) sample sets, expect to pay extra for QC data, but it is clearly worth the extra cost.

2) *Ask about certification*: Analytical laboratories can be certified by USPEA, state agencies, and private organizations to conduct environmental analyses. Such certifications require the lab adhere to certain *good laboratory practices* (GLP) such as record-keeping, sample handling and lab cleanliness, as well as perform analyses on blind samples sent by the certifying organizations. Such certifications are valuable in choosing laboratories, but certainly do not guarantee that a given data set is of high quality.

3) *Specify or include QC samples in the sample set:* Ask for triplicate analysis on a fixed number (i.e., 5) or certain fraction (10%) of the sample set as an estimate of precision. Alternately, split a well-mixed sample into three parts and submit each separately. Typical EPA practice is to duplicate every 10^{th} sample. Also inquire about (or purchase) SRMs that can be submitted along with the sample set. Often replicates and SRM's can be included "blind" in a sequentially numbered sample set, thus excluding any bias on the part of the analyst. You may also ask for certain other data to be reported, such as re-calibration data collected during the analytical run, or blank values performed during the analysis.

4) *Submit to multiple labs*: While this is an inconvenience, it is an excellent overall accuracy check, given that labs often use varying sample prep and analytical instruments. Submit a smaller subset of samples to an alternate lab, and compare to your main lab. Include some replication in each set to test precision. Don't expect values to agree too closely (±20-30% is good), but consistent, large differences indicate problems.

5) *Make them get it right*: If a problem in the data is apparent in the QC results, ask the lab for an explanation. Some analyses are very difficult and inherently noisy, especially at low levels. If you think they can do better, ask them to repeat some analyses. They should do this at no charge– you are the customer, after all. While it is probably impractical to expect every value in a large dataset of agree exactly, it is ultimately the data user, not the data generator, who has the most at stake, and the most responsibility, for the quality of the data and the conclusions drawn from it.

LITERATURE CITED

Csuros, M. 1997. Environmental sampling and analysis: A lab manual. Lewis Publishers, Boca Raton, FL.

Day, R.A., and A.L. Underwood. 1974. Quantitative analysis. Prentice-Hall, Englewood Cliffs, NJ

Keith, L. H. 1991. Environmental sampling and analysis: A practical guide. Lewis Publishers, Chelsea, MI.

NIST (National Institute of Standards and Testing). Standard reference materials program. Available at http://ts.nist.gov/ts/htdocs/230/232/232.htm. Checked February 2001.

Taylor, J.T. 1987. Quality assurance of chemical measurements. Lewis Publishers, Boca Raton, FL.

USEPA. 1998. EPA guidance for quality assurance project plans. EPA QA/G-5. Office of Research and Development, Washington DC.

Wenner, D. and W. Miller. 2001. Some recommendations for chemical stream monitoring used by citizen groups. Proc. 2001 Ga. Water Resources Conf., 26 March 2001; Kathryn Hatch, Ed. Univ. of Georgia, Athens, GA.

ACKNOWLEDGMENTS

We would like to thank the students of CRSS 4660/6660 and GEOL 4670/6670 who, over the past several years, have demonstrated how to make almost every possible analytical error (but eventually got it right). Thanks also to Dr. Sayed Hassan and the staff at the Lab for Environmental Analysis for their analytical help and support.