Quality Assurance
Good Practice, Regulation, and Law

QUANTIFYING UNCERTAINTY: CALCULATING INTERVAL ESTIMATES USING QUALITY CONTROL RESULTS

Judith A. Schofield [a]; Ken Miller [a]; Louis Blume [b]
[a] DynCorp, Systems and Solutions LLC, Alexandria, Virginia, USA.
[b] U.S. Environmental Protection Agency, Great Lakes National Program Office, Chicago, Illinois, USA.

PLEASE SCROLL DOWN FOR ARTICLE

QUANTIFYING UNCERTAINTY: CALCULATING INTERVAL ESTIMATES USING QUALITY CONTROL RESULTS

Judith A. Schofield and Ken Miller

DynCorp, Systems and Solutions LLC, Alexandria, Virginia, USA

Louis Blume

U.S. Environmental Protection Agency, Great Lakes National Program Office,
Chicago, Illinois, USA

EPA's Great Lakes National Program Office (GLNPO) is leading one of the most extensive studies of a lake ecosystem ever undertaken. The Lake Michigan Mass Balance Study (LMMB Study) is a coordinated effort among state, federal, and academic scientists to monitor tributary and atmospheric pollutant loads, develop source inventories of toxic substances, and evaluate the fate and effects of these pollutants in Lake Michigan. A key objective of the LMMB Study is to construct a mass balance model for several important contaminants in the environment: PCBs, atrazine, mercury, and trans-nonachlor. The mathematical mass balance models will provide a state-of-the-art tool for evaluating management scenarios and options for control of toxics in Lake Michigan.

At the outset of the LMMB Study, managers recognized that the data gathered and the model developed from the study would be used extensively by data users responsible for making environmental, economic, and policy decisions. Environmental measurements are never true values and always contain some level of uncertainty. Decision makers, therefore, must recognize and be sufficiently comfortable with the uncertainty associated with data on which their decisions are based. The quality of data gathered in the LMMB was defined, controlled, and assessed through a variety of quality assurance (QA) activities, including QA program planning, development of QA project plans, implementation of a QA workgroup, training, data verification, and implementation of a standardized data reporting format. As part of this QA program, GLNPO has been developing quantitative assessments that define data quality at the data set level. GLNPO also is developing approaches to derive estimated concentration ranges (interval estimates) for specific field sample results (single study results) based on uncertainty. The interval estimates must be used with consideration to their derivation and the types of variability that are and are not included in the interval.

Address correspondence to Judith A. Schofield, DynCorp, Systems and Solutions, LLC, 6101 Stevenson Avenue, Alexandria, VA 22304. E-mail: judy.schofield@dyncorp.com

The Great Lakes, which contain 20% of the world's freshwater, are a globally important natural resource that are currently threatened by multiple stressors. While significant progress has been made to improve the quality of the lakes, pollutant loads from point, non-point, atmospheric, and legacy sources continue to impair ecosystem functions and limit the attainability of designated uses of these resources. The U.S. EPA's Great Lakes National Program Office (GLNPO) instituted the Lake Michigan Mass Balance Study (LMMB Study) to measure and model the concentrations of representative pollutants within important compartments of the Lake Michigan ecosystem (U.S. Environmental Protection Agency, 1997). The LMMB Study was a coordinated effort among Federal, State, and academic scientists to monitor tributary and atmospheric pollutant loads, develop source inventories of toxic substances, and evaluate the fates and effects of these pollutants in Lake Michigan. A key objective of the LMMB Study is to construct a mass balance model for several important contaminants in the environment: PCBs, atrazine, mercury, and transnonachlor. The mathematical mass balance models will provide a state-of-the-art tool for evaluating management scenarios and options for control of toxics in Lake Michigan.

At the outset of the LMMB Study, managers recognized that the data gathered and the model developed from the study would be used extensively by data users responsible for making environmental, economic, and policy decisions. Environmental measurements are never true values and always contain some level of uncertainty. To address these issues, GLNPO employed a comprehensive suite of tools to define, control, and assess data quality. These tools included QA program planning, development of QA project plans by each of the Principle Investigators (PIs) responsible for collecting and/or analyzing samples in the LMMB study, implementation of a QA workgroup to guide and monitor QA activities, up-front and continued training, independent verification of all field and laboratory results, and development and implementation of a standardized data reporting format. As part of this QA program, GLNPO also has been applying statistical approaches to develop quantitative assessments of data quality.

For the LMMB Study, all data were categorized, reported, and evaluated by "focus groups." A focus group is defined by sample medium (e.g., lake water, tributary water, fish, plankton, sediment, etc.), by pollutant type (e.g., PCBs, atrazine, mercury, nutrients, and transnonachlor), and by the PI responsible for analyzing the samples. Data quality assessments were conducted by a focus group to reflect the distinct sampling and analytical procedures used by each PI. GLNPO is conducting quantitative assessments for each data focus group for six

data quality attributes: sensitivity, system precision, analytical precision, system bias, analytical bias, and a novel attribute, percent variability due to sampling and analytical measurement uncertainty. These assessments were presented in a paper titled, Will Anyone Ever Read the Lake Michigan Mass Balance Quality Assurance Report, at EPA's 20th Annual Conference on Managing Environmental Quality Systems (April 2–6, 2001, St. Louis, MO). These quantitative estimates reflect data quality at the focus group level.

DERIVING AN INTERVAL ESTIMATE

Study modelers are interested in an interval estimate for single study results, expressed as a range of concentrations based on uncertainty, to set initial conditions and test model outputs. Ideally, such intervals would be derived from collection and analysis of repeated replicates of a given sampling unit. Due to resource constraints, repeated replicates of a given sampling unit were not generally collected and analyzed for the LMMB study, so GLNPO has been exploring alternate approaches that involve use of quality control (QC) sample results to derive these interval estimates.

The types of QC samples available to derive interval estimates varies according to focus group, because the LMMB Study was a performance-based study in which PIs were afforded a great deal of flexibility in choosing the QC tools that would be employed to meet study objectives. The estimates that can be derived, and the variability associated with that estimate, depend on the QC data available for a given focus. Examples of QC samples and the variability included in an interval estimate based on the these QC samples are provided in Table 1. Consideration of the variability that is and is not included in each type of interval estimate is critical when interpreting these estimates.

Recovery-Based Interval Estimate

One approach to developing interval estimates is used in several of the 1600-series methods developed by the EPA Office of Water as a means by which laboratories should monitor their performance. The approach, which is described in OW's Guidance on the Documentation and Evaluation of Trace Metals Data Collected for Clean Water Act Compliance Monitoring (EPA-821-B-96-004, July 1996), uses the mean recovery and standard deviation of ongoing QC measurements (i.e., spiked field samples, spiked reagent water, standard reference

TABLE 1 QC Samples and the Variability Included in an Interval Estimate Derived From These Results

QC Sample	Description	Variability included in Inter-val Estimate
Field spiked sample (FSF)	A routine field sample (RFS) that is split in the field. This split is fortified in the field with known concentration of analyte and analyzed in the lab. The other split is analyzed without fortification.	All field and analytical activities including: sampling, sample shipment, storage and handling, and analysis, including sample matrix effects.
Field control sample (FCM)	Aliquot of reagent water or other neutral item (resin, sand) to which known quantity of target analyte is added in the field. Otherwise handled, transported, and analyzed same as RFS.	All field and analytical activities including: sampling, sample shipment, storage and handling, and analysis, without sample matrix effects.
Laboratory spiked sample (LSF)	An RFS sample that is split in the lab. This split is fortified in the lab with known concentration of analyte and analyzed in the lab. The other split is analyzed without fortification.	All analytical activities and sample matrix effects.
Laboratory control sample (LCM)	Aliquot of reagent water or other neutral item (resin, sand) to which known quantity of target analyte is added. Contains same reagents, solvents, standards, etc. as RFS.	All analytical activities without sample matrix effects.

materials, or surrogate spikes) to establish confidence bounds around analytical results. The interval is estimated as: Mean recovery $\pm$ (Standard deviation * t) *where*: *Mean and standard deviation are the mean and standard deviation of all QC sample recoveries, and t is the 97.5th percentile of the Student's t distribution with n − 1 degrees of freedom, where n is the number of QC sample results.*

The recovery-based interval estimate can be used to estimate the true value of a reported result and to construct bounds around the result. For example, if the result reported is 10 ppb and the recovery-based interval estimate is 84% +/− 25% (i.e., the mean recovery is 84% and the standard deviation of the recovery times the t statistic is 25%) then the true value will be in the range of 9.2–16.9 ppb with 95% confidence. This range is derived as follows:

Lower limit $= [10/(0.84 + 0.25)] = [10/1.09] = 9.2$ ppb
Upper limit $= [10/(0.84 - 0.25)] = [10/0.59] = 16.9$ ppb

The recovery-based interval estimate provides variability information for sampling and the analytical activities associated with the sample result, depending on the type of QC sample used to calculate the interval. If the interval is based on spiked reference matrix samples (as opposed to spiked field samples), matrix effects and associated variability will not be reflected in the estimate. Ideally, the interval estimates would be the same based on spiked field samples and spiked reference matrix samples. Any difference could be attributable to random error or sample matrix effects. Deriving an interval using LCM samples will likely result in a tighter interval than one calculated using spiked field samples because sources of variability regarding matrix effects are not reflected in the estimate. This is shown in Table 2 using QC data from the LMMB Study. For illustration, interval estimates are applied to the median of field sample results for a given focus.[1] Because median results are adjusted based on mean recovery of QC sample results, the resulting interval estimates are not necessarily centered around the median value.

Duplicate-Based Interval Estimate

An interval also can be derived based on the variability between the field sample (RFS) results and their associated field duplicate (FD) results (U.S. Environmental Protection Agency, 2001). This variability is within the pair variance, and it is estimated as the mean squared error or MSE (Neter, Wasserman, and Kutner, 1990):

TABLE 2 Comparison of Recovery Intervals Based on Different Types of Spiked Data

Focus	Median result	Spike type	# QC results	Interval	
				Percent recovery	Concentration
Fish mercury	101 ng/g	Laboratory spiked field sample	9	69.6 to 115.9%	87.1 to 145 ng/g
		Laboratory reference sample	24	87.7 to 112.4%	89.9 to 115 ng/g
Atmospheric total phosphorus	3.6 µg/L	Laboratory spiked field sample	53	81.9 to 125.8%	2.9 to 4.4 µg/L
		Laboratory high check (reference matrix standard)	162	84.6 to 113.1%	3.2 to 4.3 µg/L

$$s^2_{L,w} = MSE = \frac{1}{n}\sum\nolimits_{i\in(1,\ldots,n)} s^2_{(LOG\,RFS_i,\,LOG\,FD_i)}$$

where: n is the number of routine field sample/field duplicate pairs, and
$s^2_{(LOG\,RFSi,LOG\,FDi)}$ *is the variance between the log-transformed routine field sample and field duplicate results in pair i.*
The interval is derived using the standard deviation (the square root of S^2_w) as follows:

$$Exp\{log(Sample \quad result) \pm (s_{L,w}{}^*t)\}$$

where: $s_{L,w}$ is the square root of $S^2_{L,w}$ calculated based on log-transformed RFS and field duplicate results, and t is the 97.5th percentile of the Student's t distribution with n degrees of freedom, where n is the number of duplicate pairs.

The field sample and field duplicate results are log-transformed prior to calculating the pair variances to reflect the fact that absolute variability of analytical data increases with increasing concentration. Log transformation also may address the skewed distribution that can be observed in field data. Log transformation is not necessary for a recovery-based interval because the recovery-based interval estimate is percentage-based, and because the distribution of recoveries is more likely to follow a normal distribution than a distribution of field results. For some LMMB focuses, a small number of field duplicates were collected and analyzed. Therefore, the interval estimate may not accurately reflect the variability for all field samples. The duplicate-based interval estimate is more valid for focuses where there are a large number of field duplicates taken at a large number of study stations.

The interval estimate also can be adjusted for bias by dividing the sample result by the mean recovery of other QC sample results (such as those used to create the recovery-based interval estimate) prior to log-transformation. This adjustment reflects components of bias based on the QC sample used as presented in Table 1 and does not include *variability* associated with the bias estimate.

Table 3 provides examples of recovery-based interval estimates and duplicate-based interval estimates for several example focus groups. As in Table 2, for illustration, the interval estimate is applied to the median of field sample results for a given focus. In addition, these median results were adjusted based on the mean recovery of QC sample results for that focus, as described in the paragraph above, for both interval types; therefore, the resulting intervals are not necessarily centered around the median value.

TABLE 3 Examples of Interval Estimates for Single Study Results Based on Different Types of QC Data

Focus	Median result	Interval type	QC Type	# QC results	Interval	Width
Tributary mercury	4.5 ng/L	Recovery-based	Lab spiked sample	53	3.5 to 5.8 ng/L	2.3 ng/L
		Duplicate-based	Field duplicate	46	3.0 to 6.5 ng/L	3.5 ng/L
Open lake mercury	0.3 ng/L	Recovery-based	Laboratory performance check	68	0.20 to 0.40 ng/L	0.20 ng/L
		Duplicate-based	Field duplicate	13	0.17 to 0.42 ng/L	0.25 ng/L
			Laboratory duplicate	68	0.19 to 0.38 ng/L	0.19 ng/L
Tributary orthophosphate	0.0086 mg/L	Recovery-based	Lab spiked sample	59	0.00828 to 0.00980 mg/L	0.00152 mg/L
		Duplicate-based	Field duplicate	19	0.00335 to 0.0241 mg/L	0.0207 mg/L

235

The interval estimates based on field QC sample results are usually wider than those based on laboratory QC sample results. This occurs, in part, because the intervals based on field QC samples include variability associated with sample collection and handling as well as analytical activities, whereas the interval estimates based on laboratory QC samples reflect variability associated with the analytical activities only. These interval estimates are confidence intervals for a single known study result. For some applications, a prediction interval may be more appropriate.

The interval estimates must be used with consideration to their derivation and the types of variability that are and are not included in the interval. The available QC data for each focus will likely dictate potential approaches for estimates. The Recovery-Based Interval Estimate is one of the simplest approaches and likely will be one of the most broadly applied for LMMB data (i.e., most focuses have laboratory QC data that can be used to construct this interval). However, the variability reflected in the interval depends on the type of QC sample used to derive the interval. Current efforts are focusing on potential approaches to combine imprecision and bias variability into a single interval estimate, in order to encompass all available uncertainty information based on the QC results for a given focus group.

In the absence of repeated replicates of a given sampling unit, QC samples, such as spiked field samples and spiked reference matrices, can be used to develop an interval estimate for a single field sample result based on the recovery of associated QC samples. QC samples, such as field and laboratory duplicates, can be used to develop an interval estimate for a single field sample result based on the variability between RFS results and their associated duplicate results. Interval estimates developed with QC sample results that represent the variability associated with *all* field and analytical activities, and sample matrix effects, will likely result in wider intervals than interval estimates based on QC sample results that represent a subset of the variability associated with field and analytical activities. Consideration of the variability that is and is not included in the interval estimate is critical when interpreting interval estimates.

NOTE

1. The median was chosen arbitrarily as a single example concentration, and the interval is not intended for interval estimates for population parameters such as the mean or median.

REFERENCES

Neter, J., W. Wasserman, and M. H. Kutner. 1990. *Applied Linear Statistical Models*, Third ed., Boston, MA.

U.S. Environmental Protection Agency, Great Lakes National Program Office. October, 1997. *Lake Michigan Mass Budget/Mass Balance Work Plan*. EPA-905-R-97-016.

U.S. Environmental Protection Agency, Office of Water. July, 1996. *Guidance on Documentation and Evaluation of Trace Metals Data Collected for Clean Water Act Compliance Monitoring*. EPA 821-B-96-004.

U.S. Environmental Protection Agency, Great Lakes National Program Office. December, 2001. *The Lake Michigan Mass Balance Study Quality Assurance Report*, EPA 905-R-01-013.