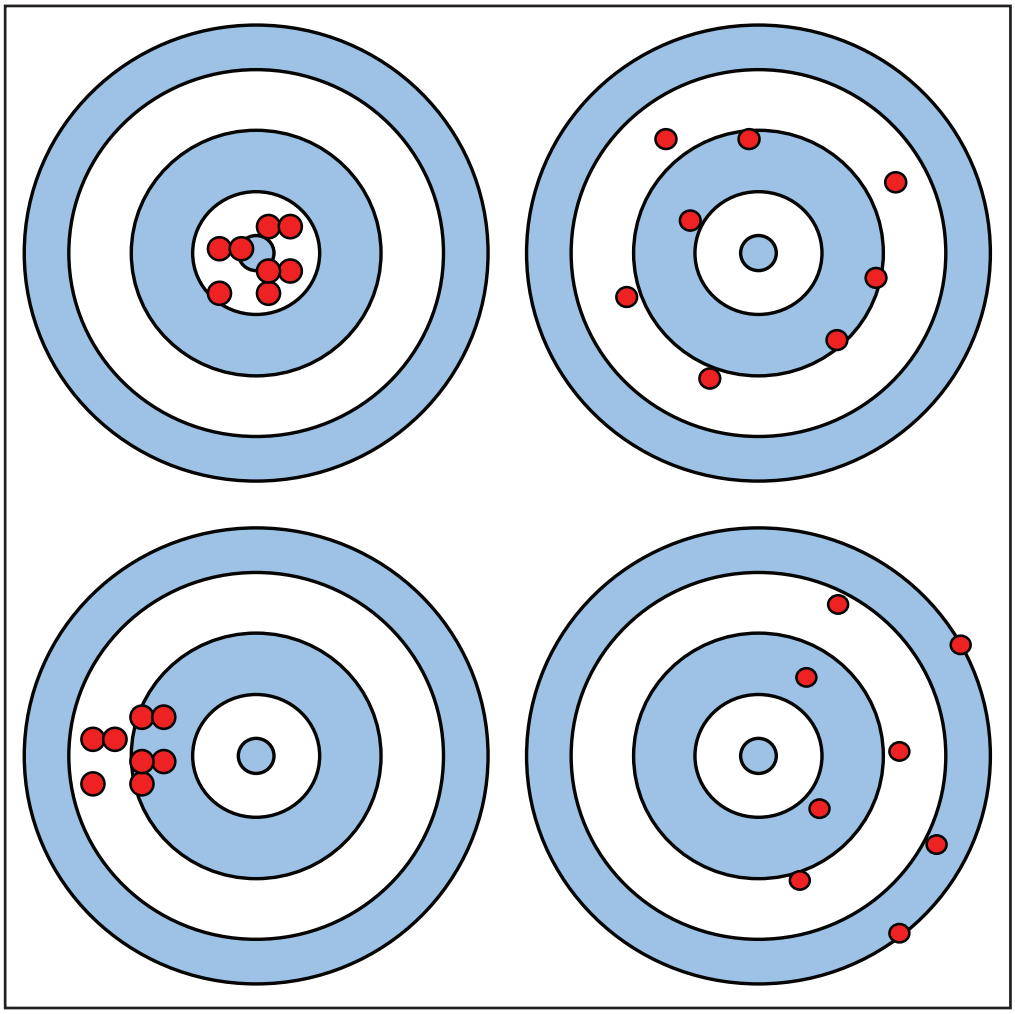
USGS
science for a changing world

Design, Analysis, and Interpretation of Field Quality-Control Data for Water-Sampling Projects

Chapter 4 of
Section C, Water Quality
Book 4, Hydrologic Analysis and Interpretation



Techniques and Methods 4–C4

U.S. Department of the Interior
U.S. Geological Survey

Design, Analysis, and Interpretation of Field Quality-Control Data for Water-Sampling Projects

By David K. Mueller, Terry L. Schertz, Jeffrey D. Martin, and Mark W. Sandstrom

Chapter 4 of
Section C, Water Quality
Book 4, Hydrologic Analysis and Interpretation

Techniques and Methods 4–C4

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod/.

Suggested citation:
Mueller, D.K., Schertz, T.L., Martin, J.D., and Sandstrom, M.W., 2015, Design, analysis, and interpretation of field quality-control data for water-sampling projects: U.S. Geological Survey Techniques and Methods, book 4, chap. C4, 54 p., http://dx.doi.org/10.3133/tm4C4.

ISSN 2328-7055 (online)

Contents

Figures

Tables

Appendix Tables

Conversion Factors

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
meter (m)	1.094	yard (yd)
	Area	
square kilometer (km2)	0.3861	square mile (mi^2)
	Volume	
liter (L)	33.82	ounce, fluid (fl. oz)
liter (L)	2.113	pint (pt)
liter (L)	1.057	quart (qt)
liter (L)	0.2642	gallon (gal)
milliliter (mL)	0.03382	ounce, fluid (fl. oz)
microliter (µL)	0.00003382	ounce, fluid (fl. oz)
	Flow rate	
cubic meter per second (m^3/s)	70.07	acre-foot per day (acre-ft/d)
	Mass	
gram (g)	0.03527	ounce, avoirdupois (oz)
kilogram (kg)	2.205	pound avoirdupois (lb)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F = (1.8 \times °C) + 32$$

Abbreviated Water-Quality Units

µg/L micrograms per liter
mg/L milligrams per liter
µg/mL micrograms per milliliter

Abbreviations and Acronyms

BBP Blind Blank Project (U.S. Geological Survey Branch of Quality Systems)
BQS U.S. Geological Survey Branch of Quality Systems
DQO data quality objective
EPA U.S. Environmental Protection Agency
IBSP Inorganic Blind Sample Project (U.S. Geological Survey Branch of Quality Systems)
IDQTF Interagency Data Quality Task Force
LCL lower confidence limit
LCS laboratory control sample

NAWQA	U.S. Geological Survey National Water-Quality Assessment Program
NELAC	National Environmental Laboratory Accreditation Conference
NWIS	U.S. Geological Survey National Water Information System
NWQL	U.S. Geological Survey National Water Quality Laboratory
OBSP	Organic Blind Sample Project (U.S. Geological Survey Branch of Quality Systems)
PE	performance evaluation (sample)
PT	performance testing (sample)
QA	quality assurance
QAPP	quality-assurance project plan
QC	quality control
QMP	quality management plan
RPD	relative percent difference
RSD	relative standard deviation
SRS	standard reference sample
TKN	total Kjeldahl nitrogen
UCL	upper confidence limit
USGS	U.S. Geological Survey
VOC	volatile organic compound

Design, Analysis, and Interpretation of Field Quality-Control Data for Water-Sampling Projects

By David K. Mueller, Terry L. Schertz, Jeffrey D. Martin, and Mark W. Sandstrom

Abstract

The process of obtaining and analyzing water samples from the environment includes a number of steps that can affect the reported result. The equipment used to collect and filter samples, the bottles used for specific subsamples, any added preservatives, sample storage in the field, and shipment to the laboratory have the potential to affect how accurately samples represent the environment from which they were collected. During the early 1990s, the U.S. Geological Survey implemented policies to include the routine collection of quality-control samples in order to evaluate these effects and to ensure that water-quality data were adequately representing environmental conditions. Since that time, the U.S. Geological Survey Office of Water Quality has provided training in how to design effective field quality-control sampling programs and how to evaluate the resultant quality-control data. This report documents that training material and provides a reference for methods used to analyze quality-control data.

Quality-control data are those generated from the collection and analysis of quality-control samples, and are used to estimate the magnitude of errors in the process of obtaining environmental data. "Bias" and "variability" are the terms used in this report for the two types of errors in environmental data that are quantified by the data from quality-control samples. Bias is the systematic error inherent in a method or measurement system. Variability is the random error that occurs in independent measurements. The types of field quality-control samples discussed in this report include blanks, spikes, and replicates. Blanks are samples prepared with water that is intended to be free of measurable constituents that will be analyzed by the laboratory; blanks are used to estimate bias caused by contamination. Spiked samples are modified by addition of specific analytes; spikes are used to determine the performance of analytical methods and to estimate the potential bias due to matrix interference or analyte degradation. Replicate samples are two or more samples that are considered to be essentially identical in composition. Replicates are used to evaluate variability in analytical results. Various sub-types of these quality-control samples are defined and discussed in this report, and guidance is provided for incorporating the proper samples into the design for a project. The concept of inference space is introduced to help determine where and when quality-control samples should be collected as well as which environmental samples are related to a set of quality-control samples. The recommended basic quality-control design incorporates project-specific considerations, such as the objectives and scale of the study, and hydrologic and chemical conditions within the study area.

The report provides extensive information about statistical methods used to analyze quality-control data in order to estimate potential bias and variability in environmental data. These methods include construction of confidence intervals on various statistical measures, such as the mean, percentiles and percentages, and standard deviation. The methods are used to compare quality-control results with the larger set of environmental data in order to determine whether the effects of bias and variability might interfere with interpretation of these data. Examples from published reports are presented to illustrate how the methods are applied, how bias and variability are reported, and how the interpretation of environmental data can be qualified based on the quality-control analysis.

Introduction

Studies of water quality in the environment require hydrologists to carefully consider the location of the sites to sample, the techniques for collecting and preserving samples, and the analytical methods to use. The goal is to obtain samples from the environment that are handled and analyzed in a manner that does not compromise how well the results represent the environment and meet the study objectives. Historically, the U.S. Geological Survey (USGS) relied on standard practices for collecting representative samples to ensure that samples collected over time and space produced results that were comparable and, therefore, of the quality needed to evaluate the environment. The only routine quality-control (QC) data were generated by the laboratories to ensure that analytical methods were providing appropriate results. In the late 1980s, a series of reports (Shiller and Boyle, 1987; Flegal and Coale, 1989; Windom and others, 1991) identified significant problems in USGS trace-element data. Initially, information about the sources and magnitude of these problems was

limited by lack of routine field QC samples; however, special studies conducted in response to these reports indicated contamination in the field was at least partly responsible for erroneous data (Rickert, 1991). As a result, an overhaul of the sampling techniques used by the USGS in water-quality studies was initiated and implemented during the early 1990s. Part of this overhaul was to include the routine collection of field QC samples.

The process of obtaining water samples includes a number of steps, all of which can contribute to differences between the analytical result and the true value in the sampled environment. The personnel, equipment, and techniques used to collect and filter samples, the bottles used for specific subsamples, any added preservatives, sample storage in the field, and shipment to the laboratory have the potential to affect how well the samples represent the environment from which they were collected. However, a single formula to define the types and numbers of required QC samples does not work for the varied scope of USGS water-quality projects. Instead of trying to craft a minimum QC requirement, the decision was made to develop a training class to teach USGS hydrologists how to design effective field QC samples into their studies. The approach that was developed for the training class has been refined over the years through more than 20 presentations of the lectures and invaluable experience gained from designing and collecting field QC samples in USGS projects. This report captures the sum of that training and experience in what is now a practice that has served the USGS well.

Basic Concepts of Measurement Errors

Three facets of evaluating data quality are (1) quality-assurance (QA) elements, (2) QC data, and (3) quality assessment. The QA elements refer to procedures that are used to manage those unmeasurable components of a project, such as sampling at the right place and time with the proper equipment and using the correct techniques. The QC data are those generated from the collection and analysis of QC samples and are used to estimate the magnitude of errors in the process of obtaining environmental data. Quality assessment is the overall process of determining the quality of the environmental data by reviewing the application of the QA elements and the analysis of the QC data.

Every measurement is subject to potential errors that can cause the result to differ from the true value and can also cause results of repeated measurements to differ from each other. Measurement error has two additive parts: (1) random error varies for each measurement (but is centered on zero), and (2) systematic error occurs in similar samples at about the same value and in the same direction (positive or negative). Random error is caused by factors that cannot be controlled, either because their sources are unknown or because their reduction is not possible within current resources. Systematic error can sometimes be reduced with changes to the sampling procedures or analytical methods if the source can be identified. Measurement errors are a part of any measurement process and should not be considered mistakes. Generally, the goal of QC data analysis is to quantify the random and systematic errors in the measurement process and, only occasionally, to reduce the overall error.

"Bias" and "variability" are the terms used in this report for the two types of errors in environmental data that are quantified by the data from QC samples. Bias is the systematic error inherent in a method or measurement system. Positive bias, typically from contamination introduced in the sample collection and analysis process, causes the results in the environmental samples to be consistently higher than what is actually present in the environment. Negative bias causes the results in the environmental samples to be consistently lower than what is actually occurring in the environment. Negative bias is common in certain analytical methods that routinely measure less than 100 percent of the actual amount of the analyte in the sample. These low measurements can be caused by degradation of an analyte between the time of collection and the time of analysis; interference from something in the sample matrix during the analytical method; or problems with performance of the method, such as incomplete extraction from the sample, losses during solvent evaporation, or sorption to containers. Variability is the random error that occurs in independent measurements. Errors that effect how well the environmental data represent the actual environment can include combinations of both bias and variability at various levels (fig. 1).

Other terms, such as "precision" and "accuracy," are commonly used in the literature to describe data quality. Precision is the degree of agreement between independent measurements; therefore, it is the inverse of variability. Accuracy is generally defined as the degree of agreement between a measured value and the true or expected value; therefore, it is a function of both bias and variability. Sometimes the term accuracy is used as a synonym for bias, although this is not precisely correct. Bias and variability were chosen for use in this report because they are the most consistently defined, used in other literature, and can be calculated directly from QC data.

There are two distinct objectives of field QC sampling that should be considered when designing QC data collection. The principal objective of field QC data is to provide overall estimates of the bias and variability of the environmental data. Estimates of the overall error are used in the assessment phase of the project to (1) support the interpretation of the environmental data without qualification when the errors are insignificant or (2) qualify the environmental data that might be affected by significant errors (either bias or variability) and identify limitations on the interpretation of the environmental data as a consequence of data quality. The QC samples designed to meet this objective provide information on errors from the combination of all procedures used to collect, process, and analyze environmental samples. The second objective of field QC data can be to locate the source of errors that have been identified as potentially affecting the environmental results, with the goal of modifying the procedures to eliminate the source of the error. The QC samples designed to meet this objective provide information on

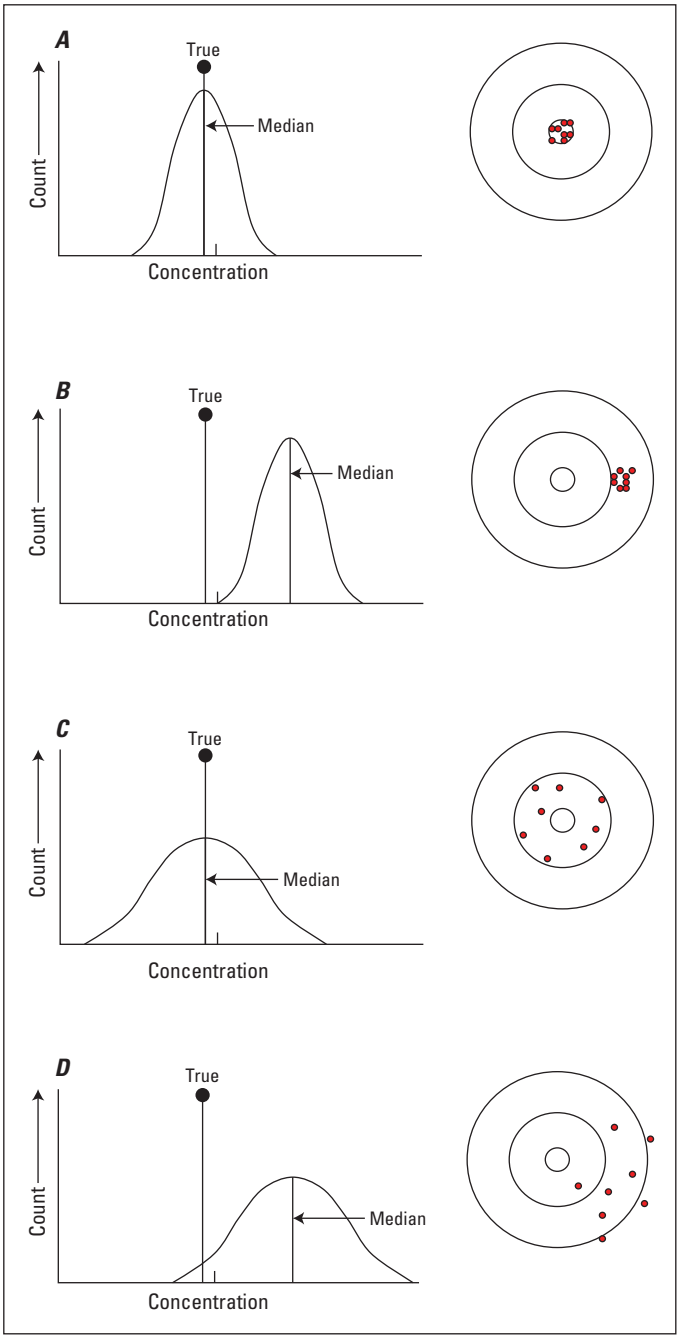


Figure 1. Plots showing the general conditions of bias (indicated by the spread of the dots) and variability (indicated by displacement of the dots): *A*, low bias and low variability; *B*, high bias but low variability; *C*, low bias but high variability; and *D*, high bias and high variability.

selected subsets of the procedures used to collect, process, and analyze environmental samples. The ability to meet either of the two QC-data objectives requires knowledge of the information provided by specific QC sample types, the potential sources of error within the overall sampling and analytical process, the techniques for using QC information in the interpretation of the environmental data, and the spatial and (or) temporal limits on the use of the QC information.

Quality Systems Established by Other Federal Agencies

The U.S. Environmental Protection Agency (EPA) provides guidance on development of a Quality Management Plan (QMP) and a Quality Assurance Project Plan (QAPP). This guidance is documented in reports published by the Office of Environmental Information in 2001 (U.S. Environmental Protection Agency, 2001a; 2001b), which were reissued by memorandum in 2006 (http://www.epa.gov/quality/qs-docs/reissue.pdf). At about the same time, the EPA Federal Facilities Restoration and Reuse Office convened an interagency task force, with the Departments of Defense and Energy, to establish a uniform policy on QA activities for site-evaluation projects at Federal facilities. These policies were documented in a series of reports published in 2005 (Intergovernmental Data Quality Task Force, 2005a; 2005b; 2005c; 2005d).

Quality Management Plans (QMPs) and Quality Assurance Project Plans (QAPPs) are formal documents that provide information about QA and QC procedures and design for programs and projects. A QMP describes the organizational structure, functional responsibilities, lines of authority, and required interactions for those planning, implementing, and assessing all activities conducted by an overall program (U.S. Environmental Protection Agency, 2001a). Within the USGS, QMPs are analogous to Water Science Center QA Plans, which document the general procedures used by Center staff to ensure data quality. A QAPP provides details about the specific QA and QC activities that will be implemented for an individual project to ensure that the results will satisfy the stated performance criteria (U.S. Environmental Protection Agency, 2001b). For most USGS projects, this information is contained in the project proposal; however, some projects might need a separate QAPP to meet cooperator requirements.

The Intergovernmental Data Quality Task Force (IDQTF) was established to address inconsistencies and deficiencies in QA and QC within and across governmental organizations. The IDQTF developed a policy that includes recommendations and guidelines for documentation and implementation of acceptable QA and QC activities for Federal agencies. The policy was developed to ensure that

- Environmental data are of known and documented quality and suitable for their intended uses, and
- Environmental-data collection meets stated requirements.

The policy was initially intended to address QA and QC for hazardous waste cleanup projects, but also was intended to be a model for other programs. The policy is considered simply "guidance" unless it is formally adopted by an agency, and each agency determines how best to implement the policy. Some agencies have used this guidance to develop data-quality objectives (DQOs). Often DQOs are based on rule-of-thumb criteria (such as a requirement that variability be

within 20 percent) rather than constituent concentrations that are environmentally significant, and the consequence of failing to meet criteria is not always defined. A different approach is to collect sufficient QC samples so that the quality of the data can be evaluated in terms of meeting the needs of a project. Methods that can be used to evaluate data quality are presented in subsequent sections of this report.

Scope and Objectives of this Report

The topics covered in this report focus on designing QC-sampling programs for water-quality projects and using QC data to evaluate the quality of environmental-sample data. Sections of this report that address QC data evaluation include background information on pertinent statistical procedures plus specific examples of QC data interpretation from other published USGS reports. The report does not cover topics such as development and implementation of QA plans or collection of QC samples. Some of these topics are covered in documents referenced in this report and also in the USGS National Field Manual for the Collection of Water-Quality Data (U.S. Geological Survey, variously dated).

Types of Quality-Control Samples

Laboratory QC samples are used to estimate bias and variability associated with sample preparation and analysis by the laboratory. Field QC samples incorporate additional sources of bias and variability associated with sample collection, processing, storage, and shipping. Four general types of field QC samples are discussed in this section: blanks, reference samples, spikes, and replicates.

The specific sources of bias or variability that can be evaluated by using these types of QC samples depend on how the QC samples are collected and prepared. Small differences in collection and preparation yield a large variety of sub-types for each of the four general types of QC samples. Unfortunately, terminology of many sub-types is inconsistent among the various institutions that collect and interpret QC samples. For example, the names (or definitions) of many of the QC sample sub-types used in this report differ from those used by the EPA and some other Federal agencies (Intergovernmental Data Quality Task Force, 2005d, p. 62–68). Therefore, QC sample names, methods of collection/preparation, and the potential sources of bias and variability must be specifically defined when designing a QC plan and in reports that describe or use QC data. Subsequent sections of this report define QC sample sub-types and provide information on how differences in QC sample collection and preparation determine which potential sources of bias or variability can be evaluated. These and other definitions of QA and QC terms are compiled in a glossary at the end of this report.

Blanks

Blanks are samples prepared with water that is intended to be free of measurable concentrations of the constituents that will be analyzed by the laboratory. Blanks are used to estimate positive bias caused by contamination. Contamination is the unintentional introduction of an analyte into the sample. Blanks estimate contamination for all or some part(s) of the sample collection and analysis process. Table 1 lists some common types of blank samples and the sources of contamination that could potentially affect them. Many other types of blanks can be collected, particularly when investigating a specific source of contamination. These types include sample-bottle blanks, preservative blanks, filter blanks, and cooler blanks. Typically, the blank type is named for the targeted source of contamination being assessed, but because all blanks measure more than one potential source of contamination, a variety of blank types are usually needed to isolate a specific source of contamination.

General procedures for the preparation of blanks are described in the USGS National Field Manual for the Collection of Water-Quality Data (U.S. Geological Survey, variously dated, chap. 4, p. 136–142). Specific procedures for preparing blanks used by the USGS National Water Quality Assessment (NAWQA) program are described in Mueller and others (1997) and Koterba and others (1995). All blanks prepared for this program must use blank water that has been obtained from and certified by the USGS National Water Quality Laboratory (NWQL). The use of a common blank water among USGS water-sampling projects facilitates the pooling of blank data to evaluate potential sources of contamination. Various grades of blank water are available from the NWQL, and the selection of which to use depends on the constituents to be analyzed. Procedures for the preparation of several types of blanks (described in table 1) are briefly summarized in the following sections. Many details concerning the preparation of blanks provided in the references above have been omitted for brevity.

Field blanks are samples that are intended to document the frequency and magnitude of contamination in environmental water samples. As such, field blanks must be prepared in a manner that exposes the blank water to all of the potential sources of contamination that might affect environmental water samples (table 1). Field blanks are used to evaluate the adequacy of field and laboratory protocols. Field blanks are prepared at the field site where environmental water samples are collected, and are processed before the collection of environmental water samples (Mueller and others, 1997, p. 3).

The most challenging aspect of preparing a field blank is simulating the collection of the water sample. For groundwater, blank water is poured into a clean standpipe, which is used to represent a well. Then a pump or bailer is used to purge the standpipe and collect a sample (Koterba and others, 1995, p. 78). If a pump is used, blank water is passed though the sample tubing to the point where groundwater samples are processed and collected. For stream water, the sampler bottle, cap, and nozzle are rinsed with blank water to simulate field

Table 1. Common types of blank samples and the potential sources of contamination they assess.

[T, targeted source of contamination; X, additional (non-targeted, but unavoidable) source of contamination]

Potential source of contamination	Type of blank sample				
	Field	Equipment	Trip	Source solution	Laboratory
Field sources					
Field Environment					
Air, rain, dust, fumes	T				
Sample-collection personnel					
Dirty hands, personal care products	T	X		X	
Sample collection					
Samplers, pumps, and tubing	T	T			
Sample processing					
Splitters, filters, chambers	T	T			
Sample bottles or vials	T	X	X	X	
Sample preservation	T	X			
Equipment cleaning					
Soap, inadequate rinsing, carryover	T	T			
Transport to and from the field site					
Field vehicles, coolers	T		T	X	
Shipping to laboratory					
Coolers, commercial carriers	T	X	T	X	
Laboratory sources					
Laboratory environment and analysis	T	X	X	X	T
Other sources					
Water used to make the blank[1]	X	X	X	T	X

[1]Although certified as appropriate for preparing blanks, there is a possibility that blank water can be contaminated during shipment or storage before use.

rinsing with native stream water. It is particularly important to use enough rinse water to remove any carry-over contamination or cleaning chemicals, such as methanol (Thiros and others, 2011, p. 37; Bender and others, 2011, p.16). Additional blank water is then poured into the sample bottle, which is capped and shaken to expose the blank water to all interior surfaces. This rinse water is then discarded, as would be done for a native-water rinse in preparing for an environmental sample. Subsequent procedures for processing groundwater or stream-water field blanks are identical to those for processing environmental water samples. These processing steps can include sample splitting, filtration, and preservation. The processed blank is poured or pumped into a sample bottle, transported from the field site, and shipped to a laboratory.

Other agencies refer to field blanks as "equipment" or "rinsate" blanks (Intergovernmental Data Quality Task Force, 2005d). However, these blanks are not necessarily exposed to all sources of contamination that potentially can affect environmental samples. The USGS differentiates between field blanks and equipment blanks based on where and how each is collected.

Equipment blanks, as defined by the USGS, are samples that are intended to demonstrate that sample collection and processing equipment and equipment-cleaning procedures are not sources of contamination. Equipment blanks are prepared in a clean, controlled environment such as a laboratory in the field office. Sample collection and processing equipment is cleaned using the routine protocols, and then an equipment blank is prepared. Blank water is exposed to all of the sample collection and processing equipment in the same manner as is done for a field blank. Because the equipment blank is prepared in the field office, it is not exposed to potential contamination sources associated with the field environment or transport to or from a field site (table 1). Although equipment blanks are prepared primarily to evaluate the equipment and equipment-cleaning procedures, some other potential sources of contamination are unavoidable. These sources include sample-collection personnel, sample bottles, sample preservation, shipment, laboratory analysis, and the water used to make the blank (table 1).

In order to receive and evaluate the analytical results, equipment blanks typically are prepared months before beginning water-sampling activities. Any contamination measured in equipment blanks is evaluated in terms of project objectives and the anticipated environmental concentrations to be measured for the project. A decision then is made whether to begin water-sampling activities, to change equipment or protocols, or to revise project objectives.

Trip blanks are samples that are intended to demonstrate that the transport and shipment of samples are not sources of contamination. Trip blanks typically are collected only for volatile organic compounds (VOC). Trip-blank vials are filled with blank water in the laboratory or office prior to a sampling trip. They are transported, along with empty sample vials, to the field site; kept with environmental VOC samples during the period of sampling and sample storage; and shipped to the laboratory with the environmental samples for analysis. Although not specifically targeted, sample vials, laboratory analysis, and the water used to make the blank also are potential sources of contamination in trip blanks (table 1).

Source-solution blanks are intended to demonstrate that the water used to make the other blanks is not a source of contamination (table 1). Source-solution blanks are prepared in a clean environment at the field site by pouring blank water directly into a sample bottle or vial. The remaining blank water is then used to prepare a field blank. The blank water used to prepare a source-solution blank must be from the same source and lot as that used to prepare the field blank. The source-solution blank and the associated field blank are interpreted as a set. Although not specifically targeted, the sample-collection personnel, sample bottles or vials, sample transport in the field, sample shipment to the laboratory, and laboratory analysis also are potential sources of contamination in source-solution blanks (table 1).

Laboratory blanks are samples that are intended to demonstrate that processing and analysis by the laboratory is not a source of contamination (table 1). Laboratory blanks are prepared by analysts at the laboratory and are primarily used by analysts to assess contamination in the analytical method. Laboratories use various terms for these blanks. The USGS NWQL refers to them as "reagent blanks" because they are made from clean, reagent (blank) water and are subjected to all steps of the analytical method. The NWQL also uses the term "set blanks" because they are associated with a set of environmental samples that are all prepared and analyzed at the same time. Other laboratories differentiate between "method blanks," which go through preparatory steps and "reagent blanks," which do not (Intergovernmental Data Quality Task Force, 2005d). Laboratories use a variety of approaches to compensate or qualify reported analytical results when contamination is identified in laboratory blanks. Typically, reporting levels for all associated samples are increased to some multiple of the concentration in the blank, or results for these samples are flagged with a data-qualifier code. Data qualifiers used by USGS are available at http://help.waterdata.usgs.gov/codes-and-parameters/codes#WQ; those used by the EPA are listed in U.S. Environmental Protection Agency (2010), for inorganic analytes, and U.S. Environmental Protection Agency (2008), for organic compounds.

Reference Samples

Reference samples contain known concentrations of selected analytes. Similar to blanks, they are used to estimate bias, but in this case bias can be positive or negative (if the measured value is less than the known concentration). Reference samples are used by laboratories to evaluate the performance of analytical methods during development, to test performance of a method at a specific concentration of interest, or to monitor performance routinely. External performance programs also use reference samples to test the capability of a laboratory.

Reference samples can be prepared in a variety of ways. Laboratory control samples are prepared in individual laboratories by spiking reagent water with analytes of interest at the midpoint of the calibration curve or at a specific concentration of interest for evaluating performance of a method. Certified reference samples are prepared by an external provider and have values measured multiple times (or by different laboratories) so that each certified value is accompanied by an uncertainty at a stated level of confidence. These can be used as proficiency-testing (PT) or performance-evaluation (PE) samples to test the laboratory's ability to qualitatively identify and accurately quantitate analytes in a given matrix. If the concentrations of the analytes in the PT or PE sample are unknown to the analyst, the sample is referred to as a "blind sample." If, additionally, the identity of the sample as a PT or PE sample is unknown to the analyst, the sample is referred to as a "double-blind sample" (Intergovernmental Data Quality Task Force, 2005d). Standard reference samples (SRS) are prepared by the USGS Branch of Quality Systems (BQS) using mixtures of natural waters. These samples are analyzed by many laboratories as part of their performance evaluation. They are more completely described under "Standard Reference Samples (SRS) Project" in the Branch of Quality Systems section of this report.

The NWQL participates in a number of proficiency test programs, including those of the National Environmental Laboratory Accreditation Conference, the New York State Department of Health, and the BQS SRS inter-laboratory comparison for inorganic analytes (Maloney, 2005). Additional reference samples can be submitted by field projects, but this is not a common practice. Consequently, reference samples only are discussed in this report relative to additional QC data available from the laboratory or external method-performance programs.

Spikes

Spikes are water samples fortified (spiked) with known concentrations of analytes. Similar to reference samples, spikes are used to estimate positive or negative bias, and are used primarily to determine whether this bias is due to method performance, effects of the sample matrix, or analyte degradation during sample shipment and storage. Bias for spikes is termed "recovery": the concentration measured in the sample expressed as a percentage of the known concentration that was added to the sample. Calculation and interpretation of spike recovery is presented in the "Analysis and Interpretation of Data for Spikes" section of this report.

Spikes are defined by the location where the spike solution is added to the sample, either in the field or in the laboratory, and by the type of water that is spiked, either environmental (matrix) water or blank (reagent) water (table 2). Field matrix spikes and laboratory matrix spikes are used to estimate recovery bias in an environmental water sample. Low recovery (negative bias) in matrix spikes could be caused by a variety of factors including matrix effects, degradation, and analytical performance. Matrix effects are the chemical, physical, and biological characteristics of environmental water that might interfere with or compromise chemical analysis of the sample. Field reagent spikes and laboratory reagent spikes are used to estimate recovery bias of the analytical method in

Table 2. Common types of spiked samples and the potential sources of bias they assess.

[T, targeted source of bias]

Potential source of bias	Field matrix spike	Laboratory matrix spike	Field reagent spike	Laboratory reagent spike
Field sources				
Field environment				
Water matrix interference	T	T		
Shipping to laboratory				
Analyte degradation	T		T	
Laboratory sources				
Laboratory environment and analysis	T	T	T	T

a blank-water sample. High or low recovery for reagent spikes might indicate a problem with performance of the method. Lower recovery in the field spike than in the laboratory spike might indicate that the analyte has degraded, through loss or chemical conversion, during the time between collection in the field and analysis in the laboratory. Lower or higher recovery in matrix spikes than in reagent spikes might indicate matrix effects from the environmental water sample. Matrix effects, method performance problems, or analyte degradation identified in spikes would similarly affect analytical results of environmental samples.

Replicates

Replicates are two or more water samples that are collected, prepared, and analyzed such that they are considered to be essentially identical in composition and analysis. Replicate is the general term; duplicates are two replicates, triplicates are three replicates, and so forth. Replicate environmental samples are used to estimate variability (random measurement error) of analytical results. In addition, replicate blanks, replicate reference samples, and replicate spikes can provide an estimate of the variability associated with various types of measurement bias. The process of collecting replicates might require some modification of field procedures (for example, sample volume usually must be increased); however, most aspects (such as sampling and processing equipment, and personnel) should be kept the same as for normal sampling. Unnecessary changes to field procedures can introduce sources of replicate variability that are not likely to affect routine environmental samples.

Replicates can be collected in several ways. Split replicates are made from a single sample that is collected and then subdivided into other samples. Concurrent replicates are made from multiple samples that are collected at about the same time. Sequential replicates are made from multiple samples that are collected one after another. The different types of replicates assess different sources of variability (table 3). Sequential replicates can include environmental variability within the sampled medium; therefore, they are not appropriate under certain conditions (for example, when water chemistry is changing rapidly) or if collection of the first replicate might affect the content of the second replicate (for example, fish shocking or bed-sediment sampling). Other agencies refer to split replicates as "subsample" replicates and also use the term "co-located" replicates, but they do not distinguish whether these are collected concurrently or sequentially (Intergovernmental Data Quality Task Force, 2005d).

Sometimes samples that might be called replicates are used to investigate some difference in the data-generation process. Samples used for this purpose are defined herein as "irreplicates" to emphasize that these samples are not used to assess variability. Typically, the goal of collecting irreplicates is to assess the comparability of data that have been generated through different methods. Depending on the results of the comparability assessment, data generated differently may be pooled for analysis. Some examples of the use of irreplicates are to:

- Compare analyses by different laboratories or analytical methods,
- Compare samples collected using different sampling equipment or techniques,
- Compare preserved with unpreserved samples,
- Compare filtered with unfiltered samples,
- Compare suspended-sediment samples from a cone splitter with conventionally collected sediment samples,
- Compare an auto sampler with traditional sampling methods, or
- Compare the effects of sample holding times.

Other agencies use the term "split samples," defined as subsamples that are sent to different laboratories or analyzed using different methods (Intergovernmental Data Quality Task Force, 2005d). Such split samples are a type of irreplicates.

Table 3. Common types of replicate samples and the potential sources of variability they assess.

[T, targeted source of variability; X, additional (non-targeted, but unavoidable) source of variability]

Potential sources of variability	Replicate type		
	Spilt	Concurrent	Sequential
Field sources			
Sample collection		T	T
Sample splitting and filtering	T	T	T
Temporal change in sampled medium			X
Laboratory sources			
Laboratory environment and analysis	T	T	T

Design of a Field Quality-Control Sampling Program

Designing a field QC sampling program is the process of deciding what types of QC samples are needed, how many of each type of QC sample are needed, and when and where the QC samples should be collected in order to adequately and efficiently assess the quality of the environmental data. The QC design needed for a project is dependent on the objectives of the project and consideration of the potentially important sources of bias and variability in the laboratory results. Bias and variability are estimated using results of QC samples, and inferences about the quality of the environmental data are made using these estimates.

Designed QC samples are selected by the investigator to meet the technical needs of the project. How adequately data quality must be determined depends on the questions being addressed by the project. In addition, QC sample design may be influenced by nontechnical considerations such as:

- Is this a routine project or something more complex?
- Is there great interest in the project?
- Could the conclusions be controversial?
- Are there human health and welfare issues?
- What is at stake? What is the cost of being wrong?

For some water-sampling projects, QC is prescribed. Some aspects of the design, usually the numbers and types of QC samples, are specified by a program administrator, such as the USGS NAWQA national leadership team, or a regulatory authority, such as the EPA. Even if the types and numbers of QC samples are prescribed, when and where to collect them might need to be designed, and additional designed QC samples might be necessary to adequately determine the quality of the environmental data or to locate data-quality problems.

A complete suite of QC samples generally cannot be collected in association with each environmental sample; therefore, some set of QC samples will need to be considered applicable to a larger set of environmental samples. Determining which QC samples best relate to which environmental samples relies on the concept of "inference space."

Inference Space

Inference space is a concept used in the design of experiments (for example, see Anderson and McLean, 1974). In many ways, the design of a QC program is the design of an experiment to determine the quality of the environmental data. The inference space is the location in time and space within which the results of the experiment are valid. In terms of QC, inference space is used to determine which QC samples can be related to which environmental samples. These related QC samples represent the same conditions, in terms of potential bias and variability, under which the environmental samples were collected.

For example, a field blank collected at the same time and location as an environmental sample is usually considered to represent the sources of contamination that potentially affect the environmental sample. This blank might also represent potential sources of contamination for samples collected at about the same time in a similar setting using similar equipment and sampling procedures. Inference space defines the extent of this representation.

For some regulatory projects, such as EPA Superfund or Department of Defense installation investigations, the inference space of QC samples is explicitly specified. For example, Superfund investigations must include a field blank with each "set" of environmental samples (U.S. Environmental Protection Agency, 1989). Typically, a set is all samples shipped in the same cooler, so the inference space for the blank in the cooler is the environmental samples in the cooler. If the field blank is clean, then concentrations reported for all the environmental samples are assumed to be unaffected by contamination. If the concentration of a particular analyte in the field blank is greater than detection, then concentrations of that analyte in all the environmental samples are subject to qualification or an elevated reporting level (U.S. Environmental Protection Agency, 1989, p. 5-16 and 5-17).

If inference space is not externally specified, as is the case for most USGS studies, the project chief or program manager must determine the QC samples that are needed to represent the range of environmental conditions that are expected to be encountered. One approach is to identify important factors or variables that could affect bias or variability in samples collected over this range of conditions. The QC sampling is then stratified so that a set of samples is collected within each combination of factors. Each of the sets of QC data is then used to represent the potential bias and variability in an associated group of environmental samples. If some factors turn out to be unimportant, QC samples can be pooled and associated with a larger group of environmental samples.

Classification of Quality-control Samples by Use: Basic or Topical

The various sub-types of QC samples may be classified as "basic" or "topical" QC samples dependent on the intended use of the QC information. Basic QC samples measure all of the potential sources of bias or variability that might affect environmental samples and are used to estimate the overall quality of the environmental data. Topical QC samples measure a limited number of sources of bias or variability; thus, they cannot be used to estimate the overall quality of environmental data. Topical QC samples are intended to measure some specific, targeted aspect of bias or variability. These samples are typically used to (1) investigate the causes of data-quality problems, (2) assess the comparability of methods

or equipment, (3) determine whether sampling equipment and protocols are adequate to initiate environmental sampling, or (4) verify that blank water is suitable for preparing blanks or cleaning equipment. When used to investigate the cause of data-quality problems, a series of topical QC samples often are collected at the same time as a basic QC sample and are interpreted as a set in order to isolate the cause of the problem.

Basic QC samples include field blanks, field matrix spikes, and field replicates; most other types of QC samples, including reference samples, are topical. Water-sampling projects should always collect basic QC samples to document the quality of the environmental data. Topical QC samples might be needed, but often are not. Only basic QC samples will be discussed in the design, analysis, and interpretation sections of this report.

Design Considerations for Blanks

How, when, where, and why field blanks are collected control the potential sources of contamination that might be included. For example, a field blank could be collected either before or after the environmental sample. Does the sequence of preparation influence the sources of potential contamination that could affect the field blank? Consider the following two sequences.

Sequence 1: the field blank is prepared before an environmental sample is collected:

1. Arrive at field site,
2. collect field blank,
3. collect environmental sample,
4. clean equipment at field site, and
5. depart from field site.

Sequence 2: the field blank is prepared after an environmental sample is collected:

1. Arrive at field site,
2. collect environmental sample,
3. clean equipment at field site,
4. collect field blank, and
5. depart from field site.

The field blank collected before the environmental sample includes one additional source of potential contamination: storage and transport of the sampling equipment and supplies prior to arrival at the field site. Storage time could be short (for example, if a previous sample was collected the same day) or could be much longer (for example, if the blank was collected before the first sample of a monthly field trip). Unless the specific sequence of activities for preparing the blank is described, data users will not know which sources of potential contamination are relevant.

Design Considerations for Spikes

In general, spikes need to be included in QC sample designs only if environmental samples will be analyzed for organic compounds. Analyses of inorganic constituents usually are not much affected by matrix interference, nor do these constituents tend to degrade if properly preserved. Many methods for organic constituents do not include preservatives, so the use of field matrix spikes is important to evaluate analyte stability and degradation during sample shipment and storage. QC designs should include samples spiked at the field site (field matrix spikes) to get the most information about potential bias from analyte degradation as well as matrix interference on the analytical method.

Spike recoveries can have large errors if the background concentration in the environmental water is similar to or greater than the expected concentration added to the spiked sample. These errors are caused by variability associated with two analytical results: one for an environmental sample (to determine background) and one for the spiked sample (to determine recovery). Computed recovery is more meaningful for matrix spikes in which the spiked addition increases analyte concentration by at least five times over background. A good QC design should avoid spiked samples when environmental concentrations are expected to be high, relative to the spiked addition. If this is not possible, the amount of spiked material should be increased.

Field matrix spikes are the only spikes required in a basic QC design. If recovery in these spikes is outside acceptable limits, laboratory matrix and reagent spikes should be added in order to determine whether the cause is analyte degradation, matrix interference, or analytical performance. If recovery in the field matrix spike is lower than in the laboratory matrix spike, the likely cause is analyte degradation during sample shipment and storage. If recoveries in both matrix spikes are different from recovery in the reagent spike, the likely cause is matrix interference. If recoveries in all three spikes are similar and outside acceptable limits, the likely cause is analytical performance. Bias due to analyte degradation or analytical performance can affect all environmental samples. Bias due to matrix effects might be limited to environmental samples from that particular matrix, which could be constrained by sampling location or time period. If laboratory matrix spikes are required, a separate sample must be submitted with a request for spiking in the laboratory. Laboratory reagent samples are routinely prepared, and results can be requested from the laboratory.

For some VOCs, typical field spiking procedures are inadequate, and the only option is to rely on laboratory matrix spikes. In this case, the laboratory matrix spikes should be held for several days between preparation and analysis in order to simulate sample shipment from the field site and account for potential degradation.

Design Considerations for Replicates

The type of field replicates (split, concurrent, or sequential) in a QC design should be selected to incorporate the most likely sources of sampling and analytical variability and the least amount of environmental variability. Split replicates are prepared so as to avoid all environmental variability, but this process also excludes potential variability due to sample collection. Sequential replicates always have the possibility of including some environmental variability, though in some cases this can be assumed negligible. Exactly concurrent replicates can be impossible to collect. The choice of which type of replicate to collect generally must be based on a variety of logistical considerations. For example, groundwater replicates collected using a submersible pump are usually collected sequentially; however, this typically adds negligible environmental variation, particularly if replicate samples for each group of analytes are collected immediately one after the other. For stream water, it might not be reasonable to collect concurrent replicates under high-flow conditions, so split replicates could be the best option. Also, samples for VOCs cannot be split because the samples must not be exposed to ambient air. Therefore, replicates for these analytes must be collected either concurrently (if two sets of sampling equipment are available) or sequentially. It is best to use only one type of field replicate throughout a project so that any difference in replicate collection is not a factor in estimated variability.

Hydrologic and Chemical Considerations

Sources of bias and variability can be influenced by hydrologic and chemical conditions at sampling sites during the time of sample collection. The QC samples need to be distributed over a range of conditions in order to ensure adequate evaluation of the potential effects of bias and variability. For example, variability for samples collected at stream sites might be greater during high runoff or storm events than during base flow; therefore, separate sets of replicates are needed to evaluate these different levels of variability. In other words, the inference space for these replicates will be limited in part by the streamflow condition during sample collection. Similarly, sources of bias can vary by location and season, depending on where and when different levels of contaminants might occur. For some analytes, samples collected in agricultural areas during times of pesticide or fertilizer application could be subjected to high positive bias. The blanks used to evaluate this bias need to be collected at locations and times that adequately represent this application condition. The inference space for blanks that represent such special conditions probably will not include locations or time periods where these conditions do not occur.

Some chemical constituents and environmental media are prone to greater bias or variability. Constituents that commonly occur in the field or laboratory environment, including ammonia, trace metals, VOCs and, plasticizers, can cause high contamination bias. Contamination also can be a problem for many constituents in samples of low-ionic strength water, such as precipitation. For studies that include these constituents or media, QC sample design should emphasize field blanks, in order to ensure that contamination can be adequately evaluated. Some constituents, primarily organic compounds, are subject to negative bias due to low analytic recovery. Low recovery of these constituents can be exacerbated in samples that have high concentrations of dissolved organic carbon, such as water from swamps and wetlands or sewage-treatment effluent. For studies that include these constituents or media, QC designs should emphasize field matrix spikes. Analytical results for suspended sediment and sediment associated constituents, such as phosphates and hydrophobic organic compounds, can be affected by high variability. Chemical constituents in samples of bottom sediment and biological tissues also are subject to high variability. For studies that include these constituents or media, QC designs need an emphasis on replicates.

The expected concentrations of analytes in environmental samples also should be considered in the design of the field QC program. High concentrations at a sampling site can be a source of carry-over contamination if equipment cleaning is not adequate or not done properly; therefore, blanks should be targeted before sampling the subsequent site. High concentrations can also affect computation of recovery if the spiked amount is small by comparison. For this reason, spikes should be targeted at times and locations where the background concentration is low relative to the spiked addition. Conversely, replicates should be targeted at times and locations where concentrations are expected to be high (during storm runoff, for example), or at least greater than the analytical reporting level. Computation of variability is not possible if the concentration in one or more replicates is reported as a censored value (less than the reporting level). Replicates in the low range of concentrations are common in most studies; high-range replicates are less common. Targeting high concentrations is important for getting a balanced set of replicate data.

Considerations Based on Study Objectives and Scale

The objectives of a study must be considered in determining how the environmental data will be analyzed, and thus what environmental samples must be collected. Similarly these objectives need to be considered in the QC sampling design. Water-quality projects can be categorized according to their objectives; some of these categories are:

- Reconnaissance projects: short-term studies to characterize current conditions at a particular location or within a specified area,
- Compliance-monitoring projects: studies to identify possible exceedances of a standard or criterion,
- Trend-analysis projects: long-term studies of possible changes in water quality over time.

Each type of project has its own needs for QC data. Reconnaissance projects are conducted over a short timeframe, so QC sampling must be concentrated early in the study, or even before environmental sampling begins, in order to ensure that there are no problems with sampling equipment, cleaning procedures, or sample processing. Compliance monitoring is generally longer term, so QC samples should be distributed throughout the project to ensure that any possible changes in data quality can be considered in a presentation of monitoring results. If the concentrations of standards or criteria are high relative to potential contamination, blanks might be less important than spikes and replicates that target the concentration of the water-quality standard. Trend analysis is particularly sensitive to changes in bias over time, so blanks and spikes are important and should be distributed over the entire time of environmental sampling. Reference samples are an option for evaluating method performance over time if a spiking material is not available (as is the case for most inorganic constituents); however, reference samples are routinely analyzed by most laboratories, so these do not necessarily need to be submitted from the field. Replicates are less important because the large number of environmental samples typically collected for these studies can overcome the effects of variability in statistical analysis for trends.

The spatial and temporal scale of a project can affect the inference space over which a set of QC samples can be assumed to apply. Spatial factors that might need to be considered in QC sampling design include land use, geology, altitude, soils, slope, crop types, point sources, stream size, and climate. The QC samples should be distributed among areas with differences in these spatial factors. This distribution could be in proportion to:

- The importance of the area to study objectives,
- The number of environmental samples collected from the area,
- The variability in environmental concentrations expected for the area, and
- The potential effects of bias or variability on data from the area.

Temporal factors that might need to be considered include seasonal, diurnal, annual, and decadal cycles; and the timing of floods, droughts, and groundwater pumping. The QC samples could be distributed during various time periods in several ways:

- Equally spaced throughout all time periods,
- In proportion to the number of environmental samples collected during each time period,
- Higher proportions during the initial phase of the study,
- In proportion to the importance of each time period to study objectives, or
- In proportion to the variability expected in environmental concentrations during each time period.

Selection of one or more of these designs for distribution of QC samples can be based on study objectives, as discussed above.

Design of a QC sampling program should also consider anticipated study results. If low concentrations of target analytes are expected to occur throughout the study area, then the QC design should emphasize field blanks in order to determine whether contamination of samples has the potential to affect the reported concentrations in environmental samples or the interpretation of study results. If concentrations of target analytes are expected to be close to a regulatory standard or some other numerical threshold, the QC design should emphasize replicates at concentrations near the standard in order to determine whether variability of results might affect the determination of compliance with, or exceedance of, the standard. This design also should include samples spiked at concentrations near the standard in order to demonstrate that these concentrations can be reliably measured in environmental samples. If concentrations of target analytes are expected to be different in samples collected under various environmental conditions, the QC design should emphasize samples that can be used to determine whether reported concentrations in environmental samples from all conditions have similar data quality. For example, if concentrations in groundwater under oxidizing conditions are expected to be greater than concentrations under reducing conditions, then field spikes are needed to determine whether there is a matrix effect associated with either water type.

Overall Approach

A QA plan for water-quality studies should focus on basic QC samples: field blanks, field matrix spikes, and field replicates. In general, the only necessary topical QC sample is an equipment blank collected before environmental sampling begins. This sample is used to ensure that sampling equipment and procedures are not sources of contamination for any target analytes.

The basic QC sampling design should be stratified over the various inference spaces that have been identified within the study area and time frame. Within each identified inference space, QC sampling should be randomized, so that each time and site for collection of an environmental sample has an equal chance of being selected for QC sampling. However, selection of specific types of QC samples also should consider the expected environmental conditions. (For example, blanks should be collected at sites or times when sources of contamination are expected to be present. Replicates should be collected only at times and sites where concentrations are expected to exceed analytical method detection limits. Spikes should be collected only at times and sites where concentrations are expected to be low relative to the spiked addition.) Other considerations for the design of QC samples include:

- Focus QC sampling early to identify any problems that must be corrected,
- Focus QC sampling after a change in sampling equipment, sampling procedures, or laboratory methods,
- Focus QC sampling on any expected issues concerning potential bias or variability,
- Focus some QC sampling on unusual hydrologic conditions, and
- Reallocate some QC samples to focus on any issues of bias or variability that become apparent as sampling proceeds.

Evaluate the QC data in relation to environmental data immediately following release of results from the laboratory. This evaluation is particularly important during the initial stage of the project in order to identify, and perhaps fix, any problems before the bulk of the environmental samples have been collected. If possible, use the same sampling personnel, equipment, sample-processing procedures, and laboratory methods throughout the project in order to avoid differences in the potential sources of bias or variability that might complicate data interpretation. When these differences are unavoidable, such as for long-term projects or large programs, the QC design should include topical QC samples (irreplicates) that can be used to verify the comparability of field and laboratory methods.

As the project progresses, continue to evaluate the basic QC data in order to determine whether any target analytes are particularly subject to high bias or variability at certain sites or under certain conditions. If so, consider rearranging the QC design to focus less on the good-quality samples and provide more data to evaluate these problematic conditions. Also, evaluate the environmental data as they become available in order to identify what significant findings are emerging from the study. Ensure that the QC results are adequate to support these findings so they can be reported with confidence in the data base and interpretive reports.

Other Sources of Quality-Control Data

In addition to data from field QC samples, information available from the analyzing laboratory can aid in evaluation of the quality of environmental data. Contract laboratories often provide QC information, such as results from set blanks and laboratory matrix spikes, in a report on results from analysis of environmental samples. For projects that submit samples to the NWQL, there are several sources of QC information available on the USGS websites. These sources include the NWQL itself and the BQS.

National Water Quality Laboratory

Routine QC data produced by the NWQL include results for several types of QC samples and for QC analytes (surrogates and internal standards) (Maloney, 2005). Instrument QC samples are used to calibrate the analytical instrument initially, to monitor contamination from the analytical instrument, and to verify calibration linearity and response during the analytical run. Typical instrument QC samples included in an analytical run or batch of samples include calibration standards, instrument blanks and carryover blanks, and continuing calibration verification standards. The analyst uses the instrument QC samples to ensure that the analytical instrument is producing acceptable qualitative and quantitative data. If data are outside acceptance criteria, the analyst can either take corrective action and then reanalyze the samples if possible, or report results with appropriate qualifiers. Results from instrument QC samples are generally not useful for data users in interpreting field data, so they are not reported to the data user with sample results.

Preparation or set QC samples are typically assigned to a set of field samples during preparation and other steps prior to instrument analysis. Set QC samples are reagent blanks and reagent spikes. These samples are prepared using reagent-grade blank water, and go through the same preparation, extraction, and other steps as field samples before analysis on the instrument. In this regard, they are similar to field QC samples, and thus are the most useful internal-laboratory QC data for comparison with field blanks and spikes.

Routine QC data are available from the NWQL in a variety of ways, depending on the need and stage of a project. These QC data can be retrieved from the NWQL internal website; however, access is restricted to connection using the USGS network. Set QC data can be reviewed for a particular sample using the Sample Status application (http://wwwnwql.cr.usgs.gov/USGS/sampstatus/). This application is useful for reviewing set-blank results in comparison to an equipment blank or single field blank. Laboratory QC data for a larger group of samples can be reviewed by the QC sample results application (http://nwqlqc.cr.usgs.gov/). This application provides summary statistics, graphs of results, and downloadable files for individual analytes, all of which are useful for comparing field matrix spikes or field blanks to laboratory reagent spikes or blanks analyzed during a specific time. Unfortunately, this application only allows retrieval of QC data by analyte and is not very efficient for multi-analyte methods. Routine laboratory QC data for a group of analytes in a method can only be obtained by special request to the NWQL for a retrieval of the QC data from the laboratory information system.

Surrogates and internal standards are two types of QC analytes commonly used in all gas and liquid chromatography methods (most methods for organic constituents). These QC analytes are similar in that they are added to all samples, but

are used for slightly different purposes. Surrogates are added to samples prior to sample preparation and are used to provide information about recovery and matrix effects in the method. Internal standards are added to samples prior to instrument analysis and are used for calculating quantitative results. The number of QC analytes varies by method but is usually a small percent of the total number of analytes (3 to 4 of each for a method with 60 analytes).

Surrogates are compounds similar in physical and chemical properties to the analytes of interest but normally are not found in environmental samples. They are expected to behave similarly in a method in terms of instrument response and recovery. Typical surrogates are compounds that have been isotopically labeled or have hydrogen replaced with fluorine or bromine and have chemical structures similar to one of the analytes in the method. Surrogates are added to all environmental and QC samples at the start of sample preparation. Because surrogates are added to every sample, they provide sample-specific QC information by monitoring matrix effects and gross sample-processing errors. If all surrogates from a sample are outside acceptance limits, the analyst will take corrective action and re-analyze the sample. If re-analysis is not possible, the analyst will qualify the original results in the database. Surrogates are not used to calculate or correct analyte concentrations. Surrogate results are reported to the data user because they are useful for interpretation of possible matrix effects and understanding of qualification codes.

Internal standards are chemicals that provide good response on the analytical instrument but are not analytes in the method nor expected to be found in environmental samples. Internal standards are added to samples just prior to instrumental analysis for use in quantitative analysis calculations. This calculation method uses the response ratio of the analyte to the internal standard for calibration, and therefore compensates for changes in response between sample and calibration standard due to differences in the amount of sample injected into the instrument. The response of the internal standards is monitored throughout the analysis of a group of samples to check for changes in sensitivity or matrix effects. The internal standard results are generally not useful for interpretation of environmental sample results, so are not reported to the data user.

Branch of Quality Systems

The role of the BQS is to assure the quality of laboratory and field measurements and to supply reference materials to USGS water-quality programs and projects. Data from blind sample QA projects or programs operated by BQS are used to estimate bias and variability within the field, analytical, and measurement processes. Statistical summaries, charts, and data reports summarizing the QA data are produced and made available on the BQS website (http://bqs.usgs.gov/). Information provided by BQS, especially the information on laboratory performance over time for the wide range of analytical methods covered, can be incorporated into the overall evaluation of data quality for a project.

Standard Reference Samples (SRS) Project: The SRS Project formulates reference samples for trace elements, major ions, mercury, and nutrients using various mixtures of natural-matrix water. These samples are submitted semiannually for analysis to about 100 laboratories throughout the United States. The true concentration of each constituent is not known, but a most probable value (MPV) is determined as the median of results from all laboratories. These results also are used to calculate a measure of variability, F-pseudosigma, for each constituent. Individual results are normalized by subtracting the median and dividing by F-pseudosigma; therefore, analytical variability can be compared without bias among samples with a variety of constituent concentrations. Data for each laboratory, including the NWQL, are available from the BQS website, and project staff can evaluate laboratory performance during the time period their samples were being analyzed. Information on the variability of individual analytical methods, based on results from all laboratories, is also available for each constituent.

Inorganic Blind Sample Project (IBSP): The IBSP prepares reference samples for the majority of inorganic constituents analyzed at the NWQL. These samples are submitted as if they were environmental samples collected in the field, so their reference status is blind to the laboratory analysts. The blind samples are designed to capture the same sources of analytical variability that affect environmental samples. Data from these samples are used to evaluate method performance over time and estimate laboratory bias and variability for inorganic constituents. For example, figure 2 shows plots obtained from the BQS website for cadmium during water year 2013 (October 2012–September 2013). Results in figure 2*A* are from the method used for filtered samples; results in figure 2*B* are from the method used for unfiltered samples. The data points indicate deviation of individual results from the "known" value, determined as the median for each batch of blind samples. These data are normalized using the known concentration and F-pseudosigma for that batch; therefore, results from multiple samples with various known concentrations can be displayed on the same scale. The upper chart shows a positive (though generally acceptable) bias through most of the year, becoming almost unbiased (close to zero) in September. Results for environmental samples analyzed during this time period probably will have a slightly high bias. Also, a decreasing trend for cadmium concentrations in environmental samples could represent an artifact of analytical bias rather than a true environmental change. The lower chart shows a slight negative, but acceptable, bias throughout the year. Together, the charts indicate that environmental-sample results for filtered samples could be biased higher than results for unfiltered samples, which could produce anomalous data with dissolved cadmium exceeding total cadmium.

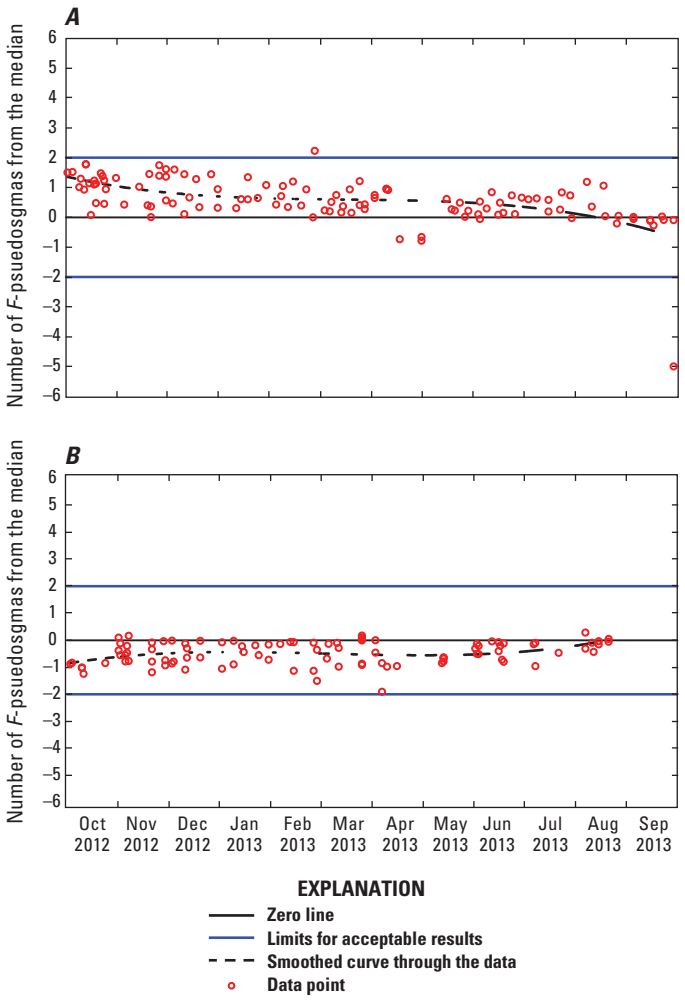


Figure 2. Example time-series charts of method performance for cadmium analysis at the U.S. Geological Survey National Water Quality Laboratory during October 2012–September 2013: *A*, results from the method used for filtered samples; *B*, results from the method used for unfiltered samples (data from http://bqs.usgs.gov/ibsp/FY13charts.shtml).

Organic Blind Sample Project (OBSP): The OBSP prepares samples spiked with known concentrations of organic constituents and submits them as blind samples for analysis at the NWQL. Data for over 600 organic analyses at the NWQL are evaluated for bias, variability, false positives, false negatives, and method performance over time. Figure 3 and table 4 show example results for the pesticide acetochlor from 2011 to 2014 obtained from the BQS website. Individual data points are in percent recovery, which is comparable for samples of various expected concentrations. Recovery was biased low until September 2012. Since then, recovery has averaged about 94 percent. Results for environmental samples collected during this period could be similarly biased. Reported concentrations are likely to be less than the actual value in the environment, particularly for samples analyzed prior to September 2012. An increasing trend for acetochlor concentrations in environmental samples during 2012 could represent an artifact of analytical bias rather than a true environmental change. Table 4 provides statistics about sample recovery, broken down by how the analytical result was reported. The number of false negatives (censored results in spiked samples) and false positives (quantified results in unspiked samples) also are listed. For the example, mean recovery of acetochlor was 87 percent with a standard deviation of 15 percent. There were no false negatives or false positives. This information can be used in evaluation of data quality for environmental samples collected during the time period shown in figure 3.

Blind Blank Project (BBP): The BBP submits blind blank samples to the inorganic and carbon sections at the NWQL, collecting contamination data on approximately 140 analytical determinations. Analytical results from the BBP are summarized weekly for the NWQL and are used to monitor for laboratory contamination. For example, figure 4 shows results obtained from the BQS website for chromium analyzed during water years 2012 and 2013 (October 2011–September 2013) using the inductively-coupled plasma method. These results include a number of false positives (values greater than the detection limit) with concentrations up to 0.2 micrograms per liter (µg/L) during June–November, 2012. The results of environmental samples analyzed during that time may be biased high by a similar amount. If field blanks analyzed during this time period had similar concentrations, the source of contamination could be from the laboratory rather than sample collection, processing, and shipping. Blanks analyzed before or after this period are unlikely to be affected by laboratory contamination; therefore, any positive bias in those blanks would more likely be due to a source in the field.

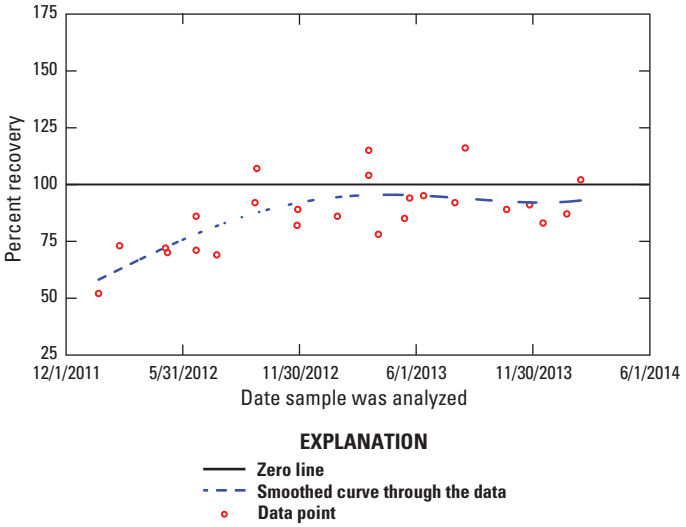


Figure 3. Example time-series chart of method performance for the pesticide acetochlor analyzed at the U.S. Geological Survey National Water Quality Laboratory (data from http://bqs.usgs.gov/OBSP/Current_Charts/Spikelevel_2001_ACETOCHLOR.html).

Table 4. Example statistics for acetochlor results from organic blind samples analyzed at the U.S. Geological Survey National Water Quality Laboratory (data from http://bqs.usgs.gov/OBSP/Current_Charts/Spikelevel_2001_ACETOCHLOR.html, accessed February 2014).

[--, no data]

Sample type	Number	Percent of total	Mean recovery	Standard deviation of recovery	False negatives		False positives	
					Number	Percent	Number	Percent
Spiked: result greater than or equal to reporting level	25	100	87	15	--	--	--	--
Spiked: result less than reporting level	0	0	--	--	--	--	--	--
Spiked: result censored	0	0	--	--	--	--	--	--
Spiked: total	25	--	--	--	0	0		
Not spiked	11	--	--	--	--	--	0	0

Analysis and Interpretation of Quality-Control Data

The QC data collected within a specified inference space can be used to estimate the bias and variability that might affect environmental samples collected within the same inference space. In this way, QC data can be considered a statistical "sample" which is used to make inferences about a population consisting of all possible QC and environmental data that were obtained in the same area during the same time using the same methods of collection, processing, and analysis. Thus, statistical methods can be used to analyze QC data in order to provide information on the potential bias and variability in environmental data. Methods are described herein for statistical analysis of blanks, spikes, and replicates. Some background is provided in the statistical concepts that underlie these methods, but for a more complete discussion, the reader is referred to Hahn and Meeker (1991) or Helsel and Hirsch (2002). Also, it must be noted that statistical estimation of bias and variability might not be appropriate in all situations, and that in this case other, non-statistical methods should be used.

Following an introduction to basic statistical concepts, guidance is provided for applying statistical and other interpretive methods to the analysis of blank, spike, and replicate data. Each subsequent section includes examples of the analysis and interpretation of QC data from published USGS reports. In each example, the project and QC sampling design are described, then the QC data analysis is summarized, and any implications of this analysis on the interpretation of environmental data are presented.

Statistical Concepts

It is not possible, physically or financially, to measure all occurrences of every characteristic of interest in environmental studies. For some characteristics, any direct measurement is impossible. Thus, statistical methods are necessary to make estimates of these characteristics. Such estimates can be less than satisfying, and even the subject of disbelief or derision. Mark Twain popularized a statement attributed to Benjamin Disraeli that "there are three kinds of lies: lies, damned lies, and statistics" (Twain, 1907). On the other hand, Twain's contemporary H.G. Wells is reported to have stated that "a certain elementary training in statistical methods is becoming as necessary for everyone living in this world today as reading and writing" (Wells, 1938). Regardless of which of these outlooks is accepted, statistical analysis is an important tool for turning hydrologic data into useful information.

A few basic statistical concepts provide adequate background for understanding the methods used to analyze QC data. Statistical analysis is based on individual observations or measurements that can be combined into a dataset, referred to as a "statistical sample." The theoretical set of all possible observations or measurements is called the "population." The statistical sample is a subset of these possible observations, selected and measured in a way such that conclusions about the sample can be extended to the entire population. Analysis of a statistical sample is predicated on the assumption that the observations are valid. For water-quality data, observations

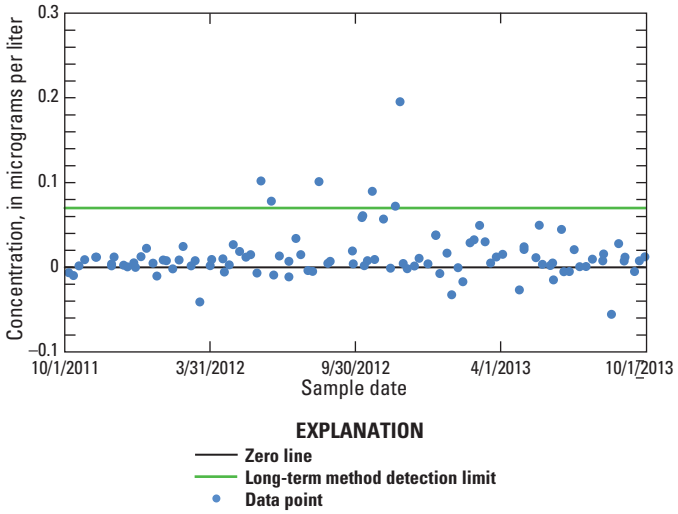


Figure 4. Example time-series chart showing blind-blank results for dissolved chromium in water analyzed at the U.S. Geological Survey National Water Quality Laboratory (current blind-blank charts are available at https://bqs.usgs.gov/ibsp/Blind%20Blank%20Charts).

must represent the environment from which the water sample was obtained. Use of standard sampling procedures, such as those described in the National Field Manual (U.S. Geological Survey, variously dated), and laboratory methods, such as those documented by the NWQL or the EPA, can help ensure that observed data are representative.

Statistical analyses can be divided into two broad categories: descriptive and inferential. Descriptive statistics provide summaries of the available data, for example:

- The mean concentration of nitrate in samples collected at a stream-sampling site,
- The maximum flood during 10 years (yr) of record, or
- The median pesticide concentration in samples collected from a well.

These are characteristics of a set of observations. By themselves, they do not provide any information about the population from which the observations were obtained. Inferential analysis is an attempt to estimate characteristics of a population that is not completely sampled. Some examples of inferential statistics are:

- The time-trend for nitrate concentration in a stream,
- The 1-percent annual exceedance probability, or 100-yr flood, at a stream site, or
- The difference between pesticide detections in two aquifers.

Inferential analysis is based on probability and uncertainty. Statistical probability is the likelihood or relative frequency of the occurrence of an event, expressed as a number from 0 to 1 or as a percentage from 0 to 100. Uncertainty is the variability of a statistically estimated value, expressed as imprecision or error.

Confidence Intervals

The uncertainty of an inferential statistic often is indicated by reporting a range of values, referred to as a "confidence interval." Confidence intervals are constructed to contain an unknown characteristic of the population, such as the mean, median, standard deviation, or a percentile, with a specified probability. The width of the confidence interval is the uncertainty due to estimation of a population characteristic based on sample data. For example, assume the mean concentration for a set of observations (the statistical sample) is 10 milligrams per liter (mg/L). This is a descriptive statistic. Based on this mean and an associated measure of variability, one might determine that the 90-percent confidence interval for the mean concentration in the entire population is between 8 and 12 mg/L. The population mean cannot be known exactly, but there is only a 10-percent chance that it is outside the range of plus or minus 2 mg/L from the sample mean. The 10-percent chance that the interval estimate is incorrect is the statistical probability of error. The range of the interval (8–12 mg/L) is the uncertainty in the estimated mean.

For any confidence interval, there are two possibilities:

1. The interval does not contain the true value of the population characteristic. This is an error. The probability of this error is defined as α, and is referred to as the "significance level."
2. The interval does contain the true value. This is correct. The probability of being correct is $1–\alpha$, referred to as the "confidence level."

The interval constructed for a specified significance level (α) is called the $100(1–\alpha)$-percent confidence interval. The significance level represents the risk we are willing to accept that the interval estimate is incorrect. For example, if α is selected to be 0.1, there is a 1-in-10 chance that the constructed 90-percent confidence interval will not contain the true population characteristic. Ten confidence intervals theoretically constructed from different samples (datasets) collected from the same population are plotted in figure 5. The true mean is included within the range of nine intervals but is outside the range of one interval. Unfortunately, the sample collectors cannot know the true mean, so they do not know whether their particular interval is the 1-in-10 that is incorrect; therefore, all must assume that their interval has a 10-percent risk of error.

Uncertainty is inversely related to the significance level; if the risk of an error is decreased, then uncertainty (the width of the interval estimate) will increase. For the same sample data, the 95-percent confidence interval will be larger, and thus will have more uncertainty, than the 90-percent confidence interval; however, the risk of error will decrease to 1-in-20 (5 percent).

Confidence Interval for the Mean

Confidence intervals for the population mean are constructed based on the mean of a sample of observations from the population, the standard deviation (a measure of variability) of the sample observations, and an acceptable level of risk that the interval will not contain the true population mean. Mathematically, this interval is computed as a 2-sided inequality:

$$\bar{x} - t_{(1-\alpha/2),(n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(1-\alpha/2),(n-1)} \frac{s}{\sqrt{n}} \qquad (1)$$

where

μ is the population mean,

$\bar{x}$ is the mean of a random sample of data,

s is the standard deviation of the sample data,

n is the sample size (number of observations),

α is the specified significance level (0 through 1), and

t is the percentage point of Student' t distribution for an area of $1–\alpha/2$ with $n–1$ degrees of freedom.

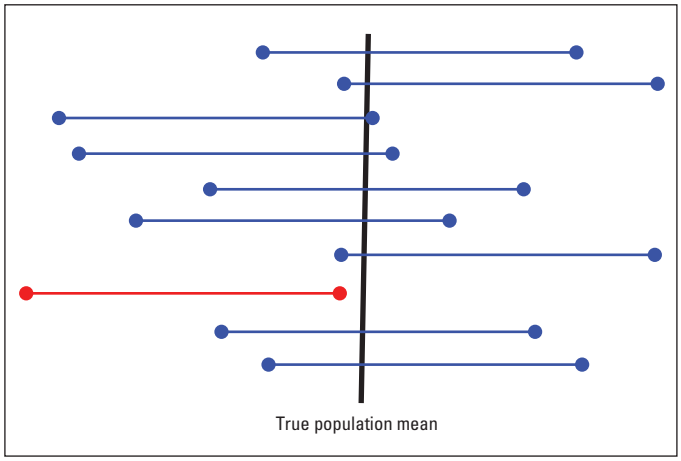


Figure 5. Theoretical confidence intervals constructed from 10 different sample datasets taken from the same population; the red line indicates an interval that does not contain the true mean.

In equation 1, standard deviation is estimated from the same data used to compute the mean. In some cases, the standard deviation of the population is known or can be assumed based on data from many previous samples. Using this known standard deviation (σ) requires a slightly different calculation of the confidence interval, with the percentage point of the standard normal curve (Z) in place of the Student's t statistic:

$$\bar{x} - Z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}} \qquad (2)$$

The Z value depends only on the specified significance level (α), not on the sample size (n).

Several characteristics of confidence intervals for the mean can be seen in equations 1 and 2. The interval is symmetric around the sample mean ($\bar{x}$), with one-half the potential error ($\alpha/2$) on each side. The values calculated on either side of the inequality, which define the range of the confidence interval, are referred to as the "confidence limits." There is equal probability that the true mean is less than the lower confidence limit (LCL) or greater than the upper confidence limit (UCL). Figure 6 illustrates these concepts by showing a 90-percent confidence interval plotted on a probability density function derived for a sample of 30 observations with a mean of 5 and a standard deviation of 6.5. The UCL and LCL are determined using equation 1 with 29 degrees of freedom ($t_{0.95,\,29} \approx 1.7$). For a 90-percent confidence interval, the significance level (α) is 0.10, so $1-\alpha/2$ is 0.95.

Confidence Interval for the Median

Confidence intervals can be constructed for many other statistics in addition to the mean. Each statistic requires a unique calculation. Many of these are described in Hahn and Meeker (1991). The median and other percentiles are statistics of particular importance in QC analyses. Confidence intervals on percentiles are calculated using the binomial probability function (B). For a given number of observations (n), ranked in ascending order of magnitude, the ranks of the upper (U) and lower (L) 100(1–α)-percent confidence limits of a specified percentile (p) can be determined by the inequality:

$$B\,(p,\,n,\,U\text{–}1) - B\,(p,\,n,\,L\text{–}1) \ge 1\text{–}\alpha \qquad (3)$$

The two binomial functions on the left side of this inequality indicate the probabilities that observed values less than rank U are no more than the population value of percentile p and that observed values less than rank L are less than the population value of percentile p. For a 90-percent confidence interval, the difference between these probabilities must be at least 0.90 (1–α). This condition is met if U is selected so that the value of the first function is at least 0.95 and L is selected so that the value of the second function is no more than 0.05. The resultant confidence interval for the population percentile is from the value of observation L to the value of observation U.

As an example, consider the median ($p = 0.5$) of 99 observed values. The sample median is simply the 50th ranked value; 49 values are less than the median and 49 are greater than the median. Constructing a 90-percent confidence interval requires finding the smallest rank U where B (0.5, 99, U–1) is at least 0.95 (95 percent probability) and the largest rank L where B (0.5, 99, L–1) is no more than 0.05 (5 percent probability). A series of upper and lower confidence limits are shown in table 5. As ranks are iterated from 51 to 59, probability increases from 58 percent to 96.5 percent. At rank 58, the probability is slightly less than 95 percent, so there is still more than a 5-percent chance that the true population median exceeds the value of the 58th ranked observation. In order to ensure no more than a 5 percent chance, the upper confidence limit for the median must be set at the 59th ranked observation. Similarly, there is a 5.4 percent chance that the true median is less than the 42nd ranked observation. The lower

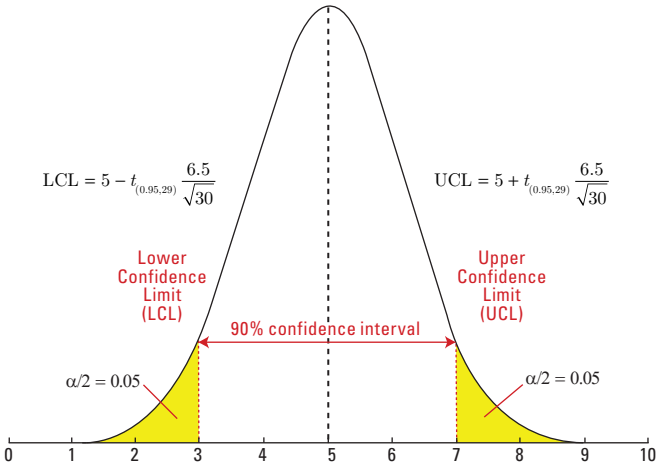


Figure 6. Probability density function for the distribution of a hypothetical mean and the associated 90-percent confidence interval (standard deviation = 6.5, number of observations = 30) (LCL, lower confidence limit; UCL, upper confidence limit; t, percentage point of Student's t distribution; %, percent; and α, significance level).

limit must be set at the 41st ranked observation so that the probability is less than 5 percent. The actual confidence level for the interval bounded by the 59th and 41st ranked observations is 96.5 – 3.5 = 93 percent. This 93-percent confidence interval is symmetric around the sample median with respect to rank, but not necessarily with respect to observed values.

Ranks of the LCL and UCL can be computed directly using the inverse binomial function (B^{-1}):

$$\text{rank } L = B^{-1}(p, n, \alpha/2) \quad (4)$$

$$\text{rank } U = B^{-1}(p, n, 1-\alpha/2) + 1 \quad (5)$$

For this example, $p = 0.5$ (the median), n (the number of observations) = 99, and $\alpha = 0.10$ for a 90-percent confidence interval. Then rank $L = B^{-1}(0.5, 99, 0.05) = 41$, and rank $U = B^{-1}(0.5, 99, 0.95) + 1 = 59$.

Upper Confidence Limit for a Percentile

In some analyses, only one confidence limit is of interest. For example, it might be important to not underestimate a specified percentile, but overestimation of that percentile is not of concern. In this case, only the upper confidence limit is needed, and all the error can be applied to the probability of underestimating the true percentile value. This is referred to as a "one-sided confidence limit" (or sometimes as an "upper confidence bound"). A one-sided upper confidence limit is determined so that:

$$B(p, n, U-1) \geq 1-\alpha \quad (6)$$

For example, the 75th percentile of 99 observations is between the 74th and 75th ranked values. To find the 1-sided upper limit with a least 90-percent confidence, ranks are iterated until the binomial probability is at least 0.90. Binomial probability at rank 80 is 89.1 percent, so there is more than a 10-percent chance that the observation at this rank will underestimate the true 75th percentile. However, at rank 81, this chance drops to 7 percent, so the one-sided upper confidence limit is the value of the observation at this rank.

Table 5. Binomial probability for selected ranks of 99 observations that the value observed at the specified rank is less than the population median.

Determination of upper confidence limit		Determination of lower confidence limit	
Rank	Binomial probability (percent)	Rank	Binomial probability (percent)
51	58.0	49	42.0
52	65.6	48	34.4
53	72.7	47	27.3
54	78.9	46	21.1
55	84.3	45	15.7
56	88.6	44	11.4
57	92.0	43	8.0
58	94.6	42	5.4
59	96.5	41	3.5

Confidence Limits for a Proportion

Proportions can be computed for a dataset by dividing the observations into groups, such as those less than or greater than a specified value. In water-quality analyses, proportions often are used to indicate the frequency of analyte detection, based on the proportion ($\hat{p}$) of quantified values (q) within the total number of observations (n) in a sample dataset. The sample proportion $\hat{p} = q/n$ is a point estimate of the unknown population proportion (ϕ). Confidence limits can be determined to provide an interval estimate of the population proportion (Hahn and Meeker, 1991, page 104):

$$\text{LCL} = \left\{1 + \frac{(n-q+1)F_{(1-\alpha/2,\, 2n-2q+2,\, 2q)}}{q}\right\}^{-1} \quad (7)$$

$$\text{UCL} = \left\{1 + \frac{n-q}{(q+1)F_{(1-\alpha/2,\, 2q+2,\, 2n-2q)}}\right\}^{-1} \quad (8)$$

These limits are based on an F-statistic with $100\alpha/2$ percent uncertainty that ϕ is either less than the LCL or greater than the UCL. The F distribution requires two values for degrees of freedom (df_1 and df_2). In equation 7, df_1 is $2n - 2q + 2$ and df_2 is $2q$; in equation 8 df_1 is $2q + 2$ and df_2 is $2n - 2q$. One-sided lower or upper confidence limits can be calculated by replacing $\alpha/2$ by α in either equation 7 or equation 8. (Note: in some statistics packages, values of the F-statistic are computed using $\alpha/2$ or α instead of $1-\alpha/2$ or $1-\alpha$ as shown above.)

As an example, assume that in a set of 20 water analyses, 5 results were quantified and the rest were censored (measured less than the detection limit). The sample proportion is 5/20 = 0.25, generally expressed as 25 percent. For a 90-percent confidence interval on the population proportion, $\alpha = 0.1$ and $1-\alpha/2 = 0.95$. The degrees of freedom for the F-statistic used in calculating the LCL (eq. 7) are: $df_1 = 2(20) - 2(5) + 2 = 32$, and $df_2 = 2(5) = 10$. From tables of the F-statistic (such as in Ott and Longnecker, 2001, p. 1105), $F_{(0.95,\, 32,\, 10)} = 2.690$. Substituting this and the values of n (20) and q (5) into equation 7 yields a lower limit of 0.104 (10.4 percent) for the population proportion. The degrees of freedom for the F-statistic used in calculating the UCL (eq. 8) are: $df_1 = 2(5) + 2 = 12$, and $df_2 = 2(20) - 2(5) = 30$. From tables of the F-statistic, $F_{(0.95,\, 12,\, 30)} = 2.092$. Substituting this value into equation 8 yields an upper limit of 0.456 (45.6 percent). Thus, based on the number of measurements and the observed number of detections, there is a 90-percent likelihood that the detection frequency would be between 10.4 and 45.6 percent in the collection of all possible samples.

Censored Data in Statistical Analyses

Water-quality data are reported as censored values if the measured result is less than the reporting level, defined by the laboratory for each method and analyte. Censored values can be problematic for statistical analyses, particularly if some

assumed value, such as zero or one-half the reporting level, is substituted for the censored result (see Helsel, 2005, for a detailed discussion). This problem is less severe for non-parametric methods, which rely on ranked results rather than specific data values. For example, confidence limits for medians and other percentiles are based on ranking data from low to high values. If the censoring level is the same for all data on any single analyte, the censored values are all given a tied rank lower than any quantified value. In this case, substitution of any value less than the lowest quantified result is appropriate for censored values. Non-parametric methods commonly are used for analysis of QC data. Simple substitution for censored values should not be used in other types of statistical analysis that are frequently applied to environmental-sample data.

Analysis and Interpretation of Data for Blanks

Blanks are used to estimate the positive bias that can be caused by extraneous contamination introduced into environmental samples during collection, processing, shipment, and laboratory analysis. Evaluation of data from field blanks depends on the inference space represented by the blanks. In general, there are two possibilities: (1) a single blank is prepared to represent potential sources of contamination that affect a specific, small set of environmental samples, or (2) multiple blanks are prepared periodically over time and space to represent potential sources of contamination that might affect a much larger set of environmental samples.

Evaluating Contamination Based on Single Blanks

Certain types of projects, including remedial investigations at Superfund sites and military installations, can require a sampling design consisting of a single blank that represents potential contamination in a specific set of environmental samples. In this design, there is a correspondence between each environmental sample and an associated field blank. If analytes are detected in the blank, then concentrations of these analytes in the associated environmental samples are evaluated to determine whether they should be considered valid. The EPA (U.S. Environmental Protection Agency, 1989) provides guidelines for comparing sample concentrations with blank concentrations. These guidelines state that for most analytes, quantified analytical results in environmental samples are considered valid only if the concentration exceeds five times the amount detected in the blank. Additionally, for common laboratory contaminants, such as acetone, methylene chloride (dichloromethane), or toluene (methylbenzene), environmental sample results are considered valid only if the concentration exceeds 10 times the amount detected in the blank. Project staff can use these guidelines but ultimately must be able to justify qualification of environmental data based on QC blank results.

Evaluation of contamination based on single blanks can lead to revision of environmental sample data from quantified values to nondetections (less than a reporting level). Revised data should be used in interpretive reports, and results can be qualified (using codes) but not changed in the USGS National Water Information System (NWIS) database (http://waterdata.usgs.gov/nwis; Dupre and others, 2013). Single blanks provide no estimates of extraneous contamination in environmental samples with concentrations that exceed the revised quantitation limits.

Blank Example 1: One Set of Blanks Associated with a Few Environmental Samples

In 2012, the USGS in cooperation with the Wyoming Department of Environmental Quality collected a sample from each of two monitoring wells near the town of Pavillion, Wyoming (Wright and others, 2012). These wells had been installed to test groundwater for potential effects of hydraulic fracturing of oil and gas wells in the area. Because results could be controversial, a large number of QC samples were included in the design. In the end, only one well could be sampled, so the project data consisted of two environmental samples (collected after different amounts of water had been purged from the well), along with replicates for each sample, various blanks, and two matrix samples spiked in the laboratory (Wright and others, 2012). A field blank was prepared immediately prior to collection of the first environmental sample. Source-solution blanks were prepared in the hotel, before traveling to the well site, and in the field (Wright and others, 2012, referred to this field-prepared source-solution blank as an “ambient” blank). A trip blank, prepared in the laboratory, was transported to the well site. All blanks and environmental samples were transported to the laboratory in the same container. The laboratory also provided results for a method blank. Thus, there were five blanks analyzed in association with the four environmental samples and replicates.

Results of all analyses were presented in a USGS Data Series report (Wright and others, 2012). Analytical results were less than reporting levels in all blank samples for 215 (92 percent) of the 234 constituents analyzed by the laboratory. Forty-three results (3.6 percent) for 17 constituents in the environmental samples and replicates had to be qualified because they were less than 5 times the maximum concentration in at least one of the associated field, ambient, or laboratory-method blanks. Data qualification involved including a qualifier code with the reported results for those constituents in data tables provided in the report. Qualifiers were not included for results that were less than the reporting level. Data values were not changed in the USGS NWIS database, but data users were cautioned in the report that these values should be treated as nondetections, and the reported concentration should be considered the quantitation limit for the analyte in that sample.

Blank Example 2: A Few Blanks Associated with a Set of Environmental Samples

Five USGS Water Science Centers participated in sampling water and sediment at 70 coastal sites on the Gulf of Mexico following the Deepwater Horizon oil spill in 2011 (Nowell and others, 2013). Samples were collected before and after the oil made landfall. These samples documented changes between "pre-landfall" and "post-landfall" conditions. The QC design included one field blank from each of the Centers during each of the two sampling periods. The intent was to associate each Center's blank with the environmental samples collected by that Center. All five field blanks were collected during the pre-landfall period, but only four were collected post-landfall. The QC data analysis had to be revised so that all blanks collected within each period were associated with all environmental samples from that period, and a detected concentration in any blank was considered evidence of potential contamination in every associated environmental sample. Although this approach might overestimate the extent of incidental contamination of some water samples, no other procedure would ensure that contamination would not be underestimated.

Results and data interpretation were reported by Nowell and others (2013). Analytical results were available for 168 constituents in at least 4 of the pre-landfall field blanks. There were 24 quantified results (detections), effecting a total of 21 constituents. Analyses of the 4 post-landfall field blanks included 584 total results for 146 constituents, of which 564 (97 percent) were censored values. There were 20 quantified values reported for 12 analytes. In addition, 31 trip blanks prepared during post-landfall sampling were analyzed at one of the laboratories. These had limited utility for comparison to environmental samples; however, quantified results reported for three analytes indicated some potential for contamination during laboratory processing and analysis. None of these three constituents was detected in field blanks analyzed at this laboratory.

For constituents detected in field or trip blanks, concentrations in environmental samples were censored at raised reporting levels on the basis of guidance from the EPA (U.S. Environmental Protection Agency, 1989, pages 5–16 and 5–17). This raised reporting level was set equal to five times the maximum concentration detected in any blank and was applied to results in all associated environmental samples. Quantified results that were less than this raised reporting level were changed to censored values (nondetections) and reported as less than the previously quantified value. For a few common laboratory contaminants (acetone, dichloromethane, diethyl phthalate, methyl ethyl ketone, and toluene), the reporting level was raised to 10 times the maximum concentration detected in the blank.

In addition, 10 constituents had one or more detections in laboratory blanks. Concentrations of these constituents in all environmental samples were more than five times the blank concentration, except for two nutrient constituents—ammonia-plus-organic nitrogen and phosphorus. Because a laboratory blank is associated with a particular set of environmental samples, censoring for laboratory-blank contamination was applied only to those environmental samples that were associated with a contaminated blank. Results for both nutrients subsequently were censored in more than one-half of the post-landfall samples, and phosphorus was censored in 2 of the 110 pre-landfall samples. (Note: censoring based on laboratory blanks is done prior to reporting sample data by some laboratories, including the NWQL, but not by all the laboratories used in this study.)

Table 6 lists the constituents that were affected by censoring on the basis of contamination in laboratory, field, and trip blanks. Eight organic compounds and two trace elements were left with no detections in either sampling period after blank-censoring. Four additional organic compounds were left with no detections in the pre-landfall period; benzene and ammonia-plus-organic nitrogen were left with no detections in the post-landfall period. Four other constituents were censored to some extent, though some results still were quantified; two of these constituents were left with only one quantified value during the post-landfall period. Overall, 223 out of a total of 1,189 results for 19 constituents were censored for data interpretation because of contamination in blanks, but 174 of these censored results were for only 5 constituents: toluene, ammonia-plus-organic nitrogen, mercury, dissolved organic carbon, and phosphorus.

It might have been possible to improve the QC design so that the effects on data interpretation would have been less severe. In a short-term project, such as this one, QC samples can identify whether contamination might have affected the environmental data, but there is little likelihood that potential sources of contamination can be identified in time to make any corrections. Thus, the effects of potential contamination on data quality could not have been eliminated, even if more blanks had been collected. However, a better QC design (even following the original design of one blank per Center per sampling period) could have provided more information about the potential effects of contamination in specific environmental samples, and fewer results might have needed to be censored.

Evaluating Contamination Based on Multiple Blanks

In many water-quality investigations, blanks are not collected in association with every environmental sample; instead, blanks are collected at a variety of environmental sampling sites throughout the study period. Potential contamination in environmental samples is estimated by statistical analysis of this set of multiple blanks. The critical assumption in this analysis is that the blanks represent exposure to exactly the same sources of contamination that could affect any environmental sample. After the samples have been collected, it is important to review the blank data to ensure that there are no obvious geographical or temporal patterns in analyte concentration. This review normally can be done by plotting the data.

Table 6. Constituents affected by censoring due to contamination detected in laboratory, field, and trip blanks (table 15 from Nowell and others, 2013).

[pre, pre-landfall; post, post-landfall]

Analyte	Time period	Samples	Results after laboratory-blank censoring		Results after field and trip-blank censoring			
			Quantified results	Detection frequency (percent)	Samples censored due to blanks	Quantified results	Detection frequency (percent)	100-percent censored[1]
Organic contaminants								
1,2,3,5-Tetramethyl-benzene	pre	60	1	2	1	0	0	yes [2]
1,2,3-Trimethyl-benzene	pre	60	2	3	2	0	0	yes [2]
1,2,4-Trimethyl-benzene	pre	60	3	5	3	0	0	yes [2]
Acetone	pre	62	5	8	5	0	0	yes [3]
Benzene	post	48	3	6	3	0	0	post only
Dichloromethane	pre	62	3	5	3	0	0	yes
Dichloromethane	post	48	4	8	4	0	0	yes
Dissolved organic carbon	pre	62	62	100	41	21	34	no
Diesel range organics	post	48	6	13	5	1	2	no
Ethylbenzene	pre	63	3	5	3	0	0	yes [3]
Naphthalene	post	48	1	2	1	0	0	yes [3]
Toluene	pre	63	15	24	15	0	0	pre only
Trichloromethane	pre	62	3	5	3	0	0	pre only
Xylene, *meta* plus *para*	pre	60	4	7	4	0	0	pre only
Xylene, *ortho*	pre	60	3	5	3	0	0	pre only
Trace and major elements, and nutrients in water								
Ammonia-plus-organic nitrogen as N	post	48	48	100	48	0	0	post only
Copper	post	48	22	46	3	19	40	no
Mercury	post	48	23	48	23	0	0	yes [2]
Phosphorus as P	post	48	48	100	47	1	2	no
Phosphorus as P	pre	68	55	81	2	53	78	no
Silver	pre	63	4	6	4	0	0	yes [3]

[1]Analytes that were 100-percent censored (no detections remaining) after blank-censoring; "pre only," detected in pre-landfall samples; "post only," detected in post-landfall samples; no, not fully censored.

[2]Data available only for one sampling period.

[3]Not detected (without blank-censoring) in the other sampling period.

For example, many field blanks were collected during sampling for the first cycle of the NAWQA Program. Although these blanks were collected by different sampling teams in different NAWQA study units, they were all collected using similar equipment and the same protocols. Thus, all these blanks were intended to represent a common inference space, and the distribution of analytical results from the blanks could be compared to results for all environmental samples. In order to check this assumption, summary statitics for blanks within each study unit were compared by using boxplots. Figure 7 shows results for dissolved ammonia in NAWQA field blanks collected at stream-sampling sites. These boxplots are arranged generally from west (on the left side of the graph) to east across the continental United States. There is no discernable spatial pattern in the blank results, so grouping all ammonia blanks into a single inference space seems justified. Blanks also need to be reviewed for differences over time. Figure 8 shows the time series of total Kjeldahl nitrogen (TKN) results in field blanks collected at stream-sampling sites within all NAWQA study units. In this case, obvious changes occurred in the distribution of results, beginning in late 1997. These changes were associated with decreases in the method detection limit that year and again in 1998 and 2000. Thus, the inference space for TKN in NAWQA field blanks must be divided into several parts, dependent on when the blanks were collected. Blank results from each inference space can be compared only to environmental samples collected during that same time period.

The next step is to evaluate the distribution of various constituent concentrations reported for field blanks within each identified inference space. The expected concentrations in a blank are zero (or less than the reporting level), so any non-zero concentration (or detection) is considered evidence of contamination. Because blanks are simply a subset of all samples within the inference space, this same distribution of contamination is assumed to affect environmental samples. The distribution of concentrations in blanks can be highly skewed; therefore, statistical techniques that rely on assumptions of normality are not applicable. These assumptions are avoided by evaluating the distribution as a series of percentiles.

The objective in analyzing data from blank samples is to characterize the frequency and magnitude of contamination in field blanks and then to infer how that distribution of contamination applies to environmental samples. This objective can be achieved by constructing an upper confidence limit (UCL) for a high percentile of contamination in the population

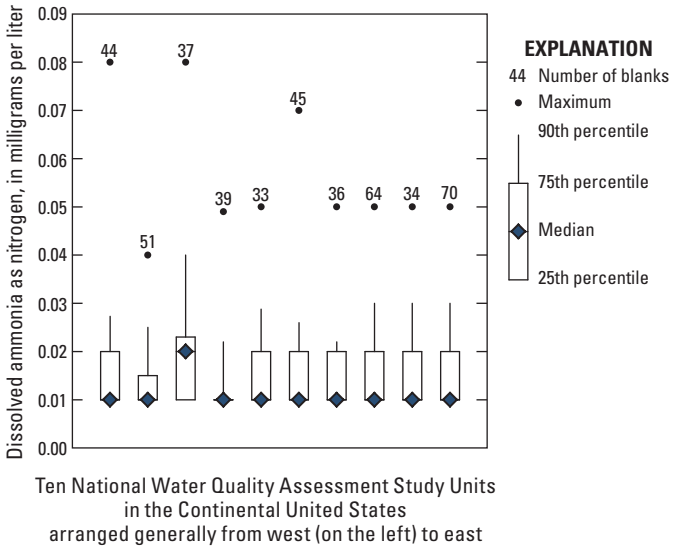


Figure 7. Example plot of the spatial distribution of ammonia concentrations in field blanks collected in 10 selected study units of the National Water Quality Assessment Program during 1992–2001 (from data compiled by Mueller and Titus, 2005).

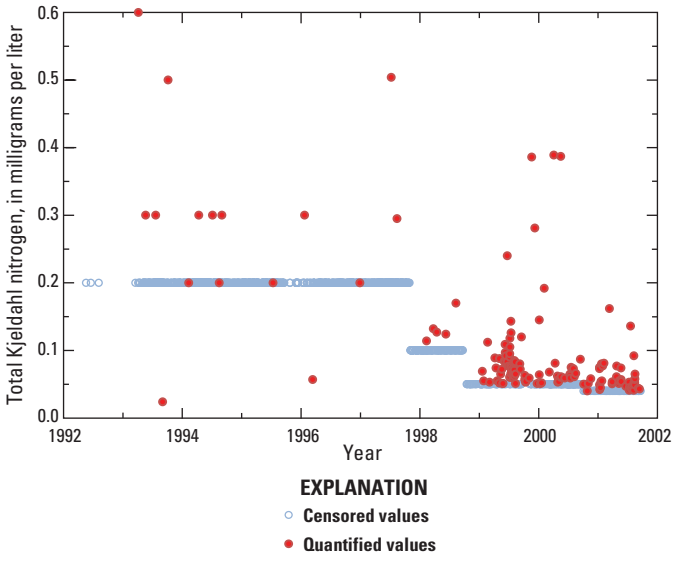


Figure 8. Example plot of the time series of total Kjeldahl nitrogen concentrations in field blanks collected as part of the National Water Quality Assessment Program during 1992–2001 (from data compiled by Mueller and Titus, 2005).

that includes blanks and environmental samples. This UCL is the maximum contamination expected in the specified percentage of samples. For example, the 90-percent UCL for the 95th-percentile concentration, estimated using data from field blanks, is the maximum contamination expected, with 90-percent confidence, in 95 percent of all samples in the population. The 90-percent confidence level indicates that there is only a 10-percent chance that the 95th-percentile of contamination has been underestimated. An alternative description is that there is 90-percent confidence that this amount of contamination would be exceeded in no more than 5 percent of all samples (including environmental samples) that were collected, processed, shipped, and analyzed in the same manner as the blank samples.

The binomial function, described in the Confidence Interval for the Median section of this report, can be used to determine a distribution-free UCL for a percentile. This method uses order statistics (based on ranking the data values from small to large) and binomial probability to determine the UCL. Equation 6 (repeated here as equation 9) is used to calculate the probability that no more than n minus U values from a total of n observations exceed the (100p)th percentile of the sampled population. The rank (U) is selected as the smallest integer such that:

$$B\,(p, n, U–1) \geq 1–\alpha \qquad (9)$$

where

α is the significance level.

The 100(1–α) percent UCL for the (100p)th percentile of contamination in the population then is determined by the measured value of the U ranked observation. For example, in a group of 100 blank samples, the 90-percent UCL for the 95th percentile can be determined as follows. First, find the smallest value of U that meets the criterion:

$$B\,(0.95, 100, U–1) \geq 0.90 \qquad (10)$$

For $U = 98$, $B = 0.882$, which is less than the criterion of 0.90, but for $U = 99$, $B = 0.963$; therefore, the 99th ranked observation is the smallest that meets or exceeds the criterion, and the 90-percent UCL for the 95th percentile is the concentration in the 99th ranked blank sample.

Figure 9 shows a conceptual example of the distributions of concentration in environmental samples and the 90-percent UCL for percentiles of concentration in field blanks. In determining plotting positions for both lines, censored values are assigned a concentration of one-half the reporting level. The 90-percent UCL for concentration in blanks is less than the reporting level (0.05 mg/L) up to the 89th percentile, and the UCL for the 95th-percentile concentration is about 0.07 mg/L. Potential contamination bias in the environmental samples is then estimated from the UCL calculated from the blank sample data. Contamination bias represents an extraneous amount of a constituent introduced during the sampling process and is in excess of any actual “contamination” present in the stream or groundwater. For the example in figure 9, contamination bias can be described as follows: Extraneous contamination is estimated, with at least 90-percent confidence, to be less than detection in at least 89 percent of all samples, and to exceed 0.07 mg/L in no more than 5 percent of all samples. This latter amount of extraneous contamination can affect 1 or 2 significant figures in concentrations reported for environmental samples up to 10 times the estimate made using field blanks (which would be 0.7 mg/L for this example). Figure 9 shows

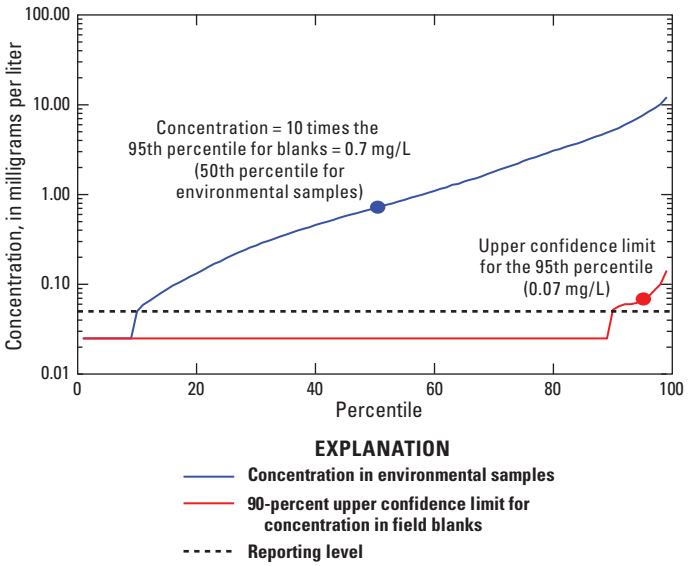


Figure 9. Conceptual example of the distributions of concentration in environmental samples and the 90-percent upper confidence limit for percentiles of concentrations in field blanks (mg/L, milligrams per liter; UCL, upper confidence limit).

that about 50 percent of reported concentrations in environmental samples were less than 0.7 mg/L, so extraneous contamination could be a substantial fraction of the reported result for many samples. However, it is also likely that extraneous contamination is negligible (less than detection) in 89 percent of all the samples. Therefore, extraneous contamination is likely to have affected no more than 11 percent of the environmental samples, and only those with concentrations less than 0.7 mg/L (thus, 5.5 percent of all samples). Those potentially affected samples should be qualified in data reporting and interpretation.

Determining the Number of Blanks to Collect

In the design phase of a project, the project staff must determine the number of field blanks that will be required to adequately estimate the potential for extraneous contamination in environmental samples. If the design is to associate a single blank with a set of environmental samples, then a blank must be collected with each set (generally those samples shipped in the same container). If the statistical approach will be used to determine percentiles of contamination, the number of field blanks should be determined based on the following two criteria:

- How much uncertainty is acceptable?
- What level of confidence is necessary?

For contamination bias, based on data from blanks, uncertainty is determined by the largest percentile (p) of contamination that can be evaluated.

$$\text{Uncertainty} = 1-p \qquad (11)$$

The extent of potential contamination cannot be known for higher percentiles. "Confidence" is the likelihood that this uncertainty has not been underestimated, and is based on the binomial probability for the selected percentile. Using equation 9, confidence can be determined:

$$\text{Confidence} = 100\,(1-\alpha) = 100\,[B\,(p, n, U-1)] \qquad (12)$$

For the percentile that can be determined from the maximum concentration reported for any field blanks, then rank (U) = sample size (n), and equation 12 can be solved to calculate the number of blanks based on the selected uncertainty (p) and confidence (α):

$$n = \frac{\log(\alpha)}{\log(p)} \qquad (13)$$

Because α and p are fractional values, the logarithms are negative and are larger (less negative) as α and p increase from 0 to 1. For a given level of confidence, the number of blanks (n) must be increased in order to achieve less uncertainty (make p larger). For a given percentile, n must be increased in order to increase confidence (make α smaller).

The number of blanks required to evaluate various percentiles of contamination are shown in figure 10 for three selected levels of confidence. For example, the 75th percentile of contamination can be estimated, with 90-percent confidence, as the maximum concentration reported for a set of eight blanks. The number of blanks must be increased to 22 in order to estimate the 90th percentile with 90-percent confidence. The 90th percentile also could be estimated with 16 blanks, but the level of confidence must be decreased to 80 percent. Increasing confidence to 95 percent would require 29 blanks.

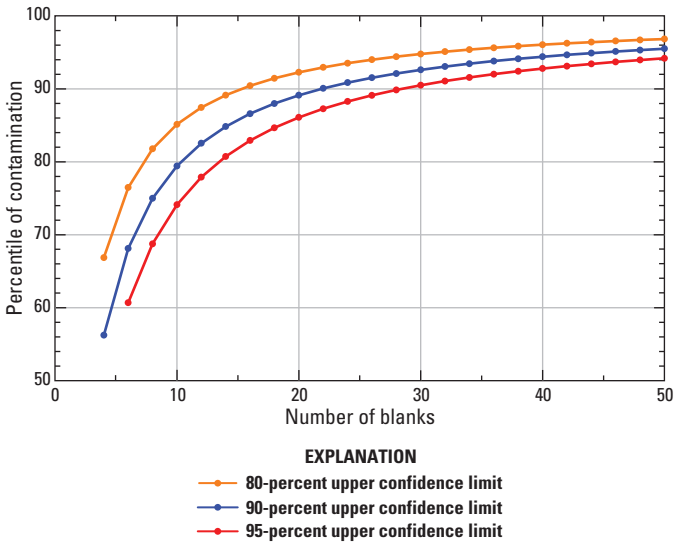


Figure 10. Number of blanks required to determine selected upper confidence limits for a specified percentile of contamination.

Blank Example 3: A Few Blanks Collected for More Than One Set of Environmental Samples

During 2006–07, the USGS Colorado Water Science Center conducted a study of the groundwater quality, age, and probability of contamination associated with land use in the Eagle River valley-fill aquifer upstream from Dotsero, Colorado (Rupert and Plummer, 2009). Samples were collected one time from 61 wells and quarterly for 1 yr from 10 surface-water sites. The QC samples included five field blanks prepared using groundwater sampling equipment and three field blanks prepared using surface-water sampling equipment. Table 7 presents a summary of laboratory results for selected constituents in the groundwater blanks and in the 61 primary environmental samples plus 2 replicates.

Too few blanks were available for a meaningful statistical analysis of the potential for extraneous contamination of samples collected for this project. With five blanks, the 90th percentile of contamination can be estimated with only 41 percent confidence, and the highest percentile that can be estimated with 90-percent confidence is the 63rd (eq. 12). Thus, the project investigators chose to compare the range of blank results to the range of environmental results. They state (Rupert and Plummer, 2009, p. 27):

> "The field-blank, replicate, and cation-anion balance data indicate that the major-ion and nutrient data are of high quality and suitable for quantitative analysis of ground- and surface-water quality. Concentrations of major ions and nutrients in the field blanks are mostly below laboratory reporting levels. Calcium, magnesium, nitrite, ammonia, boron, iron, and manganese were detected in a few field blanks, but in general, the concentrations were much smaller than those compounds detected in native and replicate samples. Analysis of the major-ion and nutrient field blanks indicated that the decontamination procedures of the sampling equipment were effective at cleaning the equipment between sampling sites and that there was no contamination occurring during sample collection, transport, and analysis that could affect the results of this report."

Additional interpretations of the QC data could have been made. For example, the range of field-blank results indicate there could be substantial quantitative uncertainty in low-range concentrations of iron and manganese reported for environmental samples.

Blank Example 4: A Few Blanks Collected for More Than One Set of Environmental Samples

Another project conducted by the USGS Colorado Water Science Center was an evaluation of the water quality of Vallecito Reservoir in southwestern Colorado. Samples were collected during 1999–2002 from the reservoir, its two major inflows, and its outflow at about monthly intervals from April through November (Ranalli, 2008). Nutrients and trace elements were analyzed in nine field blanks prepared with the equipment used to sample the reservoir, though not all constituents were analyzed in every blank. Potential extraneous contamination in environmental samples was estimated statistically using results from the blanks (table 8). All results for most constituents in the blanks were less than the reporting level. For these constituents, table 8 indicates that there is at least 90-percent confidence that extraneous contamination is no greater than the reporting level in at least 60–75 percent of all reservoir samples (depending on the number of blanks). Results for six constituents were greater than reporting levels. For example, nickel and zinc were detected in two of eight blanks. There is 90-percent confidence that extraneous contamination is no greater than the maximum concentrations of these constituents (2.6 µg/L for nickel and 8.3 µg/L for zinc) in at least 75 percent of all samples. The uncertainty is that contamination could be greater in up to 25 percent of all samples. At these concentrations, contamination could affect most reported results for nickel and zinc in many of the environmental samples from the reservoir. Therefore, environmental results for these constituents were considered potentially biased and were excluded from interpretation of project data.

The project investigator interprets these blank data as follows (Ranalli, 2008, p. 58):

> "The results of the analysis of the field blank samples can then be compared to environmental concentrations to determine the likelihood that contamination has affected the interpretation of the environmental data in samples collected from Vallecito Reservoir. For all constituents measured in Vallecito Reservoir, with the exception of total nickel and zinc, there is a high degree of confidence that contamination is not greater than the reporting level or is only slightly greater than the reporting level. However, because the environmental concentrations are close to the reporting level for all constituents and the range of samples that could be affected by contamination varies from 25 to 40 percent, some effects from contamination cannot be ruled out. For example, with respect to orthophosphate, there is 92 percent confidence that contamination is less than the reporting level in at least 75 percent of samples; however, the potential contamination in the remaining 25 percent is not known. Although there is no evidence that contamination accounts for any of the measured orthophosphate, environmental concentrations are low in comparison to the reporting level, and some effects from contamination cannot be ruled out. The environmental concentrations of total iron and total manganese, however, are usually at least 10 times greater than the reporting level. It is unlikely that contamination as great as 4 mg/L in 25 percent of all samples (total iron) and as great as 0.4 mg/L in 25 percent of all samples (total manganese) will affect the interpretation of the total iron and manganese concentrations in the environmental samples. The amount of potential contamination for total nickel and total zinc…exceeds the concentration of these constituents in most of the environmental samples so that contamination probably has compromised the interpretation of the environmental data for these two constituents."

Table 7. Summary of selected constituent data in field-blanks and environmental groundwater samples from the Eagle River valley-fill aquifer upstream from Dotsero, Colorado, 2006–07 (data from Rupert and Plummer, 2009, tables 2 and 6).

[N, nitrogen; mg/L, milligrams per liter; µg/L, micrograms per liter; n, number of samples; <, less than]

Constituent	Units	Reporting level	Field Blanks (n = 5)		Environmental Samples (n = 63)	
			Number of censored values	Range of quantified values	Number of censored values	Range of quantified values
Ammonia as N	mg/L	0.01	3	0.010–0.012	49	0.01–1.97
Nitrite plus nitrate as N	mg/L	0.03	5	<0.03	7	0.04–5.42
Calcium	mg/L	0.02	2	0.03–0.014	0	3.9–621
Magnesium	mg/L	0.01	4	0.01	0	0.72–321
Boron	µg/L	0.9	1	1.9–2.6	0	5–332
Iron	µg/L	3	3	3.1–10.8	12	3–8,364
Manganese	µg/L	0.1	3	0.20–0.34	9	0.1–1,500

Table 8. Summary of upper confidence limits for contamination by nutrients and trace elements in specified percentiles of samples from Vallecito Reservoir, near Bayfield, Colorado, based on data for field blanks (data from Ranalli, 2008, table 10).

[mg/L, milligrams per liter; µg/L, micrograms per liter; <, less than]

Constituent	Number of blanks	Achieved level of confidence	Evaluated percentile of contamination	Upper confidence limit[1]
Ammonia, dissolved, mg/L	9	92-percent	75th	<0.015
Nitrate + nitrite, dissolved, mg/L	9	92-percent	75th	**0.014**
Ammonia + organic nitrogen, dissolved, mg/L	9	92-percent	75th	**0.233**
Ammonia + organic nitrogen, total, mg/L	9	92-percent	75th	<0.2
Phosphorus, total, mg/L	9	92-percent	75th	<0.008
Phosphorus, dissolved, mg/L	9	92-percent	75th	<0.006
Orthophosphate, dissolved, mg/L	9	92-percent	75th	<0.007
Aluminum, dissolved, µg/L	8	90-percent	75th	<15
Aluminum, total, µg/L	8	90-percent	75th	<10
Arsenic, dissolved, µg/L	7	92-percent	70th	<1
Arsenic, total, µg/L	8	90-percent	75th	<1
Cadmium, dissolved, µg/L	7	92-percent	70th	<0.3
Cadmium, total, µg/L	8	90-percent	75th	<0.3
Chromium, dissolved, µg/L	7	92-percent	70th	<0.4
Chromium, total, µg/L	8	90-percent	75th	<0.4
Copper, dissolved, µg/L	7	92-percent	70th	**3**
Copper, total, µg/L	8	90-percent	75th	<2
Iron, dissolved, µg/L	8	90-percent	75th	<4
Iron, total, µg/L	8	90-percent	75th	<4
Lead, dissolved, µg/L	7	92-percent	70th	<2
Lead, total, µg/L	8	90-percent	75th	<2
Manganese, dissolved, µg/L	8	90-percent	75th	<2
Manganese, total, µg/L	8	90-percent	75th	<0.4
Mercury, dissolved, µg/L	5	92-percent	60th	<0.2
Mercury, total, µg/L	8	90-percent	75th	<0.2
Nickel, dissolved, µg/L	7	92-percent	70th	<0.4
Nickel, total, µg/L	8	90-percent	75th	**2.6**
Potassium, dissolved, µg/L	7	92-percent	70th	<0.2
Potassium, total, µg/L	8	90-percent	75th	**0.22**
Silver, dissolved, µg/L	7	92-percent	70th	<0.8
Silver, total, µg/L	8	90-percent	75th	<0.8
Zinc, dissolved, µg/L	8	90-percent	75th	<2
Zinc, total, µg/L	8	90-percent	75th	**8.3**

[1]Values in bold are greater than the reporting level.

Blank Example 5: Many Blanks Collected for a Large Program

A VOC dataset, compiled from NAWQA samples collected during 1996–2008, contained more than 7,000 environmental samples, 704 field blanks, and 472 source-solution blanks (table 9). In addition, 5,167 laboratory set blanks were analyzed for VOCs during this time period. The QC design was nationally consistent, so environmental and QC samples were collected using similar equipment and procedures at each type of sampling site (surface water, domestic and public-supply wells, or monitoring wells), and it is valid to group national results. Bender and others (2011) compared the statistical distributions of concentrations in environmental samples, field blanks, source-solution blanks, and laboratory set blanks, in order to place each compound into one of four "contamination categories" (table 10). This comparison was made using graphs of the distributions based on the conceptual model previously shown in figure 9.

The assumptions of this statistical comparison are that the field blanks and environmental samples are affected by the same potential sources of extraneous contamination and the same potential to experience a certain magnitude of contamination. These assumptions seemed valid for many, but not all, VOCs. The contamination categories listed in table 10 were based on the relation between the distribution of concentrations in field blanks (based on the 90-percent UCL for each percentile) and the distribution of concentrations in environmental samples, as illustrated by the examples in figure 11. Compounds in category 1, 2, or 3 are considered to meet both assumptions, though the effects of contamination bias differ among categories. Compounds in category 4 do not meet the second assumption; the effect of contamination on field blanks seems greater than on environmental samples, so field blanks were considered to be non-representative of the potential sources of contamination for these compounds.

Table 9. Number of samples collected for analysis of volatile organic compounds by the National Water-Quality Assessment Program during October 1996 through December 2008 (Bender and others, 2011, table 2).

Sample medium	Site type	Environmental samples	Field blanks	Source-solution blanks
Surface water	Surface water	1,497	129	54
Groundwater	Domestic and public-supply wells	3,042	278	225
Groundwater	Monitoring wells	2,639	297	193
Total		**7,178**	**704**	**472**

The example plots shown in figure 11 are for VOCs in samples from domestic and public-supply wells (Bender and others, 2011, p. 17 and 22). Fifty-four of the 87 analyzed VOCs were not detected in any field blanks (contamination category 1), as illustrated by lack of a red line in the graph of 1,1-dichlorethene (fig. 11*A*). Quantified results for these 54 VOCs in environmental samples are considered to be free of contamination bias. Therefore, data interpretation based on VOC concentrations in environmental samples from domestic and public-supply wells was considered valid. Thirty-three VOCs were detected in at least one field blank. For 10 of these, the distribution of concentrations in field blanks was

Table 10. Description of contamination categories and the potential for contamination bias in environmental samples based on the relation between the 90-percent upper confidence limit for percentiles of concentrations in field blanks and the distribution of concentrations measured in environmental samples (Bender and others, 2011, table 3).

Contamination category	Relation between field blanks and environmental samples	Interpretation of the potential for contamination in environmental samples	Number of compounds in category		
			Domestic and public-supply wells	Monitoring wells	Surface-water sites
1	No detections in any of the field blanks.	Quantified results for environmental samples are essentially free of contamination bias.	54	43	54
2	Detections in field blanks, but the distribution is lower (at least an order of magnitude) and negligible in comparison to concentrations in environmental samples.	Quantified results for environmental samples with larger concentrations are not markedly affected, but low concentrations might be affected by contamination.	10	8	16
3	Detections in both field blanks and environmental samples and the distributions of concentrations are similar (within an order of magnitude).	Quantified results for environmental samples are likely affected by contamination bias.	7	7	10
4	Detections in field blanks have a distribution of concentrations markedly higher (at least an order of magnitude) than the concentration distribution of environmental samples.	The potential for contamination bias in environmental samples cannot be determined by this method. Field blanks are considered non-representative of the potential sources of contamination to the environmental samples.	16	29	7

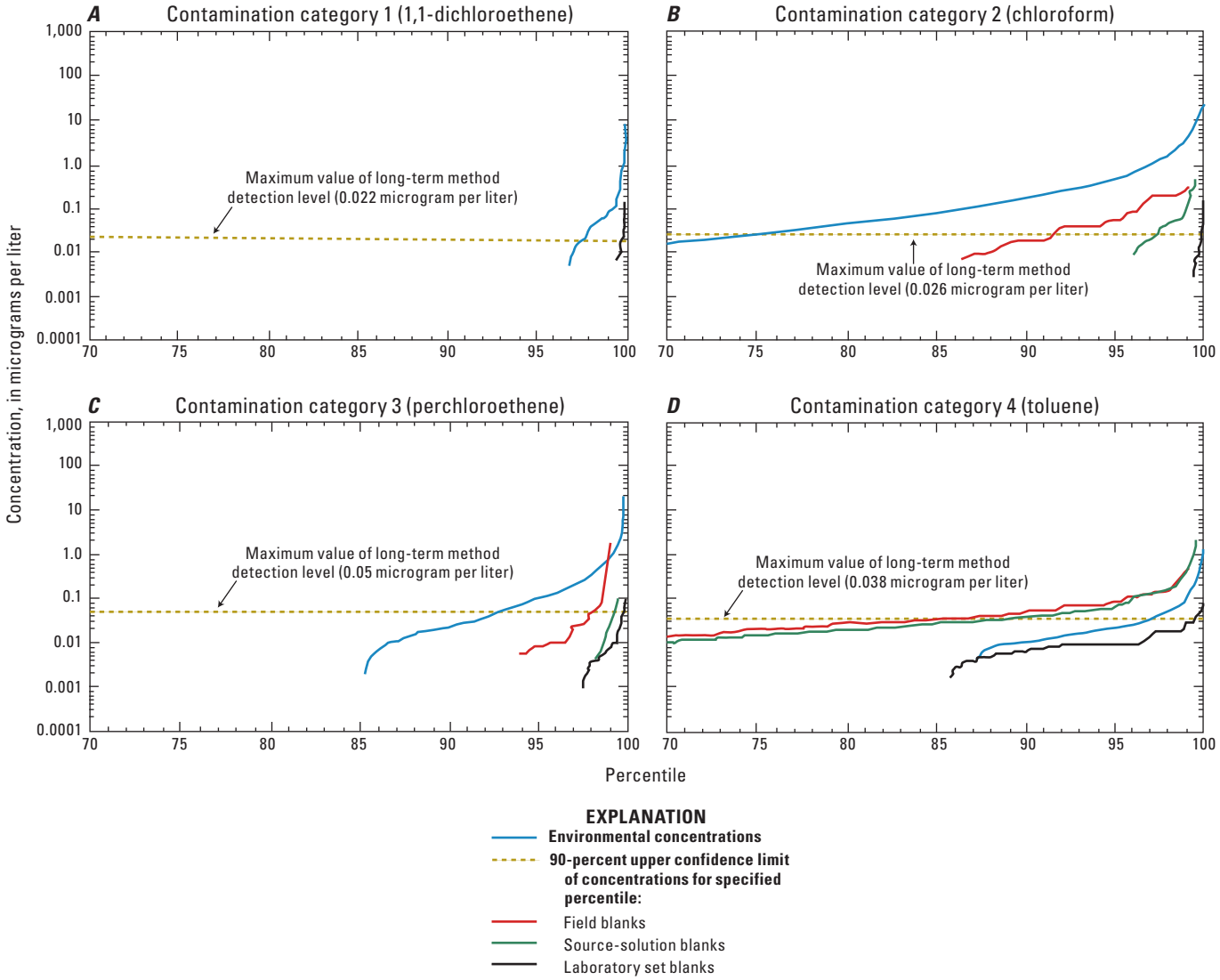


Figure 11. Examples of the distribution of volatile organic compound (VOC) concentrations in environmental samples from domestic and public-supply wells, and in field, source-solution, and laboratory set blanks: *A*, contamination category 1 (1,1-dichlorethene); *B*, contamination category 2 (chloroform); *C*, contamination category 3 (perchloroethene); and D, contamination category 4 (toluene). Nondetections are used to determine percentiles, but are not shown on the graphs. (From Bender and others, 2011, fig. 4).

at least an order of magnitude smaller than the distribution in environmental samples (contamination category 2), as illustrated by the graph of chloroform (fig. 11*B*). Low concentrations of these 10 VOCs in environmental samples might be affected by contamination, but bias in results for larger concentrations can be considered negligible. For seven VOCs, the distribution of concentrations in field blanks was similar to the distribution in environmental samples (contamination category 3), as illustrated by the graph for perchloroethene (fig. 11*C*). Quantified results for these VOCs in environmental samples are likely to be affected by extraneous contamination, and any interpretation of these data requires substantial qualification. For 16 VOCs, the distribution of concentrations in field blanks was markedly greater than the distribution in environmental samples (contamination category 4), as illustrated by the graph for toluene (fig. 11*D*). This graph also shows that source-solution blanks were contaminated at similar levels as field blanks, indicating that the blank water was contaminated by some source that did not affect the environmental samples. The potential for bias in environmental samples cannot be determined for these category-4 VOCs because the field blanks are considered non-representative of the potential sources of contamination. In the specific example of toluene, the laboratory-blank data could be used to estimate a minimum potential for bias (without consideration of contamination from field sources). Contamination from laboratory sources was not necessarily a similar problem for other compounds in category 4.

Analysis and Interpretation of Data for Spikes

Spikes are used to estimate the positive or negative bias that can affect the measured results for environmental samples because of analyte degradation or problems with the analytical methods. This bias is estimated by determining the recovery of known concentrations of the analytes in the spiked sample. Calculation of recovery for matrix spikes requires a separate environmental sample to determine the background concentration of the analyte in the unspiked matrix. Recovery in the spiked matrix samples can be compared to some criteria or to typical recovery for the analytical method based on laboratory reagent spikes.

Calculating Spike Recovery

Spike recovery is calculated from the concentration of an analyte in a spiked matrix sample in comparison to the concentration in a background environmental sample:

$$R = \frac{C_{spike} - C_{env}}{C_{expected}} \times 100 \qquad (14)$$

where

R is recovery, in percent,

C_{spike} is the concentration of the analyte in the spiked matrix sample, in micrograms per liter,

C_{env} is the concentration of the analyte in the background environmental sample, in micrograms per liter, and

$C_{expected}$ is the concentration of the spiked analyte expected in the sample, in µg/L.

The expected concentration of the analyte that was spiked into the matrix sample is calculated from the concentration in the spike solution, the volume of spike solution added, and the volume of the matrix sample:

$$C_{expected} = \frac{V_{sol} \times C_{sol}}{V_{sample}} \qquad (15)$$

where

V_{sol} is the volume of spike solution added to the sample, in milliliters (mL),

C_{sol} is the concentration of the analyte in the spike solution, in micrograms per milliliter (µg/mL), and

V_{sample} is the volume of the matrix sample, in liters (L).

Note that in equation 15 the volume of the spike amount is in milliliters, though sometimes the volume reported on field forms or in databases is in microliters, and that the sample volume is in liters, though it might be reported in databases in milliliters. Recovery for spiked reagent samples is calculated using the same equations (14 and 15), except the background concentration in equation 14 is assumed to be zero.

Results for analyte concentrations in environmental samples are often reported as a censored value (less than a reporting level). When these results need to be used in the spike-recovery calculation (eq. 14), they are re-coded either as zero, as the reporting-level value, or as some fraction (for example one-half) of the reporting level. Generally there is little difference in the recovery calculated using any of these recoded values; however, it is good practice to compute a range of recoveries by setting the background concentration to zero for one end of the range and to the reporting-level value for the other end. If the difference is negligible, then either (or some mid-point value) can be used for subsequent analysis.

As an example of recovery calculation, consider the analytical results for atrazine in the following paired samples:

- Environmental: 0.05 µg/L
- Field-matrix spike: 0.14 µg/L

The field matrix spike was prepared by adding 100 µL of a spike solution containing 1.0 µg/mL of atrazine to a 932 mL sample. The expected concentration is calculated using equation 15:

$$C_{expected} = \frac{100\ \mu L\left(\frac{1\ mL}{1000\ \mu L}\right) \times 1.0\ \frac{\mu g}{mL}}{932\ mL\left(\frac{1\ L}{1000\ mL}\right)} = 0.11\ \mu g/L \qquad (16)$$

Recovery is calculated using equation 14:

$$R = \frac{0.14 - 0.05}{0.11} \times 100 = 82\ \text{percent} \qquad (17)$$

This calculated recovery can be compared to the laboratory control limits for atrazine, which were plus or minus 30 percent (70 through 130 percent) for this example. Because recovery of atrazine in the matrix spike is within these control limits, there is no evidence of additional bias due to the sample matrix. However, a single matrix spike provides no information on the uncertainty of the estimated recovery, so it is common to collect multiple matrix spikes in order to determine a confidence interval.

Evaluating Recovery Bias using Multiple Spikes

Five examples of recovery are listed in table 11. Example 1 shows the typical case where all recoveries are (close to) 100 percent. Method performance is good, and there is no evidence of a matrix effect or analyte degradation. Results from analysis of environmental samples are unlikely to be biased because of either of these conditions. Example 2 shows low recovery in both matrix spikes but not in either reagent spike. This pattern indicates a matrix effect. Example 3 shows low

recovery in both field spikes but in neither laboratory spike. This is the typical pattern for analyte degradation. Example 4 shows the lowest recovery in the field matrix spike and better, but still low, recovery in the laboratory matrix spike and the field reagent spike. This indicates a combination of matrix and degradation effects. Both of these effects influence recovery in the field matrix spike. The laboratory matrix spike is influenced by the matrix effect but is not influenced by analyte degradation. The field reagent spike is influenced by degradation, but not the environmental matrix. Finally, example 5 shows low recovery in all spikes. This indicates that analytical performance is less than optimal, but there is no effect from the sample matrix or analyte degradation.

If only field-matrix samples are collected for a project, comparison is possible only with laboratory reagent spikes, prepared routinely to measure method performance. An example of such a comparison for three pesticide analytes is shown in table 12. Recovery of atrazine in the field matrix spike is slightly high, but not so high as to cause concern. Recovery of diazinon is low, but within the 95-percent confidence interval of the laboratory reagent spikes; therefore, this probably represents normal method performance rather than any field issue. Recovery of malathion is low and outside the 95-percent confidence interval of the laboratory reagent spikes. This might indicate a matrix effect or analyte degradation. Other types of spikes are needed to determine which of these effects is occurring. Before committing project resources, it is wise to consult the laboratory chemists, who might be able to provide some insight as to which effect is more likely. Also, additional field matrix spikes should be collected to confirm the low recovery.

Confidence intervals for recovery can be estimated by collecting multiple matrix spikes during a project. For example, concentrations of the pesticide chlorpyrifos in six field matrix spikes and corresponding environmental samples are shown in table 13. The matrix spikes were prepared by adding 100 µL of a spike mixture containing 1.0 µg/mL of chlorpyrifos to the sample volumes listed in the table. Expected concentrations for the spiked samples were computed using equation 15, and spike recoveries were computed using equation 14. Mean recovery is 45.5 percent with a standard deviation of 9.4. The 95-percent confidence interval for mean recovery, determined using equation 1 with a t value of 2.57 for 5 degrees of freedom, is 35.4–55.6 percent. This entire range is less than the lower control limit for laboratory spikes (59 percent for the NWQL in 2013). The low bias in recovery from field spikes indicates the possibility of either analyte degradation between sampling and analysis or interference from something in the sample matrix. Because chlorpyrifos is not known to degrade during normal sample holding time, a matrix effect is the more likely cause. This finding would warrant additional investigation because matrix interference might not be the same for all sampling sites or even for all times of sample collection at a single site. This example illustrates the need to evaluate QC data as they become available so that pertinent changes can be made to the sampling design.

Determining How Many Spikes to Collect

In the design phase of a project, the project staff must determine the number of field spikes that will be required to adequately estimate the potential for low or high recovery bias in environmental samples. A statistical approach is appropriate even if only a few spikes are collected, as long as the distribution of recovery values is approximately normal. The number of spikes required to meet project objectives should be determined based on the following two criteria:

- What level of confidence is necessary?
- How much uncertainty is acceptable?

Table 11. Examples of recovery in spiked samples.

Spike type	Recovery (percent)				
	Example 1	Example 2	Example 3	Example 4	Example 5
Field matrix spike	100	25	25	25	25
Laboratory matrix spike	100	25	100	50	25
Field reagent spike	100	100	25	50	25
Laboratory reagent spike	100	100	100	100	25

Table 12. Example of recovery of three pesticide analytes in one field matrix spike and 40 laboratory reagent spikes.

Compound	Recovery in field matrix spike (percent)	Recovery in Laboratory Reagent Spikes (percent)			
		Mean recovery	Standard deviation	Lower 95-percent confidence limit	Upper 95-percent confidence limit
Atrazine	109.8	102.2	2.7	96.9	107.6
Diazinon	83.0	82.2	1.1	80.0	84.5
Malathion	56.2	93.2	16.6	60.1	126.3

Table 13. Example of analyte recovery in a set of field matrix spikes and environmental samples.

[µg/L, micrograms per liter; mL, milliliter]

Chlorpyrifos (µg/L)		Spiked sample volume (mL)	Expected concentration (µg/L)	Recovery (percent)
Environmental sample	Spiked sample			
0.005	0.055	946	0.106	47.3
0.009	0.066	966	0.104	55.1
0.006	0.058	930	0.108	48.4
0.002	0.032	921	0.109	27.6
0.003	0.050	954	0.105	44.8
0.005	0.056	975	0.103	49.7

Confidence is the likelihood that potential low recovery bias has not been overestimated and potential high recovery bias has not been underestimated. This confidence is determined by constructing an interval about the mean recovery. If a 90-percent confidence interval is constructed, there is only a 10-percent chance that the true recovery is outside the interval. Because confidence intervals are symmetric about the mean, a general description is

$$CI = \bar{x} \pm d \tag{18}$$

where

CI is the overall width of the confidence interval,

$\bar{x}$ is the mean recovery from field spikes, in percent, and

d is the half-width of the confidence interval.

Uncertainty is the half-width of the interval; the actual bias due to low or high recovery could be any value between the low and high bound. From equation 2, the half-length is determined as

$$d = Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}} \tag{19}$$

where

σ is the standard deviation of recovery known from previous analysis (for example, a large number of laboratory spikes).

Solving equation 19 for n yields:

$$n = \left(\frac{Z_{(1-\alpha/2)}\sigma}{d} \right)^2 \tag{20}$$

Equation 20 is used to determine the number of spikes required to achieve an uncertainty of d with a confidence of $1-\alpha/2$. In order to decrease uncertainty without changing confidence ($1-\alpha/2$), the number of spikes (n) must be increased.

The number of spikes required to achieve various percentages of uncertainty in recovery are shown in figure 12 for three selected levels of confidence. For this plot, the value of σ in equation 20 was set to 13 percent, which was the median standard deviation for recovery of 50 pesticides in laboratory spikes analyzed at the NWQL during 2013. Five spikes are required to estimate mean recovery within about plus or minus 10 percent at the 90-percent confidence level. Increasing the sample size to 18 will decrease the uncertainty to plus or minus 5 percent.

Examples of Analyzing Spikes

Typically, an adequate number of spikes can be collected to determine mean recovery, even for small projects. Thus, statistical procedures can be used in most cases. However, in rare cases, projects include only one or a few samples, and a different approach is needed to evaluate potential recovery bias.

Spike Example 1: One Set of Spikes Associated with One Set of Environmental Samples

The Pavillion project was described previously in "Blank Example 1" in "Evaluating Contamination Based on Single Blanks." Samples collected from the monitoring well included two environmental samples (collected after different amounts of water had been purged from the well), along with replicates for each sample, various blanks, and two laboratory matrix spikes (Wright and others, 2012). All samples were analyzed by a non-USGS laboratory. Field spike kits were not available, so matrix samples were spiked in the laboratory.

The laboratory provided recovery percentages for the spiked matrix samples, but provided no information on typical recoveries expected for the analytical method. Project staff decided that recoveries within 70–130 percent were acceptable. If recovery of a constituent was outside this range in either matrix spike, results for that constituent in all samples were flagged with a data qualifier. Recoveries were available for 210 constituents, and were within 70–130 percent for

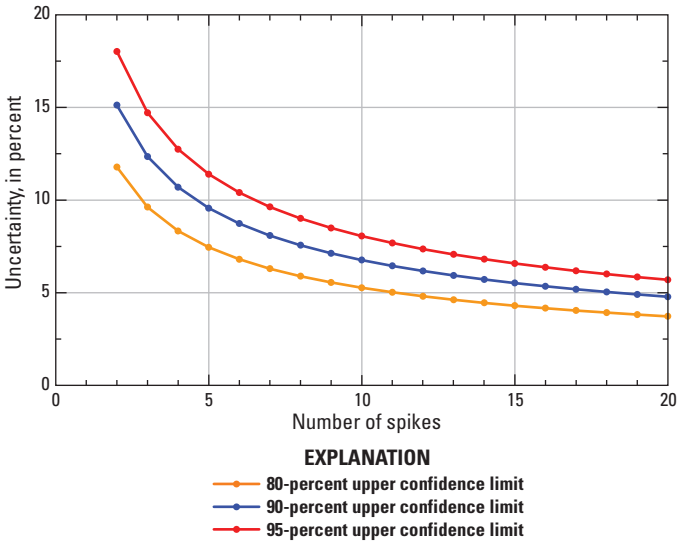


Figure 12. Number of spikes required to determine selected confidence limits for uncertainty in mean analyte recovery, based on a known standard deviation of 13 percent.

195 (93 percent) of those constituents in both spikes. For the other 15 constituents in the two environmental samples and two replicates, 42 results were qualified as having potentially low bias, because recovery was less than 70 percent in one or both spikes, and 16 results were qualified as having potentially high bias, because recovery was greater than 130 percent in one or both spikes (Wright and others, 2012, p. 21). These data qualifiers were included in the report (although not in the database); no interpretation of the environmental data was provided.

Spike Example 2: A Few Sets of Spikes Associated with More than One Set of Environmental Samples

The Gulf of Mexico oil spill project was described previously in "Blank Example 2" in "Evaluating Contamination Based on Single Blanks." Each of the five USGS Water Science Centers collected one matrix sample, for laboratory spiking, during each of the two time periods. The pre-landfall samples were spiked at the USGS NWQL and analyzed for 85 organic compounds. Mean recovery for individual analytes ranged from about 52 to 134 percent. The post-landfall samples were spiked at a non-USGS laboratory and analyzed for 107 compounds. Mean recovery for individual analytes ranged from about 19 to 124 percent. The two laboratories had 41 analytes in common, so recovery results were available for a total of 151 compounds. For the common analytes, differences in recoveries were generally small.

As in the previous Pavillion example, information on typical recoveries for these analytes was unavailable, so the project staff decided that acceptable limits were 70–115 percent. Mean recovery was within these limits for 134 (89 percent) of the 151 compounds. Mean recovery was less than 70 percent for 8 compounds in the spikes analyzed at the NWQL and for an additional 8 compounds in spikes analyzed by the other laboratory. Because recoveries were low, concentrations reported for these compounds in environmental samples could underestimate the true concentrations. Mean recovery for one compound exceeded 115 percent, so concentrations reported in environmental samples could be somewhat higher than the true concentrations. Concentrations were not recovery-corrected in the database, but recovery bias was noted in data tables in the interpretive report (Nowell and others, 2013, p. 32–33).

Spike Example 3: Many Spikes Collected for a Large Program

The VOC dataset compiled for the NAWQA program was described previously in "Blank Example 5" in "Evaluating Contamination Based on Multiple Blanks." A subset of spiked samples from this dataset was analyzed to evaluate VOC recoveries in groundwater and surface-water samples collected during 1997–2001 (Rowe and others, 2005). Results were available for 85 VOCs in a total of 428 spiked samples, including 149 field matrix spikes, 107 field matrix spike replicates, 20 laboratory matrix spikes, and 152 laboratory reagent spikes.

Spiked sample results were examined graphically and using a chi-square goodness-of-fit test (Ott and Longnecker, 2001) to determine whether recoveries were normally distributed. Although the distribution of VOC recoveries generally followed a normal distribution for all spike types, some extreme outliers were apparent in the data. Thus, median, rather than mean, recovery was used as a measure of central tendency. For all spike types, 87 percent of the individual VOC recoveries were within the range of 60 to 140 percent, which was considered acceptable for the analytical method. Median recoveries for 85 individual VOCs ranged from about 63–102 percent for field matrix spikes and field matrix spike replicates, 102–135 percent for laboratory matrix spikes, and 91–119 percent for laboratory reagent spikes.

Rowe and others (2005) evaluated the potential for analyte degradation by comparing recoveries in field-matrix and laboratory-matrix spikes. Based on the nonparametric Wilcoxon Rank-Sum Test (Ott and Longnecker, 2001), median recovery in field matrix spikes was significantly lower than in laboratory matrix spikes for 83 of the 85 VOCs. This difference could indicate that degradation might be causing a bias in environmental-sample data; however the authors considered this to be unlikely because previous, more controlled studies had not indicated any VOC degradation. They state that the differences in recovery could have been caused "simply by difference in spiking technique, spiking experience, the number of different individuals involved in processing field-matrix spikes, and (or) environmental conditions when the samples were spiked" (Rowe and others, 2005, p. 24). Another possible explanation for the difference in recoveries is the discrepancy in sample size (149 field matrix spikes, but only 20 laboratory matrix spikes). A paired-sample analysis, such as the signed-rank test, on a subset of 20 field matrix spikes associated with the 20 laboratory matrix spikes might have provided more insight into matrix specific sample degradation.

Rowe and others (2005) also evaluated potential matrix effects by comparing median recoveries in laboratory-matrix and laboratory-reagent spikes. Recoveries were not significantly different for all but two VOCs, indicating that analytical results for environmental samples generally were not affected by matrix interference. The report did not include interpretation of potential degradation or matrix effects for individual VOCs, but did provide a table of median recoveries for each of the 85 compounds in each type of spiked sample.

Analysis and Interpretation of Data for Replicates

Replicates are used to measure variability, which is defined as the random error in independent measurements as the result of repeated application of the measurement process under identical conditions. Statistical evaluation of replicate variability is based on the standard deviation of measured values in the primary environmental sample and the replicate sample or samples. If only one set of a large number of replicates was collected, the standard deviation could be calculated directly; however, the general practice is to collect many sets of a small number of replicates under different conditions.

Evaluating Variability in Analyte Detection

Analysis of variability is complicated when analytes are reported as censored values (less than the reporting level) in one or more samples within a replicate set. In this case, the standard deviation of analyte concentrations in the set cannot be calculated; however, an alternative measure, variability in analyte detection, can be estimated. This variability is determined either by calculating the mean detection rate in all replicate sets or by calculating the percentage of replicate sets with inconsistent detections (sets that contain both quantified and censored values). The percentage of replicate sets with inconsistent detections is calculated as the number of replicate sets with inconsistent detections divided by the total number of replicate sets minus the number of sets with consistent nondetections (all analytical results less than the reporting level). Replicate sets with consistent nondetections are excluded from the calculation because the objective of the analysis is to evaluate the variability of detection rather than the variability of nondetection. Mean detection rate and percentage of inconsistent replicate sets are closely related estimates of variability in analyte detection. Mean detection rates that are high correspond to percentages of inconsistent replicate sets that are low, and the converse also is true. The percentage of inconsistent replicate sets is the preferred method because uncertainty can be estimated by calculating confidence limits.

Detections in a single replicate set are either consistent or inconsistent, regardless of the number of replicates in the set. Confidence limits for the proportion of measurements that have only two possible outcomes can be calculated using the method for percentage of nonconforming units (Hahn and Meeker, 1991, p. 104–105). In the context of replicate analysis, nonconforming units are sets with inconsistent detections; conforming units are sets that contain only quantified results (consistent detections). The one-sided upper confidence limit for the percentage of inconsistent replicates sets, derived from equation 8 is computed as:

$$P_U = 100\left\{1 + \frac{n - x}{(x+1)F_{1-\alpha,df_1,df_2}}\right\}^{-1} \qquad (21)$$

where

P_U is the upper confidence limit, in percent,

n is the total number replicate sets,

x is the number of replicate sets with inconsistent detections, and

F is the percentage point of the F distribution with $100(1-\alpha)$ percent confidence and degrees of freedom $df_1 = 2x + 2$ and $df_2 = 2n - 2x$.

For example, Martin (2002, p. 25) reported on a set of pesticide replicate data that included 37 sets with consistent detections of alachlor and 7 sets with inconsistent detections. The percentage of inconsistent replicate sets is simply 7 divided by the sum of 37 plus 7, or 15.9 percent. The upper 90-percent confidence limit on this percentage is:

$$P_U = 100\left\{1 + \frac{44 - 7}{(7+1)F_{0.9,16,74}}\right\}^{-1} = 25.3 \text{ percent} \qquad (22)$$

Martin (2002, p. 33) assumes that if the percentage of inconsistent detections is less than 25 percent then the variability of detection can be considered low. For the alachlor example, there is only a 10 percent likelihood that the percentage is greater than 25.3 percent, which was close enough to the criterion that detections could be considered reproducible without any qualification to interpretation of the data.

Evaluating Variability in Analyte Concentration

For many chemical constituents, variability (as standard deviation) is correlated with the mean concentration of that constituent in a replicate set. If a sufficient number of replicate sets have been collected over a range of concentrations, variability can be evaluated by estimating standard deviation as a function of concentration. Three approaches are presented for making these estimates:

- A piecewise-linear model, as used by Mueller and Titus (2005), hereinafter referred to as the "two-range model,"
- A pooled-variance model, as used by Martin (2002), and
- A bias-corrected log-log regression model.

These methods are described in the following sections and each is applied to two example datasets. The first dataset comprises concentrations of nitrite-plus-nitrate in replicates collected as part of an assessment of total nitrogen in surface water (Rus and others, 2012). (Note: Hereinafter this analyte is referred to simply as "nitrate," which is the primary component in oxygenated water.) The second dataset is atrazine concentrations in groundwater replicates collected for the NAWQA Program during 1993–2006. Results in both datasets were not subjected to typical data rounding; additional significant figures were retained in order to provide more precise values for replicate standard deviation and mean concentration. Both datasets are provided in Appendix 1.

Two-Range Model

Over a range of low concentrations, standard deviation of replicates generally is uniform, but at higher concentrations, standard deviation tends to increase in proportion to concentration (figs. 13*A* and 14*A*). Within this high range, the relative standard deviation (RSD), defined as the ratio (in percent) of standard deviation to mean concentration, is generally uniform (figs. 13*B* and 14*B*). Both datasets (nitrate and atrazine) can thus be divided into two pieces: a low concentration range for which standard deviation is approximately constant, and a high concentration range for which RSD is approximately constant. For concentrations within the low range, variability can be estimated as the average standard deviation of replicates; within the high range, variability can be estimated as the average RSD.

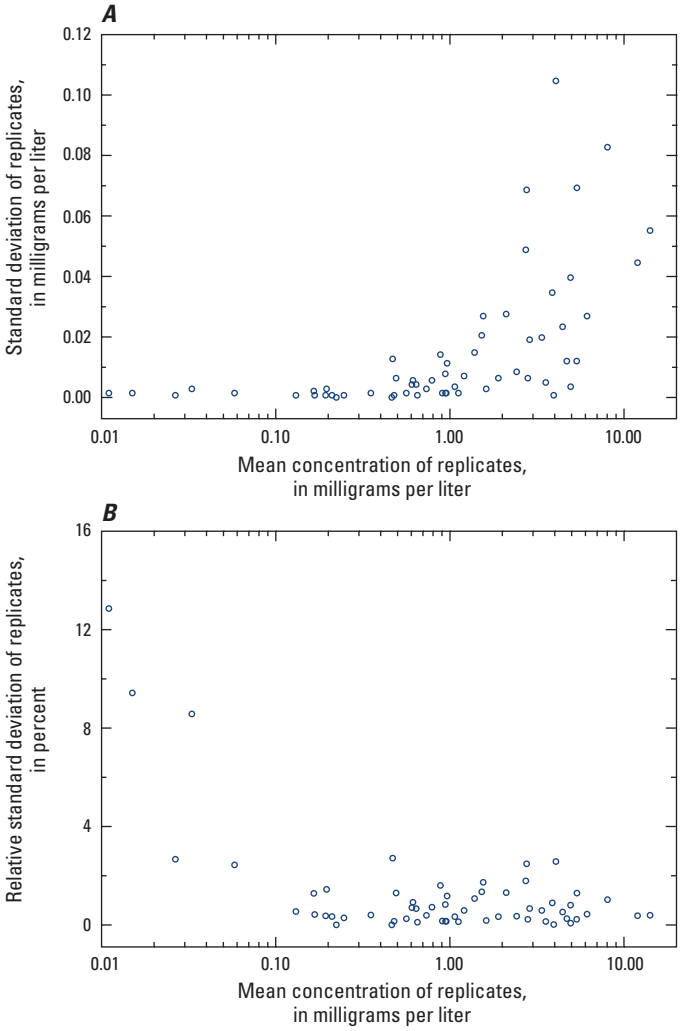


Figure 13. Plots of *A*, standard deviation and *B*, relative standard deviation, compared to mean concentration of nitrite-plus-nitrate in replicate surface-water samples collected for the total nitrogen study dataset (data from Rus and others, 2012, are compiled in table 1–1).

An appropriate boundary concentration between the low and high ranges can be selected by graphical analysis of standard deviation and RSD in relation to mean concentration. Adding a line, such as a locally weighted scatterplot smooth (LOWESS) through the center of the data (Chambers and others, 1983), can aid in the selection of the best boundary concentration. Figure 15 shows the standard-deviation and RSD data for nitrate with LOWESS curves indicating that the change in slope occurs at a mean concentration of about 0.4–0.5 mg/L. Selection of the boundary within these limits is somewhat subjective. In this example, the boundary was selected at 0.5 mg/L, as indicated by the dashed line on the plot. This selection incorporates two replicates with low standard deviations, at concentrations of 0.46 and 0.48 mg/L, into the low range and keeps one replicate with a moderately high RSD, at a concentration of 0.47 mg/L, out of the high range. Figure 16 shows the atrazine data with LOWESS curves. In this case, the curve through the RSD data (fig. 16*B)* is not helpful in selecting a boundary concentration, but the change in slope for standard deviation (fig. 16*A)* is clear at a mean concentration of about 0.03–0.07 µg/L. For this example, the boundary was selected to be 0.04 µg/L, about the mid-point of the change in slope.

After the boundary concentration has been selected and the replicate data have been divided into the two ranges, average standard deviation is calculated for replicates in the low range and average RSD is computed for replicates in the high range. For the example data,

- Average standard deviation for low-range nitrate replicates is 0.0021 mg/L,
- Average RSD for high-range nitrate replicates is 0.71 percent,
- Average standard deviation for low-range atrazine replicates is 0.0007 µg/L, and
- Average RSD for high-range atrazine replicates is 3.53 percent.

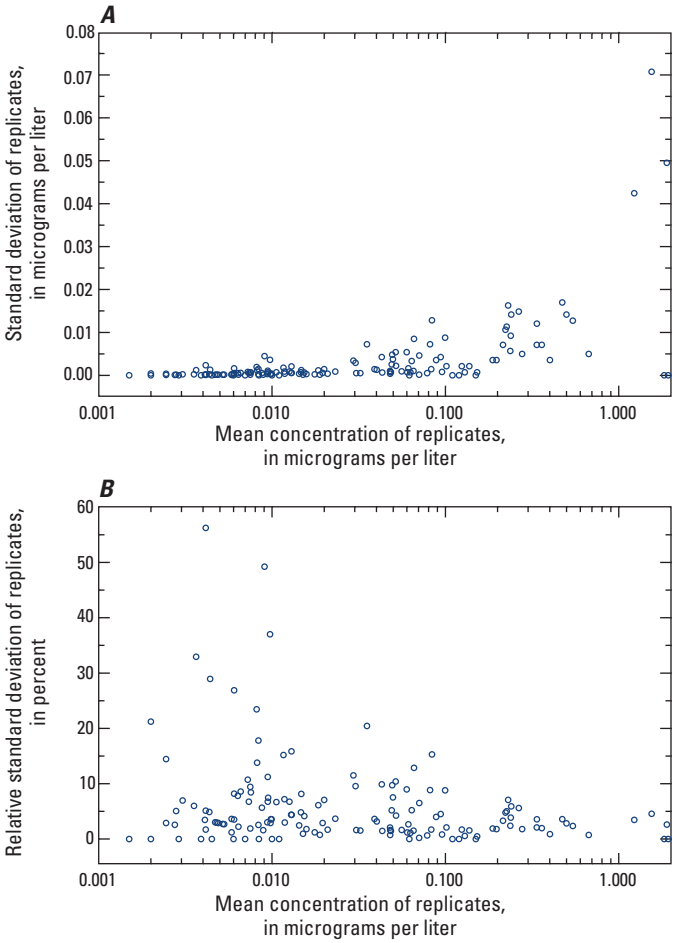


Figure 14. Plots of *A*, standard deviation and *B*, relative standard deviation compared to mean concentration of atrazine in replicate groundwater samples collected for the National Water-Quality Assessment Program, 1993–2006 (data from Martin, 2012, are compiled in table 1–2).

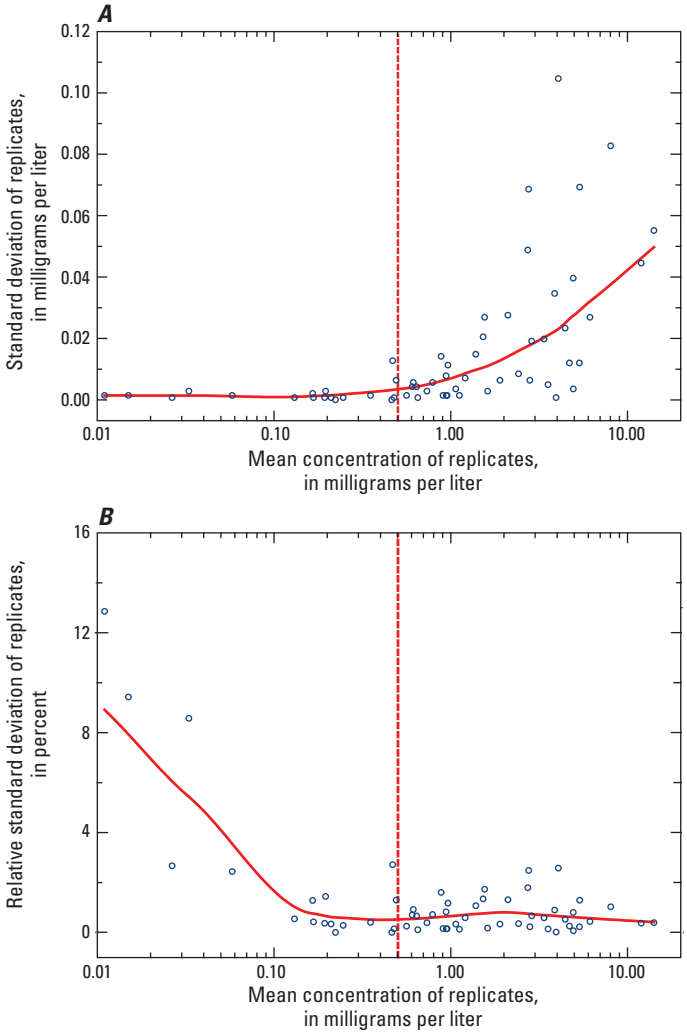


Figure 15. Plots of the nitrate example data (from fig. 13) with a LOWESS smooth (solid red line) and selected boundary between low-range and high-range concentrations (dashed red line).

These averages can be pieced together as two linear estimates on a single graph of standard deviation (fig. 17). Over the low range, the line of estimated standard deviation is horizontal; over the high range, estimated standard deviation is linear relative to mean replicate concentration, with a slope equal to the average percent RSD. (Note: The high-range lines appear curved because the scale for mean concentration is logarithmic in figure 17.) Some adjustment to the boundary concentration might be necessary if the average low-range standard deviation and high-range RSD do not intersect at the transition point. For the nitrate example, the estimated standard deviation at the transition point (0.5 mg/L) is 0.0021 mg/L using the low-range average and 0.0036 mg/L (calculated as 0.5 mg/L times 0.71 percent divided by 100) using the high range average RSD. The difference between these estimates is not decreased if the boundary is shifted to 0.4 or 0.6 mg/L. The difference could be decreased if the boundary was shifted to 0.3 mg/L, but then there would be only 13 replicate sets in the low range, so the average standard deviation could be poorly defined.

Pooled-Variance Model

Martin (2002) introduced the pooled-variance model as an alternative to the two-range model. The primary difference between these models is that the pooled-variance model splits the replicate data into more than two subsets and averages the variance, rather than standard deviation, to estimate variability within subsets. Variance is simply the square of standard deviation, but unlike standard deviations, variances are additive. When multiple statistical "samples" are combined ("pooled"), it is appropriate to use average variance but not average standard deviation.

For the pooled-variance model, the mean variance of each replicate subset is calculated as described by Anderson (1987, p. 44–45). This method requires that the individual variances within the subset be approximately equal; therefore, this requirement should be a criterion for subset selection. In practice, even approximate equality of variances within each subset is difficult to achieve, particularly at higher concentrations (for example, as in figs. 13 and 14). In addition, if replicates for many analytes are being evaluated, selecting subsets for each

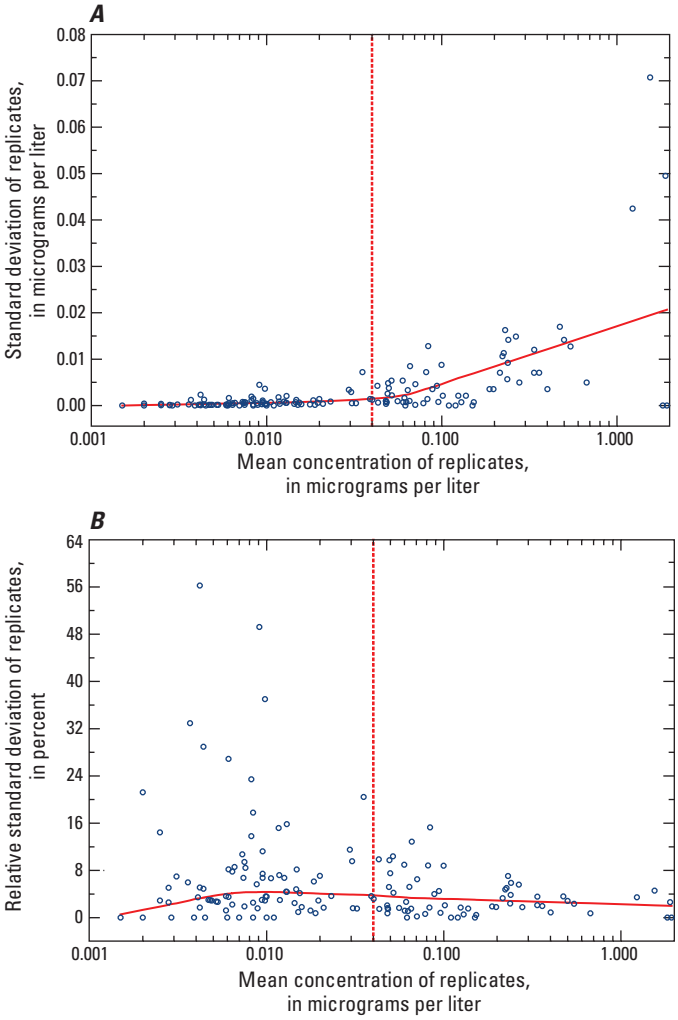


Figure 16. Plots of the atrazine example data (from fig. 14) with a LOWESS smooth (solid red line) and selected boundary between low-range and high-range concentrations (dashed red line).

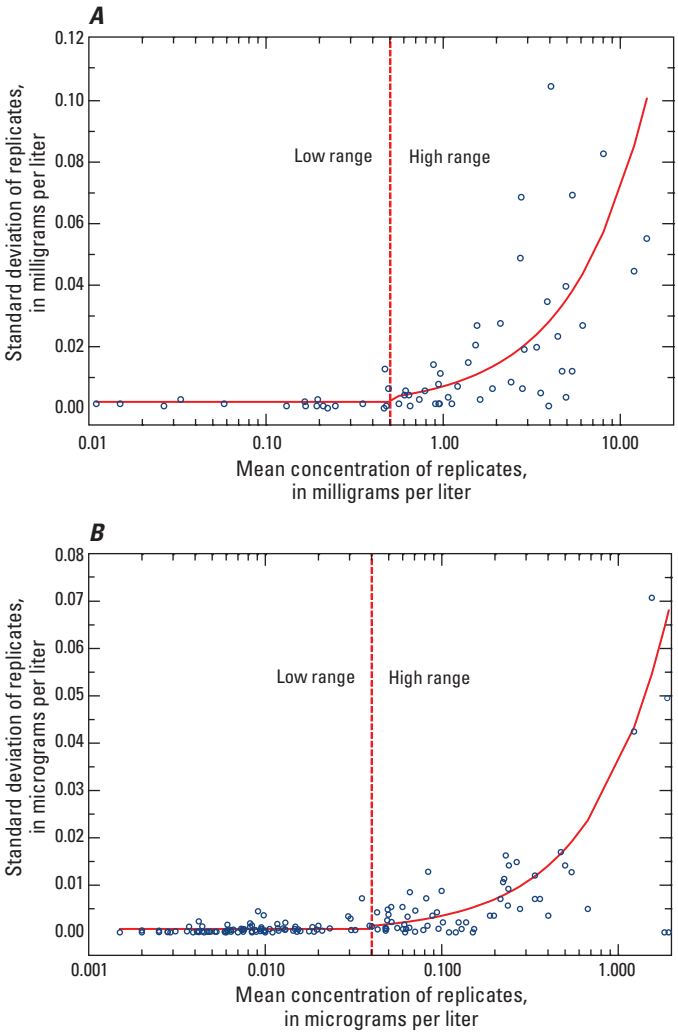


Figure 17. Plots of the *A*, nitrate and *B*, atrazine replicate data (from figs. 13 and 14) with solid red lines indicating the best estimates of standard deviation based on the two-range model.

one becomes tedious. Martin surmounted these difficulties by using a standard set of concentration ranges for each analyte. Replicates with an overall range within 0.001–10 µg/L were divided into eight overlapping subsets: less than 0.01 µg/L, 0.005–0.05 µg/L, 0.01–0.1 µg/L, 0.05–0.5 µg/L, 0.1–1 µg/L, 0.5–5 µg/L, 1–10 µg/L, and greater than 5 µg/L. The mean (pooled) variance was computed for each subset and was considered to be the best estimate of replicate variance within the central part of the subset range. An illustration of this subsetting technique is provided in figure 18. The six data subsets that have specific end-points are shown; the subsets less than 0.01 and greater than 5 µg/L are excluded. The applicable ranges of concentration for each subset also are shown. For example, the mean variance for data subset 2 (0.01–0.1 µg/L) was applied to concentrations from 0.025–0.075 µg/L; the mean variance for data subset 5 (0.5–5 µg/L) was applied to concentrations from 0.75–2.75 µg/L, and so forth. The overlapping subset ranges improve estimates of variance for concentrations that otherwise would be at the extremes of a range.

Mean pooled variance was computed for each range of replicate concentration in the nitrate and atrazine example datasets using overlapping subsets as described above. Pooled standard deviation within each range of concentration was estimated as the square-root of the mean variance. The pooled standard deviations for all ranges of concentration are plotted as step-wise lines in figure 19. The lines are generally higher

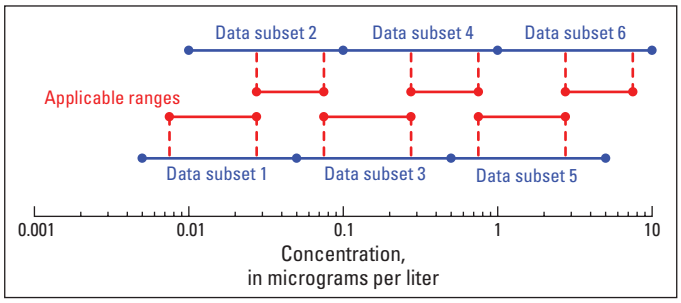


Figure 18. Illustration of the data subsetting technique used in the pooled-variance model of replicate variability.

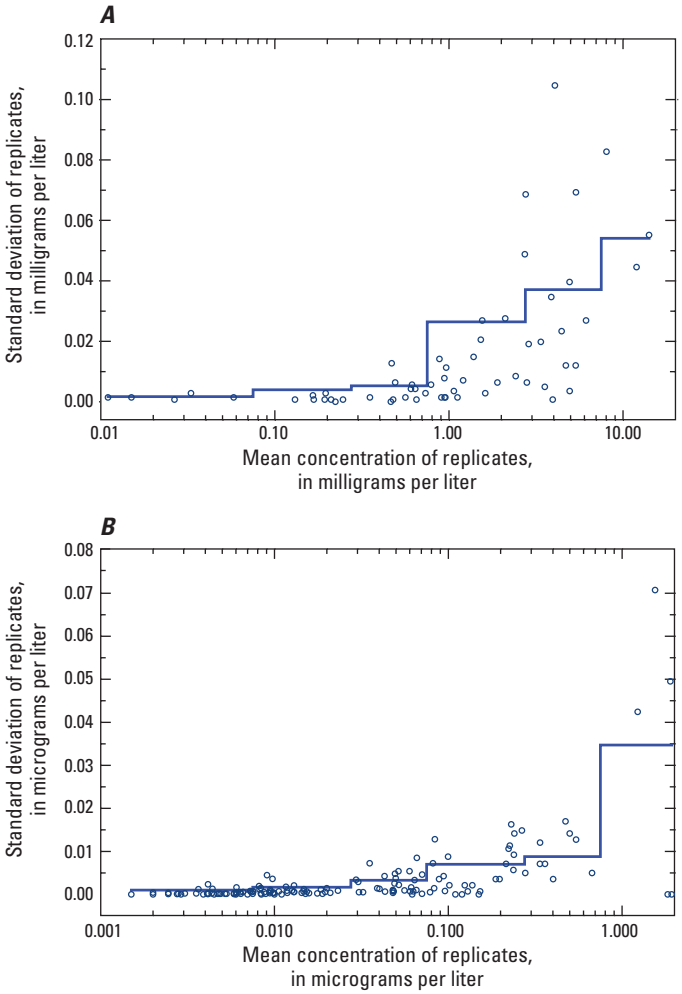


Figure 19. Plots of the *A*, nitrate and *B*, atrazine replicate data (from figs. 13 and 14) with horizontal blue lines indicating the best estimates of standard deviation at each step based on the pooled-variance model.

than the center of mass for standard deviation of individual replicate sets within each range. This happens because the pooled estimates are based on squared standard deviation, so higher values have more weight in calculation of the mean.

Bias-Corrected log-log Regression Model

Both the previous models are somewhat difficult to apply because of the need to divide the data into subsets. In addition, this division requires subjective judgment, and if the subset boundaries are changed, estimated standard deviations also will change. In order to avoid these problems, a third model has been developed. This model is based on the approximately linear relation between the logarithms of replicate standard deviation and mean concentration. Log-transformation is well justified for both these variables. Standard deviation follows a chi-square distribution, which is positively skewed with a lower bound of zero (Ott and Longnecker, 2001, p. 344). Constituent concentrations in natural water also have positive skewness and a lower bound of zero (Helsel and Hirsch, 2002, p. 2). Such distributions can be approximated by a lognormal distribution, so log-transformations of these data will be more normally distributed and more appropriate for regression analysis. Figure 20 shows least-square regression lines fit through the base-10 log-transformed nitrate and atrazine example data. The relation between replicate standard deviation and mean concentration is approximately linear for both constituents. Note that replicates with identical analytical values have a standard deviation of zero, and these are excluded from the model because the logarithm of zero is undefined. The regression line appears to be a good predictor of average replicate standard deviation for both constituents, though somewhat better for atrazine as indicated by a higher coefficient of determination (R^2) and lower standard error. In addition, the model residuals for both constituents are close to homoscedastic and are normally distributed, indicating that the log-transform was appropriate for linear regression.

Instead of the log-log model of standard deviation, a log-log model of variance might be preferable because such a model could be considered to produce "continuously pooled" estimates of variance over the entire range of concentration measurements. But that is not necessary because model results are identical for logarithms of either variance or standard deviation. The logarithm transform linearizes the relation between variance (*Var*) and standard deviation (*SD*):

$$Var = SD^2 \tag{23}$$

$$\log(Var) = 2\log(SD) \tag{24}$$

Thus, the coefficients of a log*(Var)* model will be exactly two times those of a log*(SD)* model. The estimated standard deviation from the log*(Var)* model will be one-half the estimated variance, which will be exactly equal to the log*(SD)* model estimate. Thus it is reasonable to "continuously pool" log(*SD*) values in the same way as log(*Var*). The regression model then is

$$\log(SD) = B_0 + B_1\log(C) \tag{25}$$

where

log*(SD)* is the logarithm of replicate standard deviation,

B_0 is the intercept of the regression line, estimated by least-squares,

B_1 is the slope of the regression line, estimated by least-squares, and

log*(C)* is the logarithm of mean replicate concentration.

The objective of any model of replicate variability is to estimate the mean standard deviation of replicate analyses in relation to mean replicate concentration. However, transformation of log-log model estimates back to original units provides an estimate of median standard deviation, which will be lower than the mean. Several methods have been developed to compensate for this low bias (Helsel and Hirsch, 2002, p. 255–257). The most general of these is the smearing estimate (Duan, 1983). Residuals from the log-log equation are transformed back to their original units. The mean of the transformed residuals is called the bias-correction factor. This factor (*bcf*) is multiplied by the estimated median standard deviations in order to correctly express these estimates as mean values:

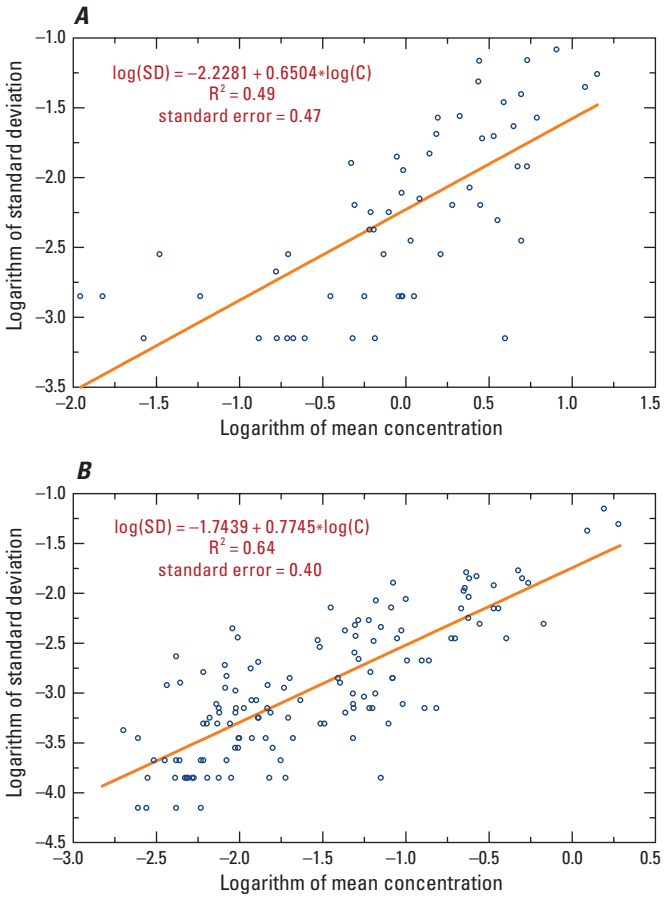


Figure 20. Plots of the base-10 log-transformed *A*, nitrate and *B*, atrazine example data (from figs. 13 and 14) with least-squares regression lines. (SD, standard deviation; C, concentration; R^2, coefficient of determination)

$$SD = bcf\ \{10^{[B_0+B_1\log(C)]}\} \quad (26)$$

Figure 21 shows the example nitrate and atrazine replicate data with curves indicating the estimated standard deviation from the bias-corrected log-log model. The bias correction factors are 1.619 for the nitrate replicates and 1.517 for the atrazine replicates, so with the regression equations shown in figure 20, the complete bias-corrected equations are:

$$\text{Nitrate: } SD = 1.619\ \{10^{[-2.2281+0.6504\log(C)]}\} \quad (27)$$

$$\text{Atrazine: } SD = 1.517\ \{10^{[-1.7439+0.7745\log(C)]}\} \quad (28)$$

Both curves provide a good approximation of average replicate standard deviation throughout the range of concentration.

Comparison of the Three Models of Variability

Each of the three models of variability has advantages and disadvantages. The two-range model is simple to apply, but requires subjective judgment about the boundary between the ranges of concentration. The statistical justification for the model is somewhat questionable because average values are computed using standard deviation rather than variance, as is the more common practice. In general, the model produces low estimates of standard deviation around the boundary concentration, because each line is affected by lower values at the extremes (low standard deviation at low concentrations and low RSD at high concentrations). In addition, estimates of standard deviation can be high at the highest concentrations if the average RSD is affected by many high values nearer the boundary. The pooled-variance model is the most complex to apply; it requires subjective judgment to determine multiple concentration boundaries. Statistical justification is good if variances are approximately equal within each range of concentration. If a systematic approach is used to set boundaries, application becomes simpler, but the risk of unequal variances increases. If variances are not equal, estimates of standard deviation can be too high or too low, particularly in the middle and upper ranges of concentration. The bias-corrected log-log model is the easiest to apply and requires no subjective judgment. It is well justified as long as the log-log relation is linear. The primary disadvantage of this model is that replicates with identical analytical values produce undefined logarithms, which cannot be used to fit the regression line, so estimates of standard deviation might be slightly high throughout the range of concentrations.

Estimated standard deviation from the three models are shown in figure 22 for the nitrate and atrazine example data. All three produce essentially equivalent estimates of standard deviation at low concentrations. The two-range model produces the lowest estimates within the mid-range of concentrations and the highest estimates at the extreme high concentrations. The pooled-variance model produces high estimates for the standard deviation of nitrate within the mid-range of concentrations and low estimates for the standard deviation

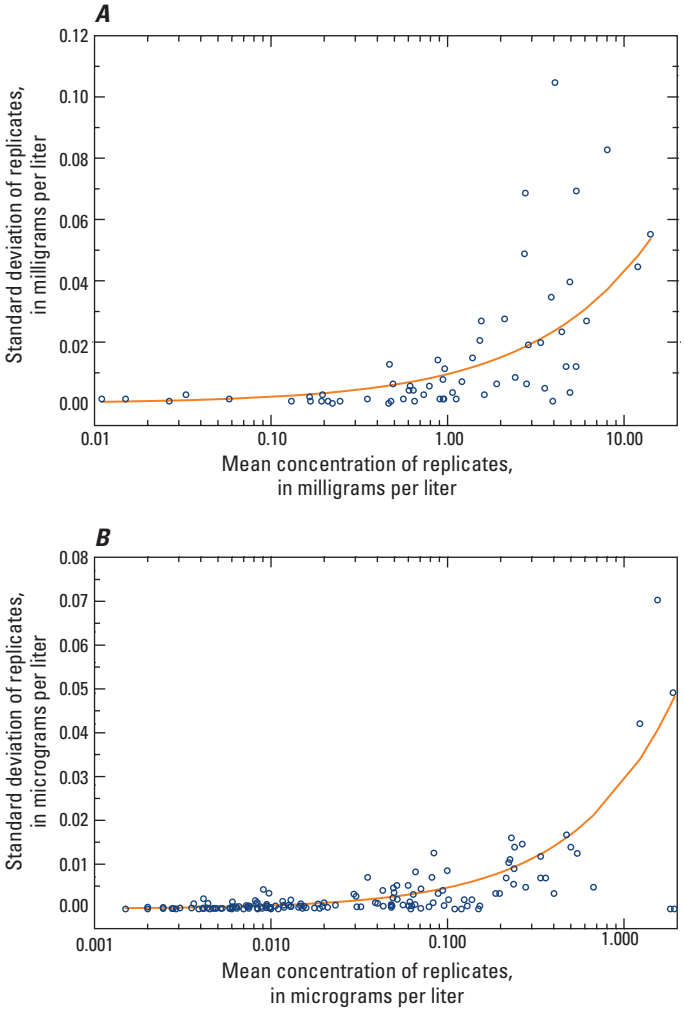


Figure 21. Plots of the *A*, nitrate and *B*, atrazine replicate data (from figs. 13 and 14) with orange lines indicating the best estimates of standard deviation based on the bias-corrected log-log model.

of atrazine at high concentrations. The bias-corrected log-log model produces reasonable estimates of standard deviation for both analytes throughout the concentration range.

Determining How Many Replicates to Collect

In the design phase of a project, the project staff must determine the number of field replicates that will be required to adequately estimate the potential variability in environmental samples. A statistical approach is appropriate even if only a few replicates are collected. The number of replicates required to meet project objectives should be determined based on the following two criteria:

- What level of confidence is necessary?
- How much uncertainty is acceptable?

Uncertainty is the potential underestimation of true variability that could result from using standard deviation based on replicate data. (Note that overestimation of true variability is

of less concern in a conservative evaluation of data quality.) Uncertainty is determined by constructing an upper confidence limit on the estimated standard deviation (Hahn and Meeker, 1991, p. 55):

$$S_U = SD\sqrt{\left(\frac{df}{\chi^2_{\alpha,df}}\right)} \quad (29)$$

where

S_U is the upper confidence limit on the true standard deviation,

SD is the standard deviation based on replicate samples,

df is degrees of freedom (defined below), and

$\chi^2_{\alpha,df}$ is the percentage point of the chi-square distribution with α uncertainty and df degrees of freedom.

In this equation, $1–\alpha$ is the confidence that the population standard deviation (variability) has not been underestimated. The expression under the radical can be redefined as $(1+\delta)$, where 100δ is the potential uncertainty, in percent. The population standard deviation could be as much as $1+\delta$ times larger than the standard deviation estimated from the replicate data:

$$S_U = SD(1 + \delta) \quad (30)$$

Substituting terms from equation 29 and then solving equation 30 for df yields:

$$\sqrt{\left(\frac{df}{\chi^2_{\alpha,df}}\right)} = 1 + \delta \quad (31)$$

$$df = (\chi^2_{\alpha,df})(1 + \delta)^2 \quad (32)$$

Typically, degrees of freedom is 1 less than the number of observations used to estimate SD; however, if SD is estimated by pooling the variance for pairs of replicates, df is equal to the sums of the degrees of freedom for each pair (Anderson, 1987, p. 45). The SD estimated from two observations has 1 degree of freedom; therefore, df becomes simply the number of replicate pairs (n), and:

$$n = (\chi^2_{\alpha,n})(1 + \delta)^2 \quad (33)$$

Equation 33 allows calculation of the required number of replicates as a function of confidence ($1-\alpha$) and uncertainty (δ). Figure 23 shows the number of replicate pairs required to achieve various levels of uncertainty for three selected levels of confidence. For example, 10 replicates are required to estimate SD within 45 percent of the true standard deviation with 90-percent confidence. The uncertainty is that the true standard deviation could be as much as 45 percent greater than the estimated value. Increasing the number of replicates to 30 will decrease the uncertainty to within 21 percent greater than the estimated standard deviation. Decreasing the uncertainty to 15 percent would require 50 replicates.

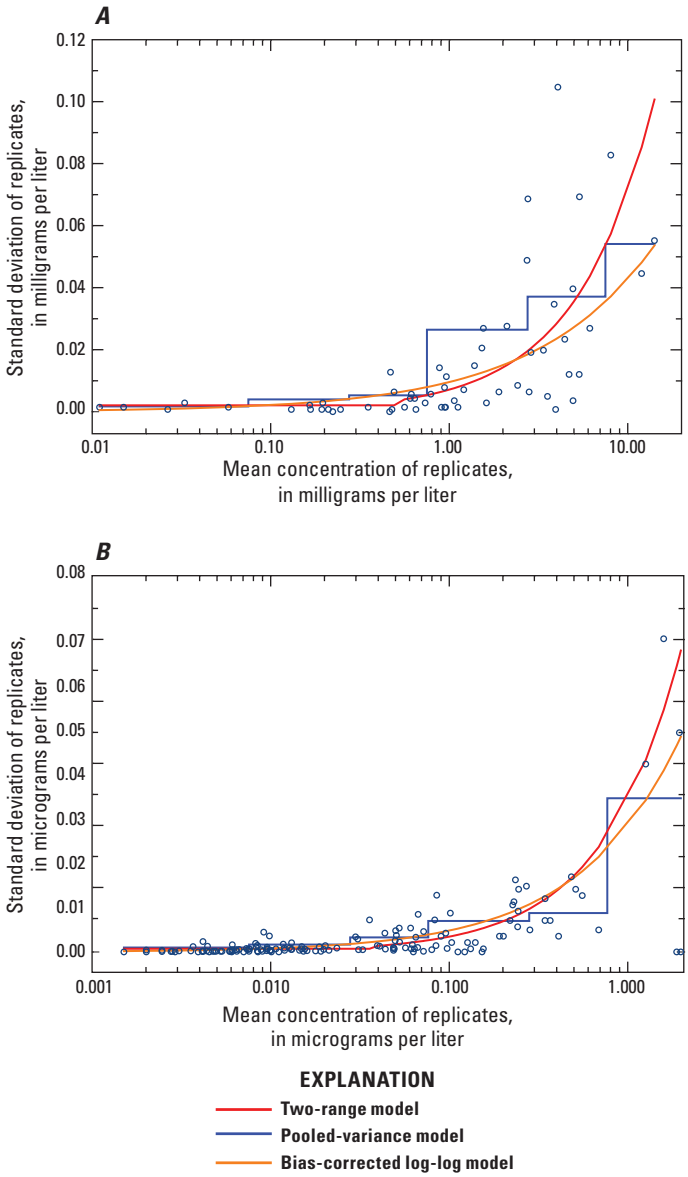


Figure 22. Plots of the *A*, nitrate and *B*, atrazine replicate data (from figs. 13 and 14) showing comparison of estimated standard deviations from the three models.

A number of problems can interfere with making a good estimate of variability from field replicate data. The distribution of constituent concentrations among sets of field replicates is not likely to be uniform, because the frequency of occurrence is typically inverse to concentration. Thus, low concentrations generally are predominant in field-replicate data, and few or no data might be available at high concentrations. In this instance, variability within the high range of concentrations might be impossible to define, particularly for the two-range model. Another issue results from laboratory rounding of the analyzed concentrations: the possible

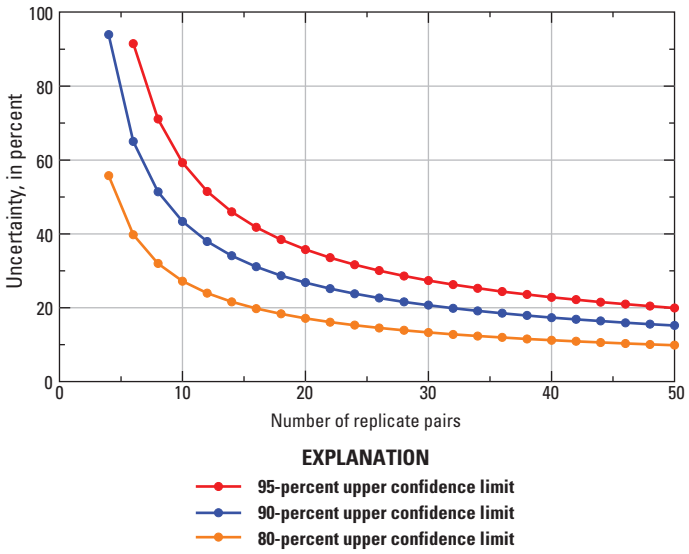


Figure 23. Number of replicate pairs required to determine selected upper confidence limits for uncertainty (potential under-estimation) in estimates of variability using the standard deviations of field replicates.

differences among rounded concentration values are not continuous but occur at discrete intervals that change with the order of magnitude of concentration. Thus, standard deviations can be defined with better resolution for low-concentration replicates than for high-concentration replicates. Again, determination of variability for the high range of concentrations might be adversely affected. Good QC designs attempt to ensure that as many replicates as possible are collected at locations and during times when concentrations are expected to be high. Also, calculation of replicate standard deviations should always be done with unrounded data, if they are available from the laboratory.

Using Replicate Variability to Evaluate Environmental-Sample Data

Variability determined from field replicates can be used to make various evaluations of uncertainty in environmental-sample data. In all these evaluations, standard deviation estimated by a model of variability is assumed to represent the true standard deviation (σ) for all samples collected within the same inference space as the replicates. If replicate variability was determined using the two-range model and the measured concentration (C) is in the low range, σ is estimated as the average standard deviation of replicates within that range. If the measured concentration is in the high range, σ is C (RSD/100). If the pooled-variance model was used, σ is the square-root of the average variance for the applicable range of concentrations containing the measured value. If the bias-corrected log-log model was used, σ is calculated using the regression equation and bias-correction factor.

One particular use of replicate variability is to estimate the uncertainty of the concentration measured in a single environmental sample. After σ has been estimated, uncertainty in the measured concentration can be determined by constructing the confidence interval for the true concentration:

$$[C_L, C_U] = C \pm Z_{(1-\alpha/2)}\sigma \quad (34)$$

where

C_L, C_U is the lower and upper limits of concentration for the 100(1–α/2) percent confidence interval,

Z is the percentage point of the standard normal curve that contains an area of 100(1–α/2) percent,

α is the probability that the confidence interval does not include the true concentration, and

σ is standard deviation of the measured concentration, independently estimated from replicate variability.

The second term, $Z_{(1-\alpha/2)}\sigma$, in equation 34 represents the error inherent in a single measurement of concentration due to field variability.

For example, consider surface-water samples collected within the inference space (same general area and during the same time) as the replicates used in the models of variability in the preceding section of this report. If nitrate in one of these samples is reported as 9.5 mg/L, standard deviation of this measurement can be estimated using the bias-corrected log-log model (eq. 27):

$$\sigma = 1.619\ \{10^{[-2.2281+0.6504\log(9.5)]}\} = 0.0414 \text{ mg/L} \quad (35)$$

The Z-value for a 90-percent confidence interval (α/2 = 0.05) is 1.645. Thus the 90-percent confidence interval (from eq. 34) is:

$$[C_L, C_U] = 9.5 \pm 1.645(0.0414) = [9.43, 9.57] \quad (36)$$

The actual concentration of nitrate in this sample is estimated, with 90-percent confidence, to be in the range of 9.43 to 9.57 mg/L.

A measured concentration also could be compared to a water-quality standard in order to estimate the probability that the true concentration in the sample exceeded the standard. In this case, the standard is set equal to the one-sided confidence limit in one of the following equations:

$$C_L = C - Z_{(1-\alpha)}\sigma \quad (37)$$

$$C_U = C + Z_{(1-\alpha)}\sigma \quad (38)$$

Equation 37 is used if C is greater than the standard; equation 38 if C is less than the standard. (If C is equal to the standard, the probability of exceedance is 50 percent.) The equation is solved

for Z, and the associated α value is determined from a table of Z (standard normal) scores. The probability that the standard has been exceeded is $100(1-\alpha)$ percent for measured concentrations greater than the standard and 100α percent for measured concentrations less than the standard. For the example, the measured concentration of nitrate was 9.5 mg/L and the drinking-water standard is 10 mg/L (http://water.epa.gov/drink/contaminants/index.cfm). Because the measured value is less than the standard, equation 38 is used:

$$10 = 9.5 + Z_{(1-\alpha)}(0.0414) \quad (39)$$

Solving for Z yields a standard-normal score of 12.1, for which α is less than 0.0001 (from tables in, for example, Ott and Longnecker, 2001, p. 1092). The probability of exceedance is equal to 100α percent. For this example, the measured concentration of 9.5 mg/L indicates there is less than a 0.01 percent likelihood that the true concentration in the sample exceeded the 10 mg/L standard.

Another use of replicate variability is to estimate the minimum difference in mean concentrations that can be measured with a specified level of confidence. This estimate is based on the standard deviation (SD_{diff}) of the difference between mean concentrations in two sets of data, X_1 and X_2:

$$SD_{diff} = \sqrt{\frac{SD^2_{X_1}}{n_1} + \frac{SD^2_{X_2}}{n_2}} \quad (40)$$

where

$SD^2_{X_1}, SD^2_{X_2}$ is variance of datasets X_1 and X_2, and

n_1, n_2 is the number of samples in datasets X_1 and X_2.

Assuming no environmental variability (that is, no actual difference in true concentrations) the only source of SD_{diff} is field variability (SD_{FV}), which is introduced by sampling and laboratory procedures and estimated from replicate variability. If the two sets of environmental data are of equal size, equation 40 becomes:

$$SD_{FV}\sqrt{\frac{2(SD_{reps})^2}{n}} \quad (41)$$

where

SD_{reps} is standard deviation estimated using one of the replicate models (two-range, pooled-variance, or log-log regression).

The confidence interval for the difference between mean concentrations (ΔC), for comparison of two sites or two time periods, is

$$\Delta C_{interval} = \Delta C \pm Z_{(1-\alpha/2)}SD_{diff} \quad (42)$$

If $\Delta C_{interval}$ includes zero, then the difference is not significant. If the only source of SD_{diff} is field variability, the difference is likely to be significant only if

$$|\Delta C| \geq Z_{(1-\alpha/2)}SD_{FV} \quad (43)$$

For example, assume groundwater samples were collected from four wells at a location and time within the inference space of the replicates from the previous example. The mean concentration of atrazine in these samples was 0.5 mg/L. The preceding equations can be used to determine the smallest increase or decrease that is likely to be statistically significant if sampling is repeated in the future. The standard deviation of atrazine at 0.5 mg/L can be estimated by using the bias-corrected log-log model (eq. 28):

$$SD_{reps} = 1.517\ \{10^{[-1.7439+0.7745\log(0.5)]}\} = 0.0160\ \mu g/L \quad (44)$$

The smallest likely variability in the mean concentration (assuming no actual differences among wells) is then estimated by substituting the result of equation 44 into equation 41:

$$SD_{FV} = \sqrt{\frac{2(0.0160)^2}{4}} = 0.011\ \mu g/L \quad (45)$$

This is the unavoidable variability due to sample collection and analysis. The smallest difference in mean concentration that would be significant with 90-percent confidence is then (from eq. 43):

$$|\Delta C| = 1.645(0.011) = 0.019\ \mu g/L \quad (46)$$

Therefore, the mean concentration in samples from the four wells next year must be either less than 0.481 µg/L or greater than 0.519 to be considered a statistically significant decrease or increase.

Examples of Analyzing a Few Replicates Collected for a Single Project

For some small projects, only a few samples might be collected, so field variability cannot be modeled even if replicates are collected with every sample. In this case replicate results can be compared to some assumed criteria in order to indicate whether environmental data generally seem acceptable or might be affected by elevated variability. The following examples illustrate several possible approaches to this type of analysis.

Replicate Example 1: Replicates Associated with One Set of Environmental Samples

The Pavillion project was described previously in "Blank Example 1" in "Evaluating Contamination Based on Single Blanks." Samples collected from the monitoring well included two environmental samples (collected after different amounts of water had been purged from the well) with replicates for each sample (Wright and others, 2012). Concentrations were reported for 244 constituents in 570 replicate-analyte pairs.

The relative percent difference (RPD) between replicate-pair results was calculated using the following equation:

$$\mathrm{RPD} = 100\left\{\frac{\text{larger result} - \text{smaller result}}{\left(\text{larger result} + \text{smaller result}\right)/2}\right\} \quad (47)$$

A criterion of 20-percent maximum difference between replicate-pair concentrations was considered to be acceptable. Variability was within this criterion for 559 (98 percent) of the replicate-analyte pairs. One pair of results each for 11 constituents exceeded the criterion, and so were qualified by adding a code to the data tables in the report. This code indicated that high variability might affect interpretation of environmental data, though no interpretation was included in the report.

Replicate Example 2: Replicates Associated with More than One Set of Environmental Samples

The Colorado Water Science Center's Eagle River study was described previously in "Blank Example 3" in "Evaluating Contamination Based on Multiple Blanks." Samples were collected one time from 61 valley-fill aquifer wells and quarterly for 1 yr from 10 surface-water sites. The QC samples included replicate pairs for six groundwater and seven surface-water samples (Rupert and Plummer, 2009, table 7). An RPD was calculated (eq. 47) for as many as 23 analytes in each replicate pair.

The RPD of replicate concentrations was very small for most analytes; mean RPD was less than 11 percent for all but ammonia (19.6 percent) and iron (25.4 percent) in surface-water samples and for all but iron (28.7 percent) and manganese (34.1 percent) in groundwater samples. The ammonia result was based on only two replicate pairs, both of which had low concentrations with little absolute difference between replicates. Project investigators concluded that "Overall, the replicate samples indicated that there was low variability (high precision) in the major-ion and nutrient analyses" (Rupert and Plummer, 2009, p. 27).

Replicate Example 3: Replicates Associated with More than One Set of Environmental Samples

The Gulf of Mexico oil spill project was described previously in "Blank Example 2" in "Evaluating Contamination Based on Single Blanks." Various numbers of replicate water samples and replicate sediment samples were collected during each time period (pre-landfall and post-landfall). Results for many analytes in water samples were censored (less than the reporting level) in one or both replicate samples; these were excluded from replicate analysis. Water samples from the pre-landfall and post-landfall periods were, for the most part, analyzed at different laboratories, so replicates from each period were compiled into separate datasets. For pre-landfall water samples, replicate pairs with quantified results were available for 21 analytes, and the number of pairs ranged from 4 to 27, depending on the analyte. For the post-landfall period, quantified results were available for 12 analytes, and the number of replicates pairs ranged from 4 to 7. Sediment samples collected during both sampling periods were analyzed at the same laboratories: inorganic constituents at the USGS Sediment Chemistry Laboratory at the Georgia Water Science Center (Norcross, Georgia) and organic compounds at a non-USGS laboratory. Quantified results were available for 31 inorganic constituents in 4 to 17 replicate pairs and for 15 organic compounds in 5 to 17 replicate pairs (Nowell and others, 2013).

Project investigators intended to evaluate variability using the two-range model, but the number of replicate results was too small to separate into low and high ranges of concentration, so variability was simply estimated as the mean RSD. This was considered a conservatively high estimate of variability, because high RSD values for low-concentration replicates were included in the calculation. Mean RSD was less than 10 percent for 13 of 21 analytes in pre-landfall replicate water samples and for 8 of the 12 analytes in post-landfall samples. For sediment replicates, mean RSD exceeded 20 percent for 27 of the 31 inorganic constituents and for 12 of the 15 organic compounds. In subsequent interpretation of environmental data, Nowell and others (2013) noted that uncertainty could be a problem for any analyte with variability (RSD) greater than 10 percent for water or 20 percent for sediment. This uncertainty limited interpretation of differences in analyte concentration between the two sampling periods.

The variability estimated from replicate data presumably was due to sampling or analytical errors. Adding more replicates could have provided more confidence in the estimate of variability; however, the sources of variability would have remained the same. The only way to decrease variability would have by using different sampling procedures or analytical methods.

Examples of Analyzing Many Replicates Collected for a Large Program

Datasets compiled for multiple projects or collected for large programs can include more than enough replicate samples so that analyte variability can be evaluated by using one of the statistical models. The NAWQA program provides a good example of such large datasets. The two following examples illustrate application of the two-range model to estimate variability of nutrient analytes and the pooled-variance model to estimate variability of pesticides.

Replicate Example 4: Many Replicates Collected for a Large Program

During Cycle I of the NAWQA program (1992–2001), more than 1,300 surface-water replicates and more than 500 groundwater replicates were collected within 52 study units around the nation (Mueller and Titus, 2005). Replicates from diverse locations could be compiled into a single dataset because they were collected and analyzed using similar methods and equipment (thus they were considered to represent a consistent inference space). Nutrient results from analysis of these replicates were evaluated in order to provide guidance on the potential effects that variability might have on the interpretation of environmental data from any study unit (Mueller and Titus, 2005). Separate evaluations were made for surface-water replicates and groundwater replicates because

these were collected using different methods. For two analytes (total Kjeldahl nitrogen and total phosphorus), the replicate data were split into three time periods, corresponding with changes in laboratory analytical methods. Variability was estimated using the two-range model; estimates for a few selected analytes are listed in table 14.

Using the estimates of variability in table 14, confidence intervals were constructed around analyte concentrations that were considered critical values for individual nutrients (table 15). Critical values included previously identified background concentrations and various water-quality standards or criteria.

The confidence intervals listed in table 15 allowed the authors to make the following statements about the effects of variability on interpretation of data from environmental samples (Mueller and Titus, 2005, p. 24–25).

> 95 percent of all measured concentrations within the range of critical values identified for ammonia in streams are expected to differ from the actual concentrations by no more than 0.25 mg/L or 9 percent of the measurement, whichever is smaller. In most circumstances, variability in this range has little effect on interpretation of ammonia data.
>
> At the highest aquatic-life criterion for ammonia (6.7 mg/L), the 95-percent confidence interval is 0.25 mg/L. Therefore, measured concentrations as high as 6.95 mg/L do not indicate exceedance of the criterion, with 95-percent confidence. Similarly, measured concentrations as low as 6.45 mg/L do not necessarily indicate compliance.
>
> For nitrite-plus-nitrate measurements at the drinking-water standard (10 mg/L), the 95-percent confidence interval is approximately 0.4 mg/L for stream samples. If laboratory results are rounded to two significant figures, a reported concentration of at least 11 mg/L would indicate an exceedance of the standard with 95-percent confidence. In this instance, the uncertainty caused by sampling variability has no real effect, because it does not change the least significant figure of the rounded value.
>
> For orthophosphate [and for] total phosphorus sampled after 1998, … differences of 0.02 mg/L would be considered significant for most individual measurements, and differences greater than 0.006 mg/L between means of 10 measurements would likely be unaffected by sampling variability. For the highest critical value for ammonia (6.7 mg/L), differences in individual measurements must exceed 0.5 mg/L to be considered significant.

Table 14. Examples of variability estimated from average standard deviation within a low range and average relative standard deviation within a high range of constituent concentrations (from Mueller and Titus, 2005, table 4).

[<, less than; >, greater than; mg/L, milligrams per liter; P, phosphorus]

Constituent	Concentration range (mg/L)	Variability	
		Value	Units
Ammonia in surface water	<0.2	0.0045	mg/L
	>0.2	1.9	percent
Nitrate in surface water	<1	0.012	mg/L
	>1	2.2	percent
Orthophosphate in surface water	<0.1	0.0027	mg/L
	>0.1	2.8	percent
Total P in surface water 1999–2001	<0.2	0.0032	mg/L
	>0.2	4	percent
Nitrate in groundwater	<1	0.043	mg/L
	>1	2.9	percent

Replicate Example 5: Many Replicates Collected for a Large Program

Another report on the quality of NAWQA data summarized pesticide results from replicate samples collected in streams and groundwater wells within the first 20 study units during 1992–1997 (Martin, 2002). Analytical data for 86 pesticides in 402 sets of surface-water field replicates and 187 sets of groundwater field replicates were used to evaluate variability. The variability of pesticide detections was assessed by calculating the mean percentage detection and the percentage of inconsistent replicates sets, as described in the section of this report on "Evaluating Variability in Analyte Detection." The variability of pesticide concentrations was assessed by using the pooled-variance model to estimate standard deviation and relative standard deviation for eight overlapping ranges of concentration.

Estimates of variability were presented in a series of tables. Selected examples for variability of detection are listed in table 16 and for variability of concentration in table 17.

Martin provided computational tools, similar to those in this report, so data users could evaluate the effects of variability on interpretations of subsets of the NAWQA data. He also drew the following general conclusions about the full dataset:

> The variability of detection for most pesticides is high at concentrations less than the minimum reporting level, but the variability of detection decreases dramatically at higher concentrations. … The overall rate of inconsistent replicate sets is 60.0 percent in the low range of concentration, 13.7 percent in the medium range, and 1.1 percent in the high range. … Inconsistent detections in replicate sets likely were caused by variability in the analytical method and by water-matrix interferences (or other loss processes) that cause false-negative errors. Consequently, estimates of the frequency of detection of pesticides in environmental water samples collected for the NAWQA

Table 15. Estimated sampling variability and confidence intervals around measured concentrations of nutrient analytes at selected critical values used to interpret environmental data (from Mueller and Titus, 2005, tables 5 and 6).

[mg/L, milligrams per liter; N, nitrogen; P, phosphorus; RSD, relative standard deviation]

Constituent	Critical value		Estimated sampling variability[1] (mg/L)	95-percent confidence interval (mg/L)	
	Concentration (mg/L)	Description		Individual measurements	Mean of 10 measurements
Ammonia in surface water	0.1	Background[2]	0.0045	0.091–0.109	0.097–0.103
	0.18	Aquatic-life criterion[3]	0.0045	0.171–0.189	0.177–0.183
	6.7	Aquatic-life criterion[3]	0.13	6.45–6.95	6.62–6.78
Nitrate in surface water	0.6	Background[2]	0.012	0.576–0.624	0.593–0.607
	10	Drinking-water standard[4]	0.22	9.56–10.4	9.86–10.1
Orthophosphate in surface water	0.05	Recommended to avoid eutrophication[5]	0.0027	0.045–0.055	0.048–0.052
Total P in surface water 1999–2001	0.1	Recommended to avoid eutrophication[5]	0.0032	0.094–0.106	0.098–0.102
Nitrate in groundwater	1.1	Background[2]	0.03	1.04–1.16	1.08–1.12
	10	Drinking-water standard[4]	0.29	9.43–10.6	9.82–10.2

[1]From table 14. For concentrations in the high range, sampling variability is concentration times RSD divided by 100.

[2]Mueller and others, 1995.

[3]Criterion varies depending on water temperature and pH (U.S. Environmental Protection Agency, 1999).

[4]http://water.epa.gov/drink/contaminants/index.cfm

[5]U.S. Environmental Protection Agency, 1986.

Table 16. Variability of pesticide detections in field replicates (from Martin, 2002, tables 5 and 6).

[MRL, minimum reporting level; ≥, greater than or equal to; <, less than; >, greater than]

Pesticide	MRL	Number of replicate sets with			Mean detection rate (percent)	Replicate sets with inconsistent detections (percent)	
		At least one detection	Consistent detections	Inconsistent detections		Measured	90-percent upper confidence limit
Mean concentration of the replicate sets: ≥ MRL and < 10 times the MRL							
Atrazine	0.001	60	50	10	90.2	16.7	24.5
Desethylatrazine	0.002	80	73	7	95.9	8.8	14.3
Simazine	0.005	99	98	1	99.5	1.0	3.9
Mean concentration of the replicate sets: > 10 times the MRL							
Atrazine	0.001	156	156	0	100.0	0.0	1.5
Desethylatrazine	0.002	82	82	0	100.0	0.0	2.8
Simazine	0.005	64	64	0	100.0	0.0	3.5

Program probably are biased low because of false-negative errors at concentrations near the minimum reporting level.

Results of correlation analyses indicate that for most pesticides and concentrations, pooled estimates of RSD rather than pooled estimates of SD should be used to estimate variability because pooled estimates of RSD are less affected by heteroscedasticity. The median pooled RSD was calculated for all pesticides to summarize the typical variability for pesticide data collected for the NAWQA Program. The median pooled RSD was 15 percent at concentrations less than 0.01 µg/L, 13 percent at concentrations near 0.01 µg/L, 12 percent at concentrations near 0.1 µg/L, 7.9 percent at concentrations near 1 µg/L, and 2.7 percent at concentrations greater than 5 µg/L.

Publication of Quality-Control Information

The primary goals for the publication of QC information are to (1) provide evidence for and insights about the sources and magnitude of bias and variability in measurements of environmental water-quality samples, and (2) explain how knowledge of bias and variability influenced the analysis and interpretation of the environmental data. This section of the report provides suggestions of the types of QC information that might be useful to include in the publications resulting from a variety of water-quality assessment projects. The guiding principal is that the types of QC information most needed are those that support specific study objectives and the conclusions of the study.

Table 17. Variability of pesticide concentrations in field replicates (from Martin, 2002, table 7).

[MRL, minimum reporting level; µg/L, micrograms per liter; ≥, greater than or equal to; <, less than]

Concentration range (µg/L)	Number of replicate sets	Median pooled standard deviation (µg/L)	Median pooled relative standard deviation (percent)
Atrazine: MRL 0.001 µg/L			
<0.01	49	0.0012	16.3
0.005 to <0.05	90	0.0014	11.8
0.01 to <0.1	80	0.0039	7.6
0.05 to <0.5	78	0.0128	7.5
0.1 to <1	62	0.0258	6.9
0.5 to <5	18	0.1396	7.1
1 to <10	12	0.1732	5.8
≥5	6	1.377	2.5
Desethylatrazine: MRL 0.002 µg/L			
<0.01	50	0.00095	18.2
0.005 to <0.05	79	0.0046	20.4
0.01 to <0.1	82	0.0061	18.5
0.05 to <0.5	42	0.0151	12.0
0.1 to <1	25	0.0258	10.8
0.5 to <5	3	0.0784	8.0
1 to <10	1	0.0919	7.6
Simazine: MRL 0.005 µg/L			
< 0.01	28	0.0010	14.8
0.005 to <0.05	98	0.0020	11.1
0.01 to <0.1	97	0.0027	8.4
0.05 to <0.5	52	0.0137	7.9
0.1 to <1	36	0.0197	8.8
0.5 to <5	12	0.1472	7.0
1 to <10	7	0.1989	9.1

Describe Institutional Quality-Assurance Programs

For some water-sampling projects, it may be appropriate to describe or cite the various quality-assurance programs that are applicable to the project. Many of these programs document various office, field, and laboratory activities that are used to ensure the quality of water-quality data. Examples of these programs include the project-specific or program-specific QA plans (for example, Mueller and others, 1997, for the NAWQA Program), USGS Water Science Center QA plans (template available at http://water.usgs.gov/owq/QAfolder/), the USGS National Field Manual (U.S. Geological Survey, variously dated), analytical method documents (for example, Sandstrom and others, 2001), Federal and State laboratory certification programs (for example, EPA Drinking Water Laboratory Certification Program: http://water.epa.gov/scitech/drinkingwater/labcert/), inter-laboratory quality programs (for example, the BQS Standard Reference Sample Project: http://bqs.usgs.gov/srs/), and external laboratory QA programs (for example, the BQS Inorganic and Organic Blind Sample Projects: http://bqs.usgs.gov/).

Define Quality-Control Terms

QC terms have not been standardized across agencies or programs. The same or similar terms might refer to different sources of bias or variability and, conversely, different QC terms might refer to the same sources of bias or variability. For example, the term “split replicates” is used by the USGS to identify a set of two or more essentially identical subsamples that have been produced by splitting (subdividing) a single sample. All subsamples are analyzed by the same method and laboratory in order to estimate the variability of sample processing and laboratory analysis. The term “split samples” is used by other agencies to identify a similar set of subsamples split from a single sample, except that each of the subsamples is analyzed by a different laboratory in order to investigate the bias between laboratories.

Clearly, it is desirable that project reports define the QC terms used in the report, explain how QC samples are collected or produced, and explain what potential sources of bias and variability are included and excluded in each type of sample.

Describe the Field Quality-Control Program

Project reports should describe the design of the field QC program and should specify which potential sources of bias and variability were considered important for assessment in view of the study objectives and possible outcomes. Examples of spatial sources include land use (agricultural, industrial, commercial, and others); land-management activities such as the presence of irrigation, tillage, and drainage; the presence of point or non-point sources of contaminants; geographic characteristics, such as altitude and climate; and the types of water sampled (streams, lakes, groundwater). Examples of temporal sources include seasonal cycles, streamflow conditions, water withdrawal or discharge schedules, and environmental conditions during sampling, such as the occurrence of precipitation, blowing dust, or aerial spraying.

Reports should include a description of the types and numbers of QC samples that were collected to assess potential sources of bias and variability and should indicate how these samples were distributed across the various sources. Explain which factors were considered in determining the types, number, and distribution of QC samples.

Summarize the Field Quality-Control Results

Project reports should summarize and interpret the field QC data in relation to the potentially important spatial and temporal sources of bias and variability targeted in the design of the field QC program. If data quality is similar across the different levels of a source, then QC samples could be pooled to estimate data quality. If data quality varies across the different levels of the source, then QC samples should be used to estimate the quality of data at each level. For example, if the frequency and magnitude of herbicide contamination in

field blanks is similar in urban and agricultural areas, then field blanks from both land uses could be pooled to estimate contamination, and land use does not need to be considered in assessing data quality. If herbicide contamination is much greater during the growing season than during the non-growing season, separate estimates of the frequency and magnitude of contamination should be made for each season, and the season needs to be considered in evaluating contamination effects on environmental data. Likewise, if any data-quality problems were identified and corrected during the course of the study, there will be two populations of data: one before and one after the correction. Data quality and its effect on interpretation of the environmental data will be different for each time period.

Tables of QC data provide little insight into data quality. Statistical summaries and graphical presentation of bias and variability are much more informative and provide evidence for or against pooling QC samples either spatially or temporally. Graphs are particularly effective for conveying data-quality information. Some pertinent graphs include (1) boxplots of data for different potential sources of bias and variability (for example, Martin and Eberle, 2011, fig. 2; Rowe and others, 2005, Appendix 1), (2) charts of data quality organized by chemical (for example, Mueller and Titus, 2005, fig. 2; Rowe and others, 2005, fig. 6), (3) scatterplots with a LOWESS smooth of data quality as a function of concentration or time (for example, Martin, 2002, fig. 4), and (4) cumulative frequency plots of data quality (for example, Bender and others, 2011, Appendix 1).

Statistical summaries of field blanks should include the number of blanks, the percentage of blanks with detections, percentiles of concentrations, and confidence limits for percent detections and percentile concentrations. Statistical summaries of field matrix spikes should include the number of spikes, the concentration spiked, median or mean recovery, variability of recovery (percentiles, standard deviation, relative standard deviation), and confidence limits for recovery statistics. Statistical summaries of field replicates should include the number of replicate sets, typical variability (standard deviation or relative standard deviation), and the applicable range of concentration for various estimates of variability.

Characterize Data Quality

Investigators must decide and define which datasets best characterize the bias and variability of the project's environmental data. These datasets might include field and laboratory QC data collected for the project, laboratory QC data from the NWQL or BQS, and QC data compiled from a national program or multiple projects within a Science Center. Small budget water-sampling projects might only be able to afford a few QC samples. Confidence intervals on statistics of bias and variability will be large for estimates based on this small QC dataset. Such projects should endeavor to use the same field protocols, equipment, supplies, analytical methods, and laboratories as are used by national programs or multiple projects within a Water Science Center. In doing so, the small amount of project QC data can be compared to the larger set of QC data for these programs or projects. If estimates of bias and variability are comparable, the small project is justified in using the larger QC dataset to characterize data quality for the small project.

The bias and variability of field data are unlikely to be less than the bias and variability of the analytical method as implemented by the laboratory. Because field QC samples are intended to assess additional sources of bias and variability attributed to field activities, the magnitude of bias (contamination) and variability measured by laboratory QC samples should be considered the minimum expected in field QC samples.

Consider Data Quality in Analysis and Interpretation

Project reports should provide evidence that data quality is adequate for all analysis and interpretation of environmental data, or explain how data analysis or interpretation had to be restricted. Evidence of adequate data quality includes quantitative indication that bias and variability were small in relation to critical data values. Examples of restricted data analysis include the following: (1) decided not to use data for some constituents, samples, sites, or time periods because of high bias or variability; (2) developed criteria for determining meaningful differences between single samples on the basis of variability of field replicates; (3) adjusted environmental data (for analysis) by some amount to account for contamination; (4) adjusted environmental data (for analysis) by some percent to account for bias in recovery. (Note that data values can be adjusted for project reporting and interpretation, but should not be changed in the NWIS database.) Examples of restricted data interpretation include the following: (1) failed to achieve or changed some project objectives or reduced project scope because of data-quality problems; (2) gave little emphasis to discussion or interpretation of some constituents, sites, or time periods because of potential bias or variability in environmental data; (3) determined that exceedance of water-quality standards or criteria was uncertain because analytical results could have negative bias or high variability; and (4) determined that the frequency of detection for some constituents might be overestimated because concentrations were similar to those in some field blanks.

Project reports should include statements that quantify the estimated bias and variability to the extent possible based on the QC data and should indicate whether the potential bias and variability based on these estimates has any effect on interpretation of the environmental data. If any results in the environmental data are determined to be of poor quality, add metadata to fields (such as sample and result comments) and codes (such as data-quality indicator codes, remark codes, and value qualifier codes) provided in NWIS to capture the outcome of the data-quality assessment (http://help.waterdata.usgs.gov/codes-and-parameters/codes#WQ; Dupre and others, 2013, Appendix A).

Ideally, the QC data can be used to show that bias and variability have a negligible effect, and that the environmental data are adequate for interpretation in support of project objectives. When high bias or variability are indicated, however, thoughtful analysis and publication of the QC information will help prevent misinterpretation or misuse of the data.

Acknowledgments

The material in this report was developed for the USGS training class "Quality-Control Sample Design and Interpretation." The authors thank the other instructors of the class for their help in developing the concepts and methods: Ed Gilroy, Stew McKenzie, Bob Broshears, Dennis Helsel, Kim Pirkey, and Amy Ludtke. Many other USGS scientists contributed to the development and dissemination of QA/QC guidance, including members of the National Water-Quality Assessment Program's QC Workgroup: Mike Koterba, Jon Scott, Greg Delzer, and Dave Bender. The report was improved by review comments provided by Alissa Coes, Greg Delzer, Mary Giorgino, Jim Kingsbury, and Dave Lorenz. We owe a special debt of gratitude to all the field hydrologists and technicians who collect the samples and to the laboratory chemists and technicians who analyze them. They play the most important role in ensuring the quality of USGS data.

References Cited

Anderson, R.L., 1987, Practical statistics for analytical chemists: New York, Van Norstrand Reinhold, 316 p.

Anderson, V.L and McLean, R. A., 1974, Design of Experiments—A Realistic Approach: New York, Marcel Dekker, Inc., 418 p.

Bender, D.A., Zogorski, J.S., Mueller, D.K., Rose, D.L., Martin, J.D., and Brenner, C.K., 2011, Quality of volatile organic compound data from groundwater and surface water for the National Water-Quality Assessment Program, October 1996–December 2008: U.S. Geological Survey Scientific Investigations Report 2011–5204, 128 p., http://pubs.usgs.gov/sir/2011/5204.

Chambers, J.M., Cleveland, W.S., Kleiner, Beat, and Tukey, P.A., 1983, Graphical methods for data analysis: Boston, Mass., Duxbury Press, 395 p.

Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: Journal of the American Statistics Association, v. 78, p. 605–610.

Dupré, D.H, Scott, J.C., Clark, M.L., Canova, M.G, and Stoker, Y.E., 2013, User's manual for the National Water Information System of the U.S. Geological Survey: Water-Quality System, Version 5.0: U.S. Geological Survey Open-File Report 2013–1054, 730 p., http://pubs.usgs.gov/of/2013/1054/pdf/OFR2013-1054_NWIS_ver5.pdf

Flegal, A.R., and Coale, K., 1989, Discussion—Trends in lead concentration in major U.S. rivers and their relation to historical changes in gasoline-lead consumption, by R.B. Alexander and R.A Smith: Water Resources Bulletin, v. 25, p. 1275–1277.

Hahn, G.J., and Meeker, W.Q., 1991, Statistical intervals—A guide for practitioners: New York, John Wiley and Sons, 392 p.

Helsel, D.R., 2005, Nondetects and data analysis—Statistics for censored environmental data: Hoboken, N. J., John Wiley and Sons, 250 p.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey, Techniques of Water-Resources Investigations, book 4, chap. A3, http://pubs.usgs.gov/twri/twri4a3.

Intergovernmental Data Quality Task Force, 2005a, Uniform federal policy for implementing environmental quality systems—Evaluating, assessing, and documenting environmental data collection/use and technology programs: U.S. Environmental Protection Agency report EPA-505-F-03-001, U.S. Department of Defense report DTIC ADA 395303, and U.S. Department of Energy report DOE/EH-0667, 114 p., accessed March 4, 2008, at http://www2.epa.gov/sites/production/files/documents/ufp_v2_final.pdf.

Intergovernmental Data Quality Task Force, 2005b, Uniform federal policy for quality assurance project plans—Evaluating, assessing, and documenting environmental data collection and use programs—Part 1, UFP-QAPP Manual: U.S. Environmental Protection Agency report EPA-505-B-04-900A and U.S. Department of Defense report DTIC ADA 427785, 154 p., accessed March 4, 2008, at http://www2.epa.gov/sites/production/files/documents/ufp_qapp_v1_0305.pdf.

Intergovernmental Data Quality Task Force, 2005c, Workbook for uniform federal policy for quality assurance project plans—Evaluating, assessing, and documenting environmental data collection and use programs—Part 2A, UFP-QAPP Workbook: U.S. Environmental Protection Agency report EPA-505-B-04-900C and U.S. Department of Defense report DTIC ADA 427486, 41 p., accessed March 4, 2008, at http://www2.epa.gov/sites/production/files/documents/part2ufp_wbk_0305.pdf.

Intergovernmental Data Quality Task Force, 2005d, Uniform federal policy for quality assurance project plans—Part 2B, Quality assurance/quality control compendium—Minimum QA/QC activities: U.S. Environmental Protection Agency report EPA-505-B-04-900B and U.S. Department of Defense report DTIC ADA 426957, 70 p., accessed March 4, 2008, at http://www2.epa.gov/sites/production/files/documents/qaqc_v1_0305.pdf.

Koterba, M.T., Wilde, F.D., and Lapham, W.W., 1995, Ground-water data-collection protocols for the National Water-Quality Assessment Program—Collection and documentation of water-quality samples and related data: U.S. Geological Survey Open File Report 95–399, 113 p., http://pubs.usgs.gov/of/1995/ofr-95-399/.

Maloney, T.J., ed., 2005, Quality management system, U.S. Geological Survey National Water Quality Laboratory: U.S. Geological Survey Open-File Report 2005–1263, version 1.3, November 9, 2005, chapters and appendixes variously paged, accessed May 16, 2015 at http://pubs.usgs.gov/of/2005/1263/pdf/OFR2005-1263.pdf.

Martin, J.D., 2002, Variability of pesticide detections and concentrations in field replicate water samples collected for the National Water-Quality Assessment Program, 1992–97: U.S. Geological Survey Water-Resources Investigations Report 2001–4178, 84 p., http://pubs.usgs.gov/wri/2001/wri01_4178/.

Martin, J.D., and Eberle, Michael, 2011, Adjustment of pesticide concentrations for temporal changes in analytical recovery, 1992–2010: U.S. Geological Survey Data Series 630, 11 p., 5 appendixes, http://pubs.usgs.gov/ds/630/.

Mueller, D.K., Hamilton, P.A., Helsel, D.R., Hitt, K.J., and Ruddy, B.C., 1995, Nutrients in ground water and surface water of the United States—An analysis of data through 1992: U.S. Geological Survey Water-Resources Investigations Report 95–4031, 74 p., http://pubs.usgs.gov/wri/1995/4031/report.pdf.

Mueller, D.K., Martin, J.D., and Lopes, T.J., 1997, Quality-control design for surface-water sampling in the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 97–223, 17 p., http://pubs.usgs.gov/of/1997/223/.

Mueller, D.K., and Titus, C.J., 2005, Quality of nutrient data from streams and ground water sampled during water years 1992–2001: U.S. Geological Survey Scientific Investigations Report 2005–5106, 27 p., http://pubs.usgs.gov/sir/2005/5106/.

National Environmental Laboratory Accreditation Conference, 2003, Appendix A—Glossary, chap. 1, p. 1A–1—1A–14, in NELAC standard, 2003, EPA/600/R-04/003, accessed July 23, 2011, at http://nelac-institute.org/docs/2003nelacstandard.pdf.

Nowell, L.H., Ludtke, A.S., Mueller, D.K., and Scott, J.C., 2013, Organic contaminants, trace and major elements, and nutrients in water and sediment sampled in response to the Deepwater Horizon oil spill: U.S. Geological Survey Scientific Investigations Report 2012–5228, 96 p., http://pubs.usgs.gov/sir/2012/5228/.

Ott, R.L., and Longnecker, M., 2001, An introduction to statistical methods for data analysis (5th ed.): Pacific Grove, Calif., Duxbury Press, 1152 p.

Ranalli, A.J., 2008, Water-quality data collected from Vallecito Reservoir, its inflows and outflow, southwestern Colorado, 1992–2002: U.S. Geological Survey Data Series 305, 76 p., http://pubs.usgs.gov/ds/305/.

Rickert, D.A., 1991, Programs and Plans—Dissolved trace element data: Office of Water Quality Technical Memorandum 91.10, accessed May 16, 2015, at https://water.usgs.gov/admin/memo/QW/qw91.10.html.

Rowe, B.L., Delzer, G.C., Bender, D.A., and Zogorski, J.S., 2005, Volatile organic compound matrix spike recoveries for ground- and surface-water samples, 1997–2001: U.S. Geological Survey Scientific Investigations Report 2005–5225, 51 p., http://pubs.usgs.gov/sir/2005/5225/.

Rupert, M.G., and Plummer, L.N., 2009, Groundwater quality, age, and probability of contamination, Eagle River watershed valley-fill aquifer, north-central Colorado, 2006–2007: U.S. Geological Survey Scientific Investigations Report 2009–5082, 59 p., http://pubs.usgs.gov/sir/2009/5082/.

Rus, D.L., Patton, C.J., Mueller, D.K., and Crawford, C.G., 2012, Assessing total nitrogen in surface-water samples—Precision and bias of analytical and computational methods: U.S. Geological Survey Scientific Investigations Report 2012–5281, 38 p., http://pubs.usgs.gov/sir/2012/5281/.

Sandstrom, M.W., Stroppel, M.E., Foreman, W.T., and Schroeder, M.P., 2001, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of moderate-use pesticides and selected degradates in water by C-18 solid-phase extraction and gas chromatography/mass spectrometry: U.S. Geological Survey Water-Resources Investigations Report 01–4098, http://nwql.usgs.gov/Public/pubs/WRIR/WRIR-01-4098.pdf.

Shiller, A.M., and Boyle, E.A., 1987, Variability of dissolved trace metals in the Mississippi River: Geochemica et Cosmochemica Acta, v. 51, p. 3273–3277.

Thiros, S.A., Bender, D.A., Mueller, D.K., Rose, D.L., Olsen, L.D., Martin, J.D., Bernard, Bruce, and Zogorski, J.S., 2011, Design and evaluation of a field study on the contamination of selected volatile organic compounds and wastewater-indicator compounds in blanks and groundwater samples: U.S. Geological Survey Scientific Investigations Report 2011–5027, 85 p., http://pubs.usgs.gov/sir/2011/5027/.

Twain, Mark, 1907, Chapters from my autobiography: North American Review, no. 186, issue 15, p. 465–474. [Available online at http://www.unz.org/Pub/NorthAmericanRev-1907jul05-00465.]

U.S. Environmental Protection Agency, 1986, Quality criteria for water 1986: Washington D.C., U.S. Environmental Protection Agency Report 440/5–86–001, Office of Water, variously paged.

U.S. Environmental Protection Agency, 1989, Risk assessment guidance for superfund volume I—Human health evaluation manual (part A) interim final: Office of Emergency and Remedial Response report EPA/540/1-89/002, 287 p., accessed March 25, 2011, at http://www.epa.gov/oswer/riskassessment/ragsa/pdf/rags-vol1-pta_complete.pdf.

U.S. Environmental Protection Agency, 1999, 1999 Update of ambient water quality criteria for ammonia: Washington D.C., U.S. Environmental Protection Agency Report EPA–822–R–99–014, Office of Water, variously paged, accessed April 1, 2004, at http://water.epa.gov/scitech/swguidance/standards/criteria/aqlife/ammonia/upload/99update.pdf.

U.S. Environmental Protection Agency, 2001a, EPA requirements for quality management plans: Office of Environmental Information report EPA/240/B-01/002, 24 p., accessed June 4, 2013, at http://www.epa.gov/quality/qs-docs/r2-final.pdf.

U.S. Environmental Protection Agency, 2001b, EPA requirements for quality assurance project plans: Office of Environmental Information report EPA/240/B-01/003, 24 p., accessed June 4, 2013, at http://www.epa.gov/quality/qs-docs/r5-final.pdf.

U.S. Environmental Protection Agency, 2008, USEPA contract laboratory program—National functional guidelines for Superfund organic methods data review: Office of Superfund Remediation and Technology Innovation report USEPA-540-R-08-01, 213 p., accessed March 6, 2014, at http://www.epa.gov/superfund/programs/clp/download/somnfg.pdf.

U.S. Environmental Protection Agency, 2010, USEPA contract laboratory program—National functional guidelines for inorganic Superfund data review: Office of Superfund Remediation and Technology Innovation report USEPA-540-R-10-011, 102 p., accessed March 6, 2014, at http://www.epa.gov/superfund/programs/clp/download/ism/ism1nfg.pdf.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9, http://pubs.water.usgs.gov/twri9A.

Wells, H.G., 1938, World Brain: Meuthuen & Co. Limited, 130 p. [Available online at http://www.ics.uci.edu/~vid/Readings/Wells_World_Brain.pdf.]

Windom, H.L., Byrd, J.T., Smith, R.G., Jr., and Huan, F., 1991, Inadequacy of NASQAN data for assessing metal trends in the nation's rivers: Environmental Science and Technology, v. 25, no. 6, p. 1137–1142.

Wright, P.R., McMahon, P.B., Mueller, D.K., Clark, M.L., 2012, Groundwater-quality and quality-control data for two monitoring wells near Pavillion, Wyoming, April and May 2012: U.S. Geological Survey Data Series 718, 23 p., http://pubs.usgs.gov/ds/718/.

Glossary

Definitions are grouped by topic and arranged within topics by pertinence, rather than alphabetically. Some definitions are quoted from other publications; these are in italics and referenced by the acronyms **EPA** (U.S. Environmental Protection Agency, 2001a), **IDQTF** (Intergovernmental Data Quality Task Force, 2005d), or the USGS **NWQL** (Maloney, 2005). Some of these quoted definitions were derived from the National Environmental Laboratory Accreditation Conference (**NELAC**) (National Environmental Laboratory Accreditation Conference, 2003). Definitions from these publications are not necessarily consistent with definitions used in this report but are provided for comparison. Some common QA/QC terms have not been specifically defined in USGS publications but are included herein with their IDQTF definition to make this list more comprehensive. A list of definitions for the variables used in equations in this report is included at the end of this section.

General QA/QC Terms

Quality assessment The overall process of assessing the quality of environmental data by reviewing the application of the quality-assurance elements and the analysis of the quality-control data.

Quality assurance (QA) Procedures used to control the non-quantifiable components of a project, such as sampling at the correct location with the proper equipment and using the appropriate methods.

(definition used by EPA, p. A–3 and IDQTF, p. 66; obtained from NELAC) *An integrated system of management activities involving planning, implementation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the client.*

Quality control (QC) Data generated to estimate the magnitude of the bias and variability in the process of obtaining environmental data.

(EPA, p. A–3 and IDQTF, p. 66; from NELAC) *The overall system of technical activities that measure the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer; operational techniques and activities that are used to fulfill requirements for quality; also the system of activities and checks used to ensure that measurement systems are maintained within prescribed limits, providing protection against "out of control" conditions and ensuring the results are of acceptable quality.*

Quality management plan (EPA, p. A–3) *A document that describes the quality system in terms of the organizational structure, functional responsibilities of management and staff, lines of authority, and required interfaces for those planning, implementing, and assessing all activities conducted.*

Quality assurance project plan (EPA, p. A–3, IDQTF, p. 66) *A formal document describing in comprehensive detail the necessary quality assurance, quality control, and other technical activities that must be implemented to ensure that the results of the work performed will satisfy the stated performance criteria.*

Data quality indicators (IDQTF, p. 7, 12, and 63) *The quantitative statistics and qualitative descriptors that are used to interpret the degree of acceptability or utility of data to the user. The principal data quality indicators are precision, accuracy/bias, comparability, completeness, representativeness, and sensitivity. Also referred to as data quality attributes.*

Data quality objectives (DQOs) (IDQTF, p. 63) *Qualitative and quantitative statements derived from the DQO process. DQOs can be used as the basis for establishing the quality and quantity of data needed to support decisions.*

Bias The systematic error inherent in a method or measurement system. The error can be positive (for example, contamination or spectral interference) or negative (for example, analyte loss or signal suppression).

Contamination bias Positive bias due to the inadvertent introduction of analytes into water samples during sample collection, processing, shipment, or analysis.

Variability Random error in independent measurements as the result of repeated application of the process under specific conditions. Variability can be statistically described by standard deviation, standard error, variance, or range in either absolute or relative terms.

Accuracy The degree of agreement between a measured value and the true or expected value. Accuracy is affected by both bias and variability, and cannot be independently determined.

> **(IDQTF, p. 13 and 62)** *The degree of agreement between an observed value and an accepted reference value. Accuracy includes a combination of random error (precision) and systematic error (bias), components which are due to sampling and analytical operations.*

Precision The degree of mutual agreement among independent measurements from the repeated application of a measurement process under identical conditions. Precision is the inverse of variability, but unlike variability, precision cannot be directly determined.

> **(IDQTF, p. 13 and 65)** *The degree to which a set of observations or measurements of the same property, obtained under similar conditions, conform to themselves.*

Sampling variability The variability introduced into sample measurements because of field procedures (collection, processing, and shipment) plus laboratory analysis.

QC-Sample Terms

Basic QC Samples QC samples that measure all of the potential sources of bias or variability that might affect environmental samples and are used to estimate the overall quality of the environmental data. Basic QC samples are field blanks, field matrix spikes, and field replicates.

Topical QC Samples QC samples that measure a limited number of sources of bias or variability; thus, they cannot be used to estimate the overall quality of environmental data.

Blank A sample prepared with water that is intended to be free of measurable concentrations of the analyte(s) of interest for determining contamination.

> **(NWQL, p. E.1; from NELAC)** *A sample that has not been exposed to the analyzed sample stream to monitor contamination during sampling, transport, storage, or analysis. The blank is subjected to the usual analytical and measurement process to establish a zero baseline or background value and is sometimes used to adjust or correct routine analytical results.*
>
> **(IDQTF, p. 62)** *A sample subjected to the usual analytical or measurement process to establish a zero baseline or background value. A sample that is intended to contain none of the analytes of interest. A blank is used to detect contamination during sample handling, preparation, and/or analysis.*

Field blank A sample of blank water that has been exposed in the field to all sampling equipment and conditions that normally are associated with the collection of an environmental sample.

> **(NWQL, p. E.3; from NELAC)** *A blank prepared on-site by filling a clean container with deionized water and appropriate preservative, if any, for the specific sampling activity being undertaken.*
>
> **(IDQTF, p. 63)** *A blank used to provide information about contaminants that may be introduced during sample collection, storage, and transport; also a clean sample, carried to the sampling site, exposed to sampling conditions, transported to the laboratory, and treated as an environmental sample.*

Equipment blank (IDQTF, p. 63) *A sample of water free of measurable contaminants poured over or through decontaminated field sampling equipment that is considered ready to collect or process an additional sample. The purpose of this blank is to assess the adequacy of the decontamination process. Also called rinse blank or rinsate blank.*

Trip blank (IDQTF, p. 68) *A clean sample of water free of measurable contaminants that is taken to the sampling site and transported to the laboratory for analysis without having been exposed to sampling procedures. Analyzed to assess the contamination introduced during sample shipment. Typically analyzed only for volatile organic compounds.*

Source-solution blank A sample of blank water taken directly from its source container without exposure to sampling equipment or conditions.

Method blank (IDQTF, p. 65) *A sample of a matrix similar to the batch of associated samples (when available) in which no target analytes or interferences are present at concentrations that impact the analytical results.*

It is processed simultaneously with samples of similar matrix and under the same conditions as the samples.

Reagent blank (IDQTF, p. 66) An aliquot of water or solvent free of measurable contaminants analyzed with the analytical batch and containing all the reagents in the same volume as used in the processing of the samples. The method blank goes through preparatory steps; the reagent blank does not.

> **(NWQL, p. E.7; from NELAC)** *A sample consisting of reagent(s), without the specified analyte or sample matrix, introduced into the analytical procedure at the appropriate point and carried through all subsequent steps to determine the contribution of the reagents and of the involved analytical steps.* [Note: NWQL also refers to these as "Method Reagent Blanks."]

Reference material A sample of sufficiently well-known composition to be used for assessment of biases in analytic methods.

> **(NWQL, p. E.7; from NELAC)** A *material or substance, one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials.*

Laboratory control sample (IDQTF, p. 64) *A sample of known composition prepared using reagent-free water or an inert solid that is spiked with analytes of interest at the midpoint of the calibration curve or at the level of concern. It is analyzed using the same sample preparation, reagents, and analytical methods employed for regular samples.*

Proficiency testing sample (IDQTF, p. 65–66) *A sample, the composition of which is unknown to the laboratory or analyst, which is provided to that analyst or laboratory to assess capability to produce results within acceptable criteria. Proficiency testing (PT) samples can fall into three categories: (1) prequalification, conducted prior to a laboratory beginning project work, to establish initial proficiency; (2) periodic (e.g., quarterly, monthly, or episodic) to establish ongoing laboratory proficiency; and (3) batch-specific, which is conducted simultaneously with analysis of a sample batch. A PT sample is sometimes called a performance evaluation sample.*

Spike (NWQL, p. E.8; from NELAC) *A known mass of specified analyte added to a blank sample or subsample; used to determine recovery efficiency or for other quality-control purposes.*

Matrix spike (IDQTF, p. 65) *A sample prepared by adding a known concentration of a target analyte to an aliquot of a specific homogenized environmental sample for which an independent estimate of the target analyte concentration is available. The matrix spike is accompanied by an independent analysis of the unspiked aliquot of the environmental sample. Spiked samples are used to determine the effect of the matrix on a method's recovery efficiency.*

Reagent spike (NWQL, p. E.7) *A synthetic matrix fortified with known concentrations of all, or a representative selection of, the method analytes. The synthetic matrix usually is the same as the method blank, for example, organic-free water or sodium sulfate. For the purpose of interpreting the corrective action guidelines described in this document, a reagent spike failure is defined as an out-of-control recovery for any relevant spiked analyte.*

Replicates Two or more environmental samples taken at the same time in the same location. They are intended to estimate sampling variability and are taken through all steps of the analytical procedure in an identical manner.

Split replicates Replicate samples prepared by taking representative portions from a single sample in the field or laboratory.

Subsample duplicates (IDQTF, p. 64) *Samples resulting from one sample collection at one sample location.*

Concurrent replicates Multiple samples that are collected in the same location at about the same time.

Sequential replicates Multiple samples that are collected in the same location one after another.

Co-located duplicates (IDQTF, p. 64) *Samples collected from side-by-side locations at the same point in time and space.*

Irreplicates Replicates used to investigate some difference in the data-generation process. Typically, the goal of collecting irreplicates is to assess the comparability of data generated differently. Irreplicates are not used to assess sampling variability.

Split samples (IDQTF, p. 67) *Two or more representative portions taken from one sample in the field or laboratory, analyzed by at least two different laboratories and/or methods.*

Variables Used in Equations

α	a specified significance level (0 through 1); the probability that a confidence interval does not include the true value
$\Delta C_{interval}$	the confidence interval for the difference between mean concentrations in two sets of data
ΔC	the difference in two mean concentrations
δ	potential uncertainty (0 through 1)
μ	the population mean
ϕ	population proportion
σ	the standard deviation of a measured concentration, independently estimated from replicate variability
χ^2	the percentage point of the chi-square distribution for a specified area and degrees of freedom
B	the binomial probability function
B_0	the intercept of a regression line, estimated by least-squares
B_1	the slope of a regression line, estimated by least-squares
bcf	the bias-correction factor
C_{env}	the concentration of an analyte in the background environmental sample, in micrograms per liter
$C_{expected}$	the concentration of a spiked analyte expected in the spiked sample, in micrograms per liter
C_{sol}	the concentration of an analyte in the spike solution, in micrograms per milliliter
C_{spike}	the concentration of an analyte in the spiked matrix sample, in micrograms per liter
CI	the overall width of a confidence interval
d	the half-width of a confidence interval
df	the degrees of freedom used to determine a statistical value (t, Z, F, or x^2)
C_L	the lower limits of concentration for the 100(1–α/2) percent confidence interval
C_U	the upper limits of concentration for the 100(1–α/2) percent confidence interval
F	the percentage point of the F distribution for a specified area and degrees of freedom
L	the rank of the lower 100(1–α)-percent confidence limit
log(SD)	the logarithm of replicate standard deviation
log(C)	the logarithm of mean replicate concentration
n	the sample size (for example, number of observations or number of replicate sets)
p	percentile
$\hat{p}$	the proportion of quantified values within the total number of observations in a dataset
P_U	the upper confidence limit, in percent
q	the number of quantified values within a dataset
R	recovery from a spiked sample, in percent
RPD	relative percent difference
s	the standard deviation of the sample data
SD	the standard deviation of a set of replicate samples
SD_{diff}	the standard deviation for the difference between mean concentrations in two sets of data
SD_{FV}	field variability, which is introduced by sampling and laboratory procedures
SD_{reps}	the standard deviation for a specified analyte concentration estimated using one of the replicate models defined under "Evaluating Variability in Analyte Concentration" in this report
SD^2x_i	the variance of dataset x_i
n_i	the number of samples (observations) in dataset x_i
SU	the upper confidence limit on the true standard deviation
t	the percentage point of Student' t distribution for a specified area and degrees of freedom.
U	the rank of the upper 100(1–α)-percent confidence limit
Var	the variance of a set of observations
V_{sol}	the volume of spike solution added to the spiked sample, in milliliters
V_{sample}	the volume of the matrix sample, in liters
x	the number of replicate sets with inconsistent detections
$\bar{x}$	the mean of a random sample of data (for example mean recovery from field spikes, in percent)
Z	the percentage point of the standard normal curve that contains a specified area

Appendix 1

Table 1–1. Nitrate plus nitrite data used in the replicate analysis example (figs. 13, 15, 17, and 19–22), compiled from Rus and others (2012).

[mg/L, milligrams per liter]

Mean concentration (mg/L)	Standard deviation (mg/L)	Relative standard deviation (percent)	Mean concentration (mg/L)	Standard deviation (mg/L)	Relative standard deviation (percent)
0.011	0.0014	12.86	0.964	0.0113	1.17
0.015	0.0014	9.43	1.068	0.0035	0.33
0.027	0.0007	2.67	1.120	0.0014	0.13
0.033	0.0028	8.57	1.207	0.0071	0.59
0.058	0.0014	2.44	1.388	0.0148	1.07
0.131	0.0007	0.54	1.525	0.0205	1.35
0.166	0.0021	1.28	1.555	0.0269	1.73
0.168	0.0007	0.42	1.618	0.0028	0.17
0.194	0.0007	0.37	1.901	0.0064	0.33
0.196	0.0028	1.44	2.109	0.0276	1.31
0.211	0.0007	0.34	2.418	0.0085	0.35
0.223	0.0000	0.00	2.731	0.0488	1.79
0.247	0.0007	0.29	2.763	0.0686	2.48
0.352	0.0014	0.40	2.805	0.0064	0.23
0.464	0.0000	0.00	2.875	0.0191	0.66
0.469	0.0127	2.71	3.373	0.0198	0.59
0.479	0.0007	0.15	3.559	0.0049	0.14
0.492	0.0064	1.29	3.873	0.0346	0.89
0.563	0.0014	0.25	3.948	0.0007	0.02
0.605	0.0042	0.70	4.064	0.1047	2.58
0.614	0.0057	0.92	4.441	0.0233	0.53
0.641	0.0042	0.66	4.695	0.0120	0.26
0.652	0.0007	0.11	4.936	0.0396	0.80
0.734	0.0028	0.39	4.941	0.0035	0.07
0.789	0.0057	0.72	5.352	0.0120	0.22
0.883	0.0141	1.60	5.367	0.0693	1.29
0.905	0.0014	0.16	6.134	0.0269	0.44
0.943	0.0078	0.83	8.049	0.0827	1.03
0.945	0.0014	0.15	11.967	0.0445	0.37
0.954	0.0014	0.15	14.123	0.0552	0.39

Table 1–2. Atrazine data used in the replicate analysis example (figs. 14, 16, 17, and 19–22), compiled from Martin (2002).

[µg/L, micrograms per liter]

Mean concentration (µg/L)	Standard deviation (µg/L)	Relative standard deviation (percent)	Mean concentration (µg/L)	Standard deviation (µg/L)	Relative standard deviation (percent)
0.0015	0.00000	0.00	0.0052	0.00014	2.72
0.0020	0.00000	0.00	0.0053	0.00014	2.67
0.0020	0.00000	0.00	0.0059	0.00007	1.21
0.0020	0.00000	0.00	0.0059	0.00021	3.63
0.0020	0.00042	21.21	0.0060	0.00000	0.00
0.0025	0.00035	14.43	0.0061	0.00049	8.18
0.0025	0.00007	2.89	0.0061	0.00049	8.18
0.0028	0.00007	2.57	0.0061	0.00163	26.88
0.0028	0.00014	5.05	0.0061	0.00021	3.51
0.0029	0.00000	0.00	0.0064	0.00049	7.79
0.0031	0.00021	6.96	0.0064	0.00014	2.21
0.0036	0.00021	5.98	0.0066	0.00057	8.57
0.0037	0.00120	32.93	0.0070	0.00000	0.00
0.0039	0.00000	0.00	0.0073	0.00078	10.73
0.0041	0.00014	3.45	0.0074	0.00049	6.73
0.0042	0.00021	5.11	0.0075	0.00071	9.43
0.0042	0.00007	1.70	0.0075	0.00014	1.89
0.0042	0.00233	56.23	0.0076	0.00064	8.43
0.0044	0.00021	4.88	0.0082	0.00191	23.43
0.0044	0.00127	28.93	0.0082	0.00113	13.80
0.0045	0.00000	0.00	0.0084	0.00021	2.54
0.0047	0.00014	3.01	0.0084	0.00148	17.78
0.0048	0.00014	2.95	0.0084	0.00000	0.00
0.0049	0.00014	2.89	0.0088	0.00049	5.66

Table 1–2. Atrazine data used in the replicate analysis example (figs. 14, 16, 17, and 19–22), compiled from Martin (2002).—Continued

[µg/L, micrograms per liter]

Mean concentration (µg/L)	Standard deviation (µg/L)	Relative standard deviation (percent)
0.0089	0.00014	1.59
0.0091	0.00445	49.22
0.0094	0.00028	3.01
0.0095	0.00064	6.73
0.0095	0.00106	11.22
0.0095	0.00071	7.44
0.0098	0.00361	36.99
0.0098	0.00028	2.89
0.0099	0.00035	3.59
0.0100	0.00035	3.55
0.0100	0.00000	0.00
0.0106	0.00071	6.67
0.0110	0.00000	0.00
0.0117	0.00177	15.17
0.0118	0.00085	7.19
0.0119	0.00035	2.98
0.0126	0.00085	6.73
0.0129	0.00057	4.39
0.0130	0.00205	15.83
0.0130	0.00057	4.35
0.0144	0.00035	2.46
0.0147	0.00071	4.81
0.0148	0.00120	8.15
0.0151	0.00014	0.94
0.0154	0.00064	4.15
0.0158	0.00028	1.79
0.0177	0.00021	1.20
0.0185	0.00113	6.12
0.0189	0.00014	0.75
0.0196	0.00057	2.89
0.0200	0.00141	7.07
0.0210	0.00035	1.69
0.0232	0.00085	3.66
0.0295	0.00339	11.51
0.0304	0.00290	9.55
0.0307	0.00049	1.61
0.0324	0.00049	1.53
0.0353	0.00721	20.43
0.0390	0.00141	3.63
0.0402	0.00127	3.17
0.0430	0.00424	9.87
0.0433	0.00064	1.47
0.0480	0.00099	2.06
0.0481	0.00035	0.74
0.0482	0.00078	1.62
0.0485	0.00071	1.46
0.0491	0.00255	5.18
0.0494	0.00481	9.73
0.0499	0.00375	7.52
0.0517	0.00537	10.39
0.0520	0.00219	4.22
0.0560	0.00092	1.64
0.0599	0.00537	8.97
0.0603	0.00071	1.17
0.0613	0.00163	2.66
0.0620	0.00000	0.00
0.0629	0.00071	1.12
0.0640	0.00332	5.20
0.0655	0.00099	1.51
0.0660	0.00849	12.86
0.0706	0.00014	0.20
0.0708	0.00460	6.50
0.0786	0.00049	0.63
0.0816	0.00721	8.84
0.0830	0.00141	1.70
0.0838	0.01280	15.28
0.0885	0.00354	3.99
0.0940	0.00424	4.51
0.0958	0.00078	0.81
0.0998	0.00877	8.79
0.1015	0.00212	2.09
0.1100	0.00000	0.00
0.1200	0.00000	0.00
0.1245	0.00212	1.70
0.1295	0.00071	0.55
0.1375	0.00212	1.54
0.1500	0.00000	0.00
0.1525	0.00071	0.46
0.1875	0.00354	1.89
0.1975	0.00354	1.79
0.2150	0.00707	3.29
0.2225	0.01061	4.77
0.2260	0.01131	5.01
0.2305	0.01626	7.06
0.2370	0.00566	2.39
0.2385	0.00919	3.85
0.2400	0.01414	5.89
0.2655	0.01485	5.59
0.2775	0.00495	1.78
0.3370	0.00707	2.10
0.3375	0.01202	3.56
0.3600	0.00707	1.96
0.4015	0.00354	0.88
0.4730	0.01697	3.59
0.5000	0.01414	2.83
0.5440	0.01273	2.34
0.6725	0.00495	0.74
1.2300	0.04243	3.45
1.5500	0.07071	4.56
1.8300	0.00000	0.00
1.8950	0.04950	2.61
1.9300	0.00000	0.00

Publishing support provided by:
Denver Publishing Service Center, Denver, Colorado

For more information concerning this publication, contact:
Chief, USGS Office of Water Quality
412 National Center
Reston, VA 20192
(703) 648-6862

Or visit the Office of Water Quality Web site at:
http://water.usgs.gov/owq/

This publication is available online at:
http://dx.doi.org/10.3133/tm4c4

ISSN 2328-7055 (online)
http://dx.doi.org/10.3133/tm4C4