Data-Quality Measures for Stakeholder-Implemented Watershed-Monitoring Programs

By Adrienne I. Greve

U.S. GEOLOGICAL SURVEY

Open-File Report 02–141

Denver, Colorado
2002

U.S. DEPARTMENT OF THE INTERIOR
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For additional information write to:

District Chief
U.S. Geological Survey
Box 25046, Mail Stop 415
Denver Federal Center
Denver, CO 80225–0046

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Box 25286
Denver Federal Center
Denver, CO 80225

CONTENTS

Glossary V
Abstract 1
Introduction 1
    Purpose and Scope 2
    Acknowledgments 2
Quality Assurance 2
    Possible Quality-Assurance Approaches for Stakeholder Groups 3
Quality Control 4
    Quality-Control Sample Design 4
        Determining the Potential Sources of Error 4
        Type of Quality-Control Samples Needed 5
        Determining the Inference Space for Quality-Control Samples 6
        Number of Quality-Control Samples Needed 6
        Distribution of Quality-Control Samples within an Inference Space 9
Quality Assessment 10
    Ongoing Quality-Assessment Measures 10
    Using Quality-Assessment Measures during Environmental Data Analysis 11
        Estimating Variability by Using Field-Replicate Quality-Control Samples 11
        Interpreting Environmental Data by Using Field-Replicate Quality-Control Samples 12
        Estimating Bias by Using Field-Blank Quality-Control Samples 12
        Interpreting Environmental Data by Using an Estimate of Bias 13
        Estimating Matrix Interaction and Sample Degradation with Field-Spike Data 13
        Interpreting Environmental Data by Using Spike-Recovery Data 14
A Data-Quality Program in the Big Thompson River Watershed 14
    Quality Assurance for the Big Thompson Watershed Forum 14
    Quality Control for the Big Thompson Watershed Forum 16
        Types of Basic Quality-Control Samples to be Collected 16
        Types of Topical Quality-Control Samples to be Collected 16
        Inference Space 17
        The Number of Quality-Control Samples to be Collected 17
        Distribution of Quality-Control Samples 17
Summary and Conclusions 18
References 18

FIGURES

1–3. Graphs showing:
    1. The number of blank samples relative to the percentile for six different confidence levels 7
    2. The percentage of the standard deviation possible for varying replicate sample sizes and upper confidence levels 8
    3. Example of a piecewise linear estimation of replicate deviation 12
4. Map showing location and land use in the Big Thompson River watershed 15
5. Flow chart showing the Big Thompson Watershed Forum monitoring program design process 16

TABLES

1. Descriptions of some of the most common types of quality-control samples 5
2. The number of blank samples needed so that the maximum detected concentration in a blank sample represents an estimate of the selected upper confidence level for the selected percentiles 7
3. The minimum number of blank samples so that the second highest concentration is an estimate of the specified upper confidence limit for the specified percentiles 8
4. Estimates of variability measured as a percentage of average standard deviation based on the number of replicate pairs collected and an upper confidence level 8
5. The minimum number of spike sets required for the specified confidence level and confidence interval half-width expressed as a proportion of the standard deviation 9

GLOSSARY

Accuracy.—The amount of agreement between a measured value and the true environmental value.

Basic quality-control sample.—A sample used to quantify most or all possible sources of bias and variability within a sampling program, including sampling, processing, transport, and analysis.

Bias.—A systematic error in a data set where values are consistently high or low.

Blank sample.—Water, free of the analyte of interest, is run through all or part of the sampling, processing, transport, and analysis procedures. Blank samples are used to estimate high bias.

Confidence.—The chance that the true environmental value is within a defined range.

Inference space.—The relation of a set of environmental samples to a given set of quality-control samples.

Precision.—The amount of agreement between independent measurements of the same quantity.

Replicate sample.—A set of samples (two or more) assumed to be identical in composition. Replicate samples are used to estimate variability.

Spike sample.—A sample fortified with a known concentration of specific constituents. Spike samples are used to estimate bias due to degradation or matrix interference.

Topical quality-control sample.—A sample used to identify possible sources of bias and variability within a specific part of the sampling program.

Uncertainty.—The chance that the true environmental value is outside a defined range.

Variability.—The random error present in independent measurements of the same quantity.

Data-Quality Measures for Stakeholder-Implemented Watershed-Monitoring Programs

By Adrienne I. Greve

Abstract

Community-based watershed groups, many of which collect environmental data, have steadily increased in number over the last decade. The data generated by these programs are often underutilized due to uncertainty in the quality of data produced. The incorporation of data-quality measures into stakeholder monitoring programs lends statistical validity to data.

Data-quality measures are divided into three steps: quality assurance, quality control, and quality assessment. The quality-assurance step attempts to control sources of error that cannot be directly quantified. This step is part of the design phase of a monitoring program and includes clearly defined, quantifiable objectives, sampling sites that meet the objectives, standardized protocols for sample collection, and standardized laboratory methods. Quality control (QC) is the collection of samples to assess the magnitude of error in a data set due to sampling, processing, transport, and analysis. In order to design a QC sampling program, a series of issues needs to be considered: (1) potential sources of error, (2) the type of QC samples, (3) inference space, (4) the number of QC samples, and (5) the distribution of the QC samples. Quality assessment is the process of evaluating quality-assurance measures and analyzing the QC data in order to interpret the environmental data. Quality assessment has two parts: one that is conducted on an ongoing basis as the monitoring program is running, and one that is conducted during the analysis of environmental data.

The discussion of the data-quality measures is followed by an example of their application to a monitoring program in the Big Thompson River watershed of northern Colorado.

INTRODUCTION

During the last decade, the number of community-based watershed groups has increased substantially (Kenney and others, 2000; River Network, 2001). More than 3,600 such groups are currently active in the United States (River Network, 2001). The groups typically focus on a single watershed and generally are composed of a combination of community members, private industry, and government agencies. A broad range of environmental issues including public education, land-use policies, water quality, habitat, and biota are addressed by these groups. In addition, many of these stakeholder groups have undertaken collaborative data-collection projects or have individual members that collect environmental samples. These projects can be funded (meaning the staff collecting, processing, and analyzing the samples are compensated for their time), entirely volunteer, or some combination of the two. The collected data can fill gaps in governmental monitoring programs such as those operated by the U.S. Environmental Protection Agency (USEPA), U.S. Geological Survey (USGS), and city or State governments and can span institutional boundaries such as State lines or city boundaries. Unfortunately, however, much of the data collected by stakeholder groups are not accepted by all potential users due to uncertainty about the quality of the data (U.S. Environmental Protection Agency, 1996). The incorporation of data-quality measures into stakeholder-implemented monitoring programs would lend statistical validity to the data and allow for more potential data users.

Environmental data, due to collection, processing, transport, and analysis, inherently has some bias and variability associated with it. In order to accurately interpret environmental data, this error must be identified and the magnitude estimated. Data-quality measures serve this purpose (Mueller and others, 1997).

Monitoring programs implemented by a single entity or agency currently have fewer data-quality concerns than programs implemented by watershed groups. These differences are due to the involvement of multiple entities, often with varying priorities, monitoring goals, or sampling and analysis protocols. Each of these factors must be accounted for through each step of the data-quality design: quality assurance, quality control, and quality assessment.

Purpose and Scope

This report is intended to provide an introduction and basic guide to data-quality measures for a stakeholder group that is initiating or operating a water-quality monitoring system. The term "stakeholder group" is used in this report to represent community-based watershed efforts. Other terms that are commonly used to describe community watershed groups include watershed councils, forums, and initiatives.

The design of a data-quality system is composed of three steps: quality assurance, quality control, and quality assessment. Each of these steps is discussed, paying particular attention to the obstacles that commonly confront stakeholder monitoring programs. The discussion of these steps is followed by an example from the Big Thompson River watershed, located in northern Colorado. Because the monitoring program in the Big Thompson River watershed has just begun (2000), only the first two data-quality steps are described.

Acknowledgments

The methods, definitions, and approach to data quality described in this report are based on those developed by the USGS and were presented to the author during a training course developed by Terry Schertz, Jeffrey Martin, David Mueller, Mark Sandstrom, and Robert Broshears. The report would not have been possible without the help of the Big Thompson Watershed Forum, most specifically Rob Buirgy.

The author would also like to thank David Mueller and Tony Ranalli for technical review, Mary Kidd for editorial review, and Joy Monson for layout.

QUALITY ASSURANCE

Quality assurance attempts to control those sources of bias and variability that cannot be directly quantified. This step is part of the design phase of a monitoring program. Quality assurance is integral to a high-quality design. Characteristics of a high-quality monitoring design include the following:

- **Clearly defined, quantifiable objectives.** All parts of a monitoring design, including the data-quality measures, are based on the informational needs or objectives of the program. In order to ensure that the monitoring program will meet the informational needs of the stakeholders, the objectives need to be clear and quantifiable. Each subsequent step in the design and implementation is evaluated on the basis of the objectives. Clearly defined objectives also ensure that all involved entities can share a common expectation of the type of data that will be produced.
- **Sampling sites that meet the objectives.** The samples collected must represent the water body of concern, as identified in the objectives. The site should be free of unique characteristics that would cause the samples to differ in composition from the stream or subbasin of interest. This means that the sites selected must be evaluated on the basis of upstream sources of the water, mixing distances if there is an upstream confluence or discharge, and the availability of some means to collect a sample both at high and low flows. For example, if a site is chosen to represent the overall quality of the water draining from a subbasin, the site probably should not be located directly downstream from a point-source discharge, where the discharge may obscure any signal or trend resulting from changes occurring farther upstream in the basin.
- **Standardized protocols for sample collection.** This step requires not only that standard methods are used but that they are appropriate for the information required to meet the stated objectives. Each

sample at each site should, ideally, be collected in an identical manner. The manner of sampling includes the sampling personnel, sampling equipment, and sample-collection methods. Standard methods limit possible sources of variability introduced during sampling. In some cases, standardization may not be possible. If, due to cost, time, or other constraints, different sets of sampling equipment, multiple sampling crews, or even different methods are unavoidable, the potential error associated with these differences must be addressed. Possible approaches to account for these differences are discussed in the following section and in the quality-control step.

Methods also must be appropriate for the objectives of the program. Some instruments and sampling methods are appropriate only within a certain range of concentration or a given level of precision. The methods chosen must produce data that meet the program objectives. Methods also should be evaluated to limit possible contamination. For example, if a water sample is to be analyzed for trace elements, certain metal samplers or processing apparatus may be inappropriate.

- **Standardized laboratory methods.** Similar to the other steps, the choice of laboratory and analytical methods should be based on the program objectives. One of the first issues that must be evaluated is the level of precision, or the detection limit, required to meet the objectives. Typically, the lower the concentrations a method can detect, the more the analysis costs. Therefore, the expected concentration and informational needs of the program objectives should be evaluated to identify the analytical method most appropriate for both for the objectives and the budget.

These steps, in addition to being addressed, also must be documented in detail. Documentation of the design not only allows for consistency in its implementation but also allows for an evaluation of the success of the stated objectives, and necessary changes can be more easily made.

Possible Quality-Assurance Approaches for Stakeholder Groups

The quality-assurance step poses the greatest challenge to stakeholder groups. First, a cooperative, stakeholder-initiated monitoring program often occurs because no individual member or entity has the resources to meet their informational goals alone. The financial support of all or most members is critical to the success of the program. Therefore, the design is most often best achieved through consensus. The monitoring design, which includes objectives, constituent list, sampling locations, sampling frequency, and sampling protocols, is intended to meet the minimum informational needs of all stakeholders. Developing a design can be a lengthy process involving open communication, compromise, and patience.

Several strategies have been described for encouraging group consensus and open communication within a stakeholder group (Natural Resources Law Center, 1996; Kenney and others, 2000; U.S. Environmental Protection Agency, 1996 and 1997; Goeldner, 1996; Buzan and others, 1996). A group leader or facilitator often can help streamline the collaborative design process. Ideally, a facilitator would not have a vested interest in the outcome of the design process but would be able to balance the various interests and priorities of the group members. In addition to a facilitator, a system of feedback and communication helps to ensure that all viewpoints and opinions are heard. Because the success of cooperative-monitoring programs relies on the support of most or all members, a system to gather input and solicit feedback throughout the design process that allows potential areas of conflict to be identified early is a key element.

Once the monitoring network has been designed, the implementation of that design can be planned. Implementation includes the choice of sampling methods, laboratories, and personnel. Quality-assurance measures require each option be evaluated to limit sources of bias and variability. Ideally, in order to limit error, samples would be collected at all sites by the same crew using the same methods and the same equipment, and the samples would be analyzed at a single laboratory. This much uniformity is difficult for stakeholder groups not only because of limited funds but also because of constraints on where the money may be spent. Individual members of a stakeholder group may already have sampling protocols, equipment, staff, and(or) a laboratory. Such entities cannot easily divert funds currently supporting equipment, staff, and laboratories to an outside contractor. In addition, “in kind” support from group members in the form of equipment, staff time, or laboratory work generally is critical to making stakeholder monitoring programs financially viable. Therefore, stakeholder groups may lack uniform protocols and may frequently use multiple sampling crews, equipment sets, and laboratories.

In order to address these challenges, stakeholder groups can attempt to limit error in their design and implementation plan within the constraints of the group. A single set of sampling protocols should be established for the entire monitoring program. The sampling crew, even if composed of staff from several entities within the stakeholder group, should be trained together or in the same manner. The type of equipment used at each site should be uniform. Preservation, transport, and laboratory analysis also should be uniform. If any of these goals cannot be achieved, additional measures (discussed in the "Quality Control" section) can be taken. For example, if multiple laboratories are to be used, it is best to have a single laboratory conduct all the analyses for a given constituent. This reduces error among sampling sites for a single constituent.

QUALITY CONTROL

Quality control (QC) is the collecting of samples and subsequent generation of data used to assess the magnitude of bias and variability in a data set due to sampling, processing, transport, and analysis. There are three general types of QC samples: blanks, replicates, and spikes.

Blank samples. Blank samples are intended to be free of the analyte of interest (Mueller and others, 1997). Therefore, the samples are used to identify contamination, also termed high bias. Bias refers to a systematic error in data such as concentrations being consistently lower or higher than the environmental concentration.

Replicate samples. Replicate samples are intended to be water samples identical in composition (Mueller and others, 1997). This allows the variability to be assessed.

Spike samples. Spike samples are water samples fortified with a known amount of the analyte of interest. Spikes are used to assess bias due to matrix interference or analyte degradation (Mueller and others, 1997).

These three types of QC samples also can be grouped on the basis of potential sources of error represented by the sample. Table 1 has descriptions of several types of blank, replicate, and spike samples as well as the grouping in which they belong based on the potential sources of the error being assessed.

Basic quality-control sample. This term describes a sample used to quantify most or all possible sources of bias and variability within a sampling program. The three types of basic quality-control samples are field replicates, field blanks, and field spikes. These three types of samples are collected in the field and are intended to assess possible sources of error occurring during sample collection, processing, transport to the laboratory, and analysis.

Topical quality-control sample. This term describes a sample used to identify possible sources of bias and variability within a specific part of the sampling program such as sampling equipment, laboratory analysis, or sample transport.

Quality-Control Sample Design

Quality-control sample design requires each monitoring program to determine how many and what type of QC samples are required to meet the informational goals or objectives of the monitoring program. In order to design a system of QC samples, a series of issues needs to be considered: (1) potential sources of error, (2) the type of QC samples, (3) inference space, (4) the number of QC samples, and (5) the distribution of the QC samples in an inference space.

Determining the Potential Sources of Error

This initial step in QC sample design is twofold. First, the potential sources of error in the monitoring program are identified. From these potential sources, the errors most likely to affect the interpretation of the environmental data should be identified. The identification of error sources acts as a guide for choosing the types of QC samples needed to quantify error in the system. Determining the potential sources of error that are likely to affect the interpretation of environmental data is based on the magnitude of the potential error and the expected environmental concentrations. For example, if the potential error is dwarfed by the expected environmental concentration, the error is not likely to affect the interpretation of the environmental data.

Some common sources of potential error include the following:

- **Multiple sampling crews.** Error may occur due to a change in sampling personnel at some point during the monitoring process or when

Table 1. Descriptions of some of the most common types of quality-control samples

Sample	Sample type	Description[1]
Field blank	Basic	Water, free of the analyte of interest, is run through all sampling and processing equipment at the stream-sampling site, stored as an environmental sample, transported, and analyzed at the laboratory.
Equipment blank	Topical	Water, free of the analyte of interest, is run through some or all sampling equipment, placed in a bottle, and analyzed at the laboratory. This sample can originate in an office or laboratory.
Laboratory blank	Topical	Water, free of the analyte of interest, is analyzed at the laboratory.
Trip blank	Topical	A bottle of water, free of the analyte of interest, is stored with the environmental samples during transport and analyzed at the laboratory.
Ambient blank	Topical	Water, free of the analyte of interest, is exposed to the ambient conditions at a sampling site, transported, and analyzed at the laboratory.
Field replicate	Basic	A field replicate is a set of samples (two or more) assumed to be identical in composition. There are several types of replicate samples including split, concurrent, and sequential replicates.
Split replicate	Basic	Two or more samples resulting from splitting a single volume of sample into multiple samples.
Concurrent replicate	Basic	Two or more samples collected at the same location at the same time. In order to collect these samples, two sampling crews are required.
Sequential replicate	Basic	Two or more samples collected at the same location, but at different times, typically one after the other.
Field spike	Basic	An environmental sample is fortified with a known concentration of specific constituents at the sampling site, transported, and analyzed at the laboratory.
Standard reference sample	Topical	A sample with known concentrations of specific constituents is analyzed at the laboratory. This differs from a laboratory spike in that the concentrations should be similar to those found in the environmental sample.
Laboratory replicate	Topical	A sample is split in the laboratory and analyzed as two separate samples.
Laboratory spike	Topical	Blank water or sample water is fortified with a known concentration of specific constituents in the laboratory.

[1]Descriptions based on A.J. Ranalli (U.S. Geological Survey, written commun., 2000), Mueller and others (1997), and Mueller (1998).

a stakeholder group chooses to split the network of sites. If these crews have different manners of sampling, the potential sources of error are different. For example, one crew may be more prone to contamination than another.

- **Differences in sampling methods.** Difference in methods, either over time or among the sampling sites, potentially will have differences in error. Some methods are more variable than others or prone to differing levels of contamination. If methods with different sources of error are used within a single inference space (see Glossary), it would inflate the estimates of variability or attach an estimate of bias to samples for which no bias may be present.
- **Different equipment.** Similar to methods and sampling personnel, different types of equipment will have different errors associated with it.
- **Different environmental sources of error.** Differences in potential environmental sources of error can be contaminants external to the stream or specific stream characteristics that cause an area to be more or less prone to bias or variability than other sites, such as low ionic strength.

Type of Quality-Control Samples Needed

In addition to the identified potential sources of error, the parameters for which the water samples will be analyzed, the expected concentrations of those parameters, and the operation of the monitoring program influence the type of QC samples needed. A QC sampling program is composed primarily of basic QC samples. A schedule of field replicates, spikes, and blanks is typically the basis of a QC program. These samples allow overall bias and variability to be estimated. Degradation and matrix interference are greater concerns for certain groups of parameters such as volatile organic compounds (VOC) and pesticides. Therefore, if a monitoring program does not include any parameters that are prone to degradation or matrix interference, field spikes can be excluded from the QC sample design.

Topical QC samples are commonly collected less often than basic QC samples and should be collected for a specific purpose. A system of topical QC samples is used in two situations. First, if the operation of the monitoring design includes a potential source error likely to affect the interpretation of environmental data, such as multiple sampling crews, sampling methods, equipment, or laboratories, a set of topical QC samples should be collected to identify differences in methods or establish comparability. For example, the collection of concurrent replicates, where two samples are collected at the same time with different crews, methods, or equipment, provides data that allow comparability to be assessed. The second situation that requires a set of topical QC samples is if errors of a magnitude that substantially affects the interpretation of environmental data are identified from basic QC sampling. For example, if a problem such as contamination is identified from basic QC samples, a system of topical QC samples can be implemented in an attempt to identify the source of the bias.

Determining the Inference Space for Quality-Control Samples

Inference space refers to the relation of a set of environmental samples to a given set of QC samples. For example, if a field blank is collected and analyzed and no detectable contamination is found, does that mean all samples that day, all samples at that sampling site, or all samples at all sampling sites are free from detectable contamination? Generally, the largest possible inference space should initially be assumed for a monitoring program. If the measures taken in the quality-assurance step (such as standardized cleaning, sampling, and transport) are implemented, the error introduced by multiple sets of equipment, sampling crews, or multiple laboratories should be limited. Following the collection of QC samples, the data can be analyzed and the inference space broken up for specific types of contamination. For example, if one sampling crew is shown to consistently contaminate samples, all blank samples, and subsequent estimates of bias, should only be associated to that crew until it can be demonstrated that the problem has been solved.

There is the possibility that a monitoring program will have components with so many differences in potential sources of error that they can be separated into different inference spaces prior to the collection of QC samples. Such a situation could include a monitoring network which combines funded and volunteer efforts. If there is more than one inference space in a sampling program, it does not mean that data from different inference spaces cannot be used together for comparisons, trends, or any other analysis. What it does imply, however, is that the errors associated with data from each of the inference spaces might be different.

Number of Quality-Control Samples Needed

The minimum number of QC samples needed will depend on the uncertainty in estimates of bias and variability that is acceptable for meeting the program goals. The more QC samples that are collected, the less the uncertainty in bias and variability estimates. However, as the number of QC samples increases, the degree of improvement in the estimates of error decreases. For example, increasing the number of QC samples from 10 to 11 will improve the estimate of error more than increasing the number from 20 to 21. A monitoring program must determine how much uncertainty can be accepted and how much confidence needs to be attained to meet the monitoring goals while staying within the available budget. The answer to these questions will depend on the streams being sampled, the goals of the monitoring program, and the schedule for data analysis. For example, if a stream has extremely high concentrations of a given constituent, contamination is not likely to be a large percentage of the environmental sample. In this case, a higher level of uncertainty and a lower level of confidence will likely still meet program goals. In a program where the primary objective is to determine compliance with a standard, the level of confidence and acceptable uncertainty will depend on how close to the standard environmental samples are expected to be. If a sample concentration is close to the standard, a higher level of confidence with a low amount of uncertainty likely will be required; however, if environmental concentrations are extremely low in comparison to a standard, more uncertainty and less confidence will still meet the program goals. The timing of data analysis also affects the number of QC samples, especially in the early stages of an ongoing monitoring program. If a program is meant to continue indefinitely but also to produce annual reports, enough QC samples should be collected prior to the first analysis for the first report to meet the minimum informational requirements. Subsequent years will have the benefit of all QC samples collected during prior sampling years, assuming a continuing inference space.

Blank-sample size. Blank samples are used to estimate bias due to sampling, processing, transport, and analysis. During the analysis of QC data, there will likely be a range of concentrations found in the

blank-sample data. It is not likely that these concentrations will be distributed normally, the distribution assumed by many statistical methods. The problem of nonnormal distributions is solved by using percentiles, which assume no underlying distribution. A percentile is a nonnormal statistical measure of variation, in this case, the variation in the concentrations of blank samples. A percentile, such as 85, refers to a data point where 15 percent of all data points are greater in value and 85 percent are less. Percentiles are calculated by ranking the data from lowest to highest. They are based on the rank of a data point rather than the concentration. The calculation of blank-sample size is based on an evaluation of the confidence with which a given percentile may be estimated. Because a percentile is determined on the basis of ranked data, the more blank samples available to be ranked, the higher the percentile that can be estimated. In addition, higher numbers of blank samples allow the desired percentile to be estimated without using the highest contamination value. In other words, the more samples that are collected, the less likely an outlier (an unusually high or low value) will strongly influence the estimate of bias. Blank-sample sizes for commonly used percentile and confidence levels are given in table 2 and figure 1. The minimum number of blank samples required to estimate a given percentile, at a given confidence level without the use of the highest ranked blank concentration, is listed in table 3. The equations from which the sample sizes in the table are calculated are equations 1, 2, and 3. The blank-sample-size estimation is based on the binomial distribution.

$$\text{given: } 100(1-\alpha) = B(p, n, y) \tag{1}$$

$$\text{if: } y \equiv n \tag{2}$$

$$\text{then: } n = \frac{\log\alpha}{\log p} \tag{3}$$

where

$100(1-\alpha)$ is confidence,

B is the binomial distribution,

p is the percentile,

n is the number of samples, and

y is the rank of the sample representing p.

[Equations 1, 2, and 3 were used for blank-sample size in Schertz, Martin, Sandstrom, Mueller, and Broshears (U.S. Geological Survey, written commun., 2000); an explanation of the binomial distribution, equations 1 and 2, is included in Ott (1993).]

Table 2. The number of blank samples needed so that the maximum detected concentration in a blank sample represents an estimate of the selected upper confidence level for the selected percentiles. For example, in order to be 75-percent confident that the highest concentration of contamination detected in a blank to represents the 75th percentile, five field-blank samples should be collected

[%, percent]

Percentile	Upper confidence level						
	60%	70%	75%	80%	90%	95%	99%
60	2	3	3	4	5	6	10
70	3	4	4	5	7	9	13
75	4	5	5	6	9	11	16
80	5	6	7	8	11	14	21
90	9	12	14	16	22	29	44
95	18	24	28	32	45	59	90
99	92	120	138	161	230	299	459

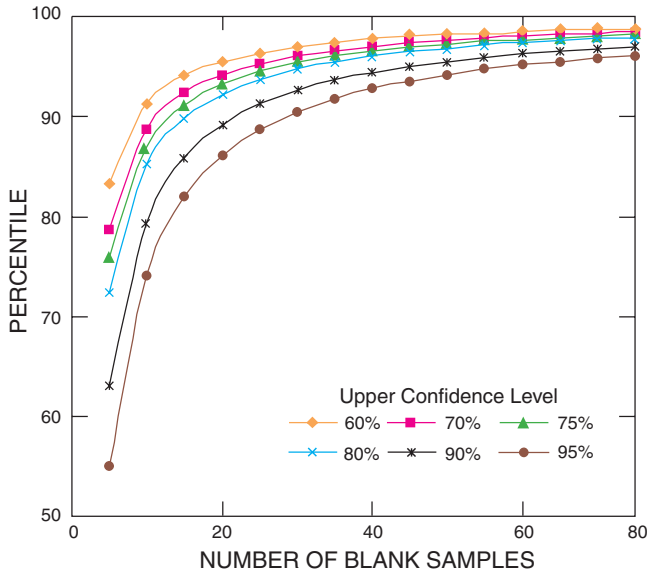

Figure 1. The number of blank samples relative to the percentile for six different confidence levels (a graphical representation of table 2).

Replicate Sample Size. Replicate samples are used to estimate variability due to sampling, processing, transport, and analysis. Variability is estimated based on the differences in detected concentration between samples of water presumed to be identical. Replicate sample-size calculations determine the resolution with which variability can be estimated. The variability is expressed as a percentage of the standard deviation of the replicate samples. The confidence in estimates of standard deviation that can be

achieved using a selected number of replicates is given in table 4 and figure 2. Equation 4, which uses the chi-square distribution, is used to generate table 4 and figure 2.

$$\delta = \left(\frac{df}{\chi^2_{\alpha,\,df}}\right)^{1/2} - 1 \qquad (4)$$

$$df = n$$

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

where

$100(1-\alpha)$ is confidence;

df is the degrees of freedom (for a pooled estimate of replicate standard deviation, $df = n$, where n = the number of replicate pairs);

δ is the uncertainty expressed as a percentage of standard deviation; and

χ^2 is the chi-square distribution.

[Equation 4 was used for replicate sample size in Schertz, Martin, Sandstrom, Mueller, and Broshears (U.S. Geological Survey, written commun., 2000).]

Field-spike sample size. Field spikes provide information about the effect of stream chemistry on analytical determination, or matrix interference, and on constituent degradation. Spikes are used to evaluate the low bias of reported concentrations. The number of spikes to be collected can be based on the width of a confidence interval around a mean spike recovery. One or more spikes (a spike set) are commonly collected to accompany an environmental sample. Analyte recovery is calculated for each spike set. Sample size requirements can be calculated on the basis of the width of a confidence interval about the mean recovery for all spike sets. A confidence interval is based on the standard deviation of the recoveries of all the spike sets collected. Because the standard deviation is not known prior to sampling, the sample size calculation is based on desired confidence level and the confidence interval half-width, expressed as a proportion of the unknown standard deviation. The smaller the proportion, the narrower the confidence interval and larger the required sample size. The

Table 3. The minimum number of blank samples so that the second highest concentration is an estimate of the specified upper confidence limit for the specified percentiles

[%, percent]

Percentile	Upper confidence level		
	80%	90%	95%
60	6	8	10
70	8	11	14
80	13	17	21
90	25	37	44
95	51	71	89

Table 4. Estimates of variability measured as a percentage of average standard deviation based on the number of replicate pairs collected and an upper confidence level

[%, percent]

Replicate pairs	Upper confidence level					
	60%	70%	75%	80%	90%	95%
5	121%	135%	144%	156%	194%	237%
10	111%	119%	124%	129%	147%	165%
15	108%	114%	117%	122%	134%	146%
20	106%	111%	114%	118%	128%	137%
25	105%	110%	112%	115%	124%	132%
30	105%	109%	111%	114%	121%	128%
35	104%	108%	110%	112%	119%	125%
40	104%	107%	109%	111%	118%	123%
45	104%	107%	109%	111%	116%	122%
50	103%	106%	108%	110%	115%	120%

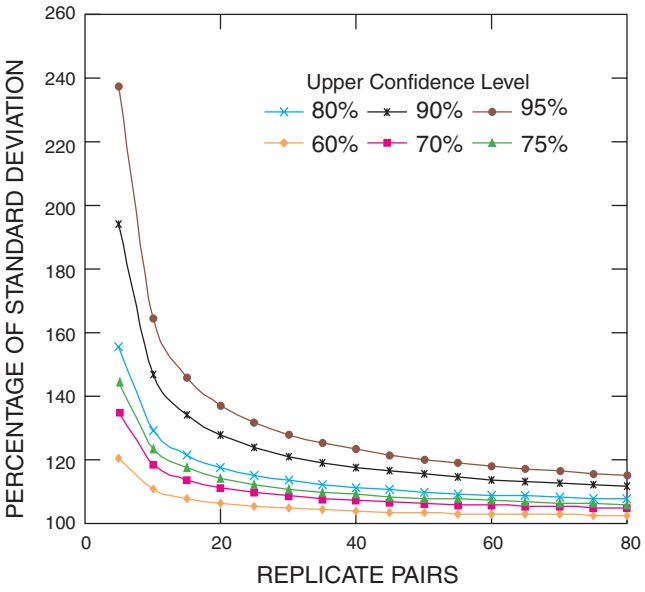

Figure 2. The percentage of the standard deviation possible for varying replicate sample sizes and upper confidence levels.

sample size for field-spike sets for common confidence levels and proportions of the standard deviation is given in table 5. The equations used to generate the sample sizes listed in table 5 follow (equations 5 and 6).

$$n = \left(\frac{Z_{(1-\alpha/2)}}{k}\right)^2 \quad (5)$$

$$k = \frac{d}{\sigma} \quad (6)$$

where

n is the number of field-spike sets;

Z is the standard normal, or Z, distribution;

$100(1-\alpha)$ is the confidence;

k is the proportion of standard deviation which equals the confidence interval half-width;

d is the confidence interval half-width; and

σ is the standard deviation of the average recovery for each spike set.

[Equation 5 is based on the calculation for a confidence interval about a mean assuming a normal distribution. It is described in several statistical texts such as Hahn and Meeker (1991) and Ott (1993).]

Field spikes are commonly collected to detect degradation or matrix interaction. However, spike data is prone to a high level of variability often due to differences in field procedure. Consistency in spike sample preparation is difficult. In some cases, such as the USGS National Water-Quality Assessment Program, a monitoring program has chosen to use laboratory spikes instead of field spikes in order to eliminate the variation due to methods in the field and focus on the effects of degradation and matrix interaction.

The number of environmental samples, as well as QC samples, is often driven by financial constraints. In addition to the planned basic QC samples, a portion of the data-quality budget needs to be set aside in order to add topical QC samples if a problem is identified. The final point to keep in mind is that these decisions commonly are made with little or no historical data. Following the first year of sample collection, it might be determined that the planned number and(or) type of QC samples should be changed. A monitoring program, including the data-quality component, must be dynamic and consistently evaluated to determine if it continues to meet the goals of the stakeholder group.

Distribution of Quality-Control Samples within an Inference Space

Within an inference space, once the number and type of QC samples have been determined, the distribution of the samples can be decided. At this point a decision can be made between random and targeted sampling or some combination of the two. Samples can be randomized or targeted in relation to the number of environmental samples, spatially within the basin, through time, over the hydrologic cycle.

Random sampling. A random sampling design distributes QC samples through time, space, and(or) among environmental samples without preference. The advantage of a randomized design is that, theoretically, all possible conditions are equally likely to be sampled. However, this might not happen if conditions that affect bias and variability are not uniformly distributed throughout the inference space.

Targeted sampling. A targeted design concentrates samples in a particular part of the sampling program. As a result, some part of the system might be overrepresented. However, there are benefits to targeting samples. Samples can be targeted on the basis

Table 5. The minimum number of spike sets required for the specified confidence level and confidence interval half-width expressed as a proportion of the standard deviation

[%, percent]

Confidence level	Confidence interval half-width expressed as a percentage of the standard deviation				
	30%	40%	50%	60%	70%
95%	43	25	16	11	8
90%	31	17	11	8	6
80%	19	11	7	5	4
75%	15	9	6	4	3
70%	12	7	5	3	3

of the hydrologic cycle, specific areas of the watershed, or in time. Quality-control samples can be targeted toward the beginning of a study for two reasons: to identify possible problems early in the program, and to have enough QC samples at the first analysis of environmental data to allow an estimate of bias and variability needed to meet the informational needs of the program. Targeting can allow some types of QC samples to be more meaningful. For example, a replicate does not provide an accurate estimate of variability if both the environmental sample and the replicate sample(s) are censored (reported as less than a given concentration). In order to avoid this situation, replicate samples, in areas where concentrations can be low, should be targeted for the time of year when the concentrations are most likely to be above the detection limit.

QUALITY ASSESSMENT

Quality assessment is the process of evaluating quality-assurance measures and analyzing the QC data. Quality assessment has two parts: one that is conducted on an ongoing basis as the monitoring program is running, and one that is conducted during the analysis of environmental data.

Ongoing Quality-Assessment Measures

These measures include the checking of data returned from the laboratories and the evaluation of field sheets in order to verify that the sampling and processing protocols are being followed.

- **Field sheet check.** Reviewing the field sheet ensures that the protocols for sample collection are being followed. The field sheet check can include the following measures: the sampling-site name and identification number, calibration data for any field measurement, and environmental conditions. Checking field sheets is especially critical if a concentration reported by the laboratory appears unusual. The field sheet may reveal that the sample was collected during an extreme event or unusual circumstance. A final step verifies field data are correctly entered into a database.
- **Environmental sample checks.** When concentrations are reported by the laboratory, a series of checks should be completed. The first set of these checks can be termed "logic checks," which include an ion balance, making sure total concentrations are greater than or equal to dissolved concentrations, and that the sum of the parts is equal to the total concentration within a specified margin for error.

  Ion balance:

  $$\textit{Concentration of major cations in milliequivalents} \cong \textit{Concentration of major anions in milliequivalents}$$

  Major cations: calcium, magnesium, sodium, potassium

  Major anions: sulfate, chloride, fluoride, carbonate, bicarbonate

  The second set of checks involves viewing the reported concentrations in the context of samples previously collected from a site. This involves verifying that the concentration makes sense at a given site for the flow condition and time of year. If any of these checks reveals a concentration that is unusual, the field sheets should be checked in order to determine if it can be explained by an extreme event in the field. If a clear explanation is not evident, the laboratory should be contacted to verify that it was not a data-entry error or analysis error. Laboratories should keep samples for a specified period of time, so reported concentrations that do not follow the typically observed concentration ranges or seasonal variations can be analyzed a second time to verify or replace the concentration in question.
- **QC sample checks.** As field QC sample concentrations are reported from the analyzing laboratory, the data should be evaluated for signs of gross contamination or other errors. If an error is suspected, the field notes should be consulted for extreme circumstances, and the laboratory should be contacted to check for data-entry errors or a sample rerun. If the QC data are deemed valid and if such bias or variability would threaten the usefulness of the environmental data, the collection of topical QC samples should be considered in order to identify the possible source of the error. If an error source is identified and the methods adjusted, samples collected after the adjustment in sampling method should be considered part of a separate inference space.

  Recoveries, reported as a percentage, should be calculated for each spike (equations 7 and 8). These recoveries then should be compared to laboratory spike data. This comparison allows the cause of potential degradation or amplification to

be narrowed down. If the laboratory spike into blank water has a low recovery, low recoveries in the environmental sample may be due to analysis procedures; however, if the blank water spike has recoveries that are near 100 percent, a poor recovery in the field spike more likely is due to matrix interaction. If the results of the spike recovery analysis reveal that the desired information is not being determined, adjustment can be made, particularly if poor recoveries are due to laboratory analysis. If laboratory methods are adjusted, the samples analyzed using the new methods should be considered part of a new inference space.

$$Recovery = \frac{(C_{spiked} - C_{unspiked}) \times 100}{C_{expected}} \quad (7)$$

$$C_{expected} = \frac{C_{solution} \times V_{spike}}{V_{sample}} \quad (8)$$

where

C_{spiked} is the concentration measured in the field-spike sample,

$C_{unspiked}$ is the concentration measured in the companion environmental sample,

$C_{expected}$ is a calculated concentration based upon the volume of water to which a spike of known volume and concentration is added,

$C_{solution}$ is the known concentration in the spike solution,

V_{spike} is the volume of spike solution added, and

V_{sample} is the volume of the spiked sample,

[Equations 7 and 8 are from the National Water Quality Laboratory (1996).]

Using Quality-Assessment Measures During Environmental Data Analysis

Quality-assessment measures taken during the analysis of the environmental data involve the estimation of bias and variability and the evaluation of inference space. These estimates can affect the interpretation of environmental data. The first step requires that the QC samples are evaluated in space and time to check if the assumed inference space is correct for each type of QC sample. Quality assessment allows for the combining of data from different inference spaces. For example, data generated by a funded monitoring program and volunteer monitoring program could be used together. In order to conduct analysis on data from different inferences spaces, error (bias and variability) must be associated with each data set. Subsequent analysis can then account for the fact that the magnitude of the bias and variability associated with each data set may vary.

Estimating Variability by Using Field-Replicate Quality-Control Samples

Replicate data are used to estimate the variability. Variability is determined by an estimate of standard deviation and can be used to define a confidence interval about a single sample concentration or a mean concentration from several samples. Variability for many chemical constituents increases at higher concentrations. This relation can be identified by plotting standard deviation against the average concentration in each set of replicates (equations 9 and 10).

$$SD = \sqrt{\frac{\Sigma[C_i - \overline{C}]^2}{n-1}} \quad (9)$$

$$\overline{C} = \frac{\Sigma C_i}{n} \quad (10)$$

where

SD is the standard deviation;

C is the mean concentration of a replicate set;

n is the number of replicate samples in the set; and

C_i is the concentration of an individual sample in the set.

If no relation is identified, an overall standard deviation should be used. If a relation is evident, it should be quantified. One of the simplest methods to define the relation is through a piecewise linear approach. Often, the relation is not a single, constant linear one but can be defined as a set of linear relations broken up by concentration range (fig. 3). Each piece of the relation can be defined through a mean standard deviation or a best-fit linear regression line. By determining an estimation of standard deviation for the full concentration range, a confidence interval can be placed on environmental data.

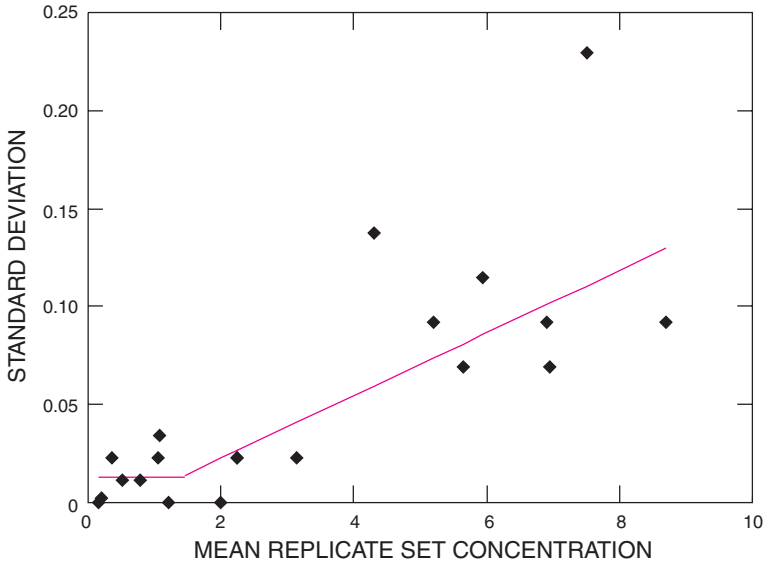

Figure 3. Example of a piecewise linear estimation of replicate deviation.

Interpreting Environmental Data by Using Field-Replicate Quality-Control Samples

Replicate data are used to estimate the uncertainty of environmental data: for example, determining the uncertainty of a single environmental sample or the minimum difference between means that can be determined with confidence. Uncertainty of a single environmental sample can be determined by using equation 11.

$$C_{interval} = C_{sample} \pm Z_{(1-\alpha/2)}SD \qquad (11)$$

where

$C_{interval}$ is the confidence interval about an environmental concentration,

C_{sample} is the concentration of an environmental sample,

Z is the standard normal distribution,

$100(1-\alpha/2)$ is the confidence, and

SD is the standard deviation of the replicate data for the concentration range in which C_{sample} fits.

[Equation 11 is presented in Schertz, Martin, Sandstrom, Mueller, and Broshears (U.S. Geological Survey, written commun., 2000) and Ott (1993).]

An estimate of the minimum difference between means that can be determined with confidence can be calculated by using equation 12. This method can be used to compare mean concentrations at different sites or different time periods.

$$\Delta C_{interval} = \Delta C \pm Z_{(1-\alpha/2)}SD_{diff} \qquad (12)$$

$$SD_{diff} = \sqrt{\frac{SD_{R1}^2}{n_1} + \frac{SD_{R2}^2}{n_2}}$$

where

ΔC is the difference between two mean concentrations,

$\Delta C_{interval}$ is the confidence interval around a difference in mean concentrations based solely on sampling variability,

Z is the normal distribution,

SD_{R1} is the standard deviation of the replicates associated with the first mean concentration, and

n_1 is the number of replicate sets in each of inference spaces associated with the first mean concentration.

[Equation 12 is presented in Schertz, Martin, Sandstrom, Mueller, and Broshears (U.S. Geological Survey, written commun., 2000).]

If the interval includes 0, a difference as large as DC is too small to identify given sampling variability.

Estimating Bias by Using Field-Blank Quality-Control Samples

Field-blank data are used to estimate bias, or contamination. Analysis is conducted on each inference space individually. The first step is to plot the

concentrations of the blanks over space (by site) and time (by sample date). This plot allows for a visual inspection of the data to see that all samples belong in a single inference space. If a marked break or systematic difference in typical levels of contamination is observed, the assumption of a single inference space should be reevaluated. The analysis of multiple blank samples from a single inference space assumes the upper confidence level for a specified percentile of the blank-sample concentrations is representative of all samples in an inference space including environmental samples. Equation (1) is used to determine the rank, *y*, that equals or exceeds the selected confidence level $(1 - \alpha)$ and percentile, *p*. The next task is to determine the concentration of the data point at the rank determined, *y*. Rank the blank-sample concentrations from low to high, and the concentration at the rank determined is the upper confidence level for the specified percentile, an estimate of overall bias.

Interpreting Environmental Data by Using an Estimate of Bias

The estimate of bias present in data from a single inference space can influence the interpretation of environmental data. The estimate of bias should be evaluated with respect to the environmental concentrations and the objectives of the monitoring program. If the estimate of bias is a large percentage of many of the environmental concentrations, the type of conclusions that can be drawn from the environmental data should be adjusted. For example, if there are low environmental concentrations and the estimate of bias is more than 50 percent of many of the samples, the certainty with which the environmental concentrations may be viewed is reduced. One possible solution would be to raise the censoring level up to a point where the estimate of bias is a lower percentage of the uncensored concentration. Another consideration is the comparison of the bias estimate to a water-quality standard. If the estimate of bias is a large percentage of the water-quality standard, compliance of environmental concentrations cannot be accurately determined.

Estimating Matrix Interaction and Sample Degradation with Field-Spike Data

The use of field-spike data is similar to that of field blanks. The data are used to determine an estimate of some systematic error in the data. The first step is to assess the assumed inference space. Spike recoveries can change due to recalibration of laboratory instruments, the use of different machines, and the different analyses. Changes in recovery due to procedural changes can be determined by plotting spike recoveries in time and by communicating with the laboratory to identify points where there were changes.

Once the inference space has been evaluated, the estimate of error is based on the mean percent recovery of field spikes (equation 13). A standard deviation then is calculated and a confidence interval constructed around the mean (equation 14) in the same manner as that used in equation 11.

$$\overline{R}_{all} = \frac{\sum_{i=1}^{n_{spike}} \overline{R}_i}{n_{spike}} \tag{13}$$

$$\overline{R}_i = \frac{\sum_{r=1}^{n_{reps}} R_r}{n_{reps}}$$

$$CI = \overline{R}_{all} \pm t_{(1-\alpha/2, n_{spike}-1)} SD_R \tag{14}$$

where

R_{all} is the mean of the average recoveries from each field-spike set,

R_i is the average recovery for a single field-spike set,

R_r is the recovery for a single field-spike sample,

n_{spike} is the number of spike sets collected,

n_{reps} is the number of spike samples in a field-spike set,

CI is the confidence interval,

t is the student's *t* distribution,

$100(1 - \alpha/2)$ is the confidence, and

SD_R is the standard deviation of R_{all}

$$SD_R = \sqrt{\frac{(\overline{R}_{all} - \overline{R}_i)^2}{n_{spike} - 1}}$$

Interpreting Environmental Data by Using Spike-Recovery Data

Recovery is used to provide an estimate of certainty to the environmental data. In all cases, the estimate of recovery should be reported with the environmental data. In some cases, the environmental data can be adjusted. If the percent recovery is near 100, the environmental data can be used without adjustment. However, if the percent recovery is poor (greater than 140 percent or less than 60 percent, for example) and reflects a large percentage of the environmental data, adjustment can be made. If the recovery is poor, but the standard deviation is small, the environmental concentrations could be adjusted to estimate 100-percent recovery. If this is done, it needs to be noted in the environmental data analysis and the spike analysis. If the recovery is poor and the standard deviation is large, the data need to be evaluated in order to determine if the program objectives can be met.

A DATA-QUALITY PROGRAM IN THE BIG THOMPSON RIVER WATERSHED

The Big Thompson River watershed is located in northern Colorado, along the east side of the Continental Divide (fig. 4). Water from the Big Thompson River and the Colorado Big Thompson Project, a water diversion project, is used for many purposes including municipal supply, irrigation, industry, recreation, and riverine habitat support. More than one-half million people depend on the Big Thompson system for drinking water. In 1996, a study by the North Front Range Water Quality Planning Association (NFWQPA) (Jeff Writer, North Front Range Water Quality Planning Association, written commun., 1996) recommended the establishment of a collaborative watershed group aimed to increase communication among stakeholders, conduct scientifically sound studies of the human effects on water quality, and educate the public to heighten the awareness of the watershed and associated water quality. The Big Thompson Watershed Forum (BTWF) was established in 1996 to satisfy this recommendation. The first year of BTWF operation focused primarily on the organization and stability of the new group. The BTWF established a consistent group of participants, hired a coordinator (facilitator), and applied for grant moneys. One of the first major projects of the BTWF was the design of a cooperative monitoring program. The BTWF was awarded a USEPA Regional Geographic Initiative grant for this purpose. The grant money was matched by contributions from five BTWF members. The monitoring design budget was used to fund a graduate student attending Colorado State University (CSU) to guide the BTWF through the design process.

Quality Assurance for the Big Thompson Watershed Forum

The design process had five components: objectives, parameters, sampling locations, sampling frequency, and cost analysis. Within each component, an iterative process was conducted through a series of meetings. First a draft, based on informal conversation, was written and presented to a small group made up of the five funding entities. Based on the needs of the five funding entities, a second draft was produced. This draft was presented to the general assembly of the BTWF. Again, feedback was solicited, which resulted in a third draft. This process was carried out for each of the first four components. The cost analysis (component 5) was then conducted. Based on the financial constraints of the BTWF, each of the first four components was revisited and the iterative process repeated (fig. 5). A detailed description of the design process and resulting monitoring network is available in Greve (1999).

Once the network design was completed, the BTWF faced another collaborative design task: choosing sampling and analysis methods. Several BTWF members were involved in water-quality monitoring; however, no member acting alone had the resources to implement the newly designed monitoring network. Rather than invest the time and energy into developing and documenting sampling protocols, the BTWF chose to cooperate with the USGS. In order for the samples to meet USGS standards, USGS protocols and methods were used (U.S. Geological Survey, 1997). This documented standardization included sampling and processing methods, sampling-site locations, and field sampling forms. The BTWF, however, could not afford to send all samples to an external USGS laboratory because three members were currently operating or working with a laboratory. As a result, four laboratories were selected. Due to the involvement with the USGS, the three laboratories not operated by the USGS were required to undergo an evaluation by the USGS Branch of Quality Systems (BQS). This evaluation is the manner in which the

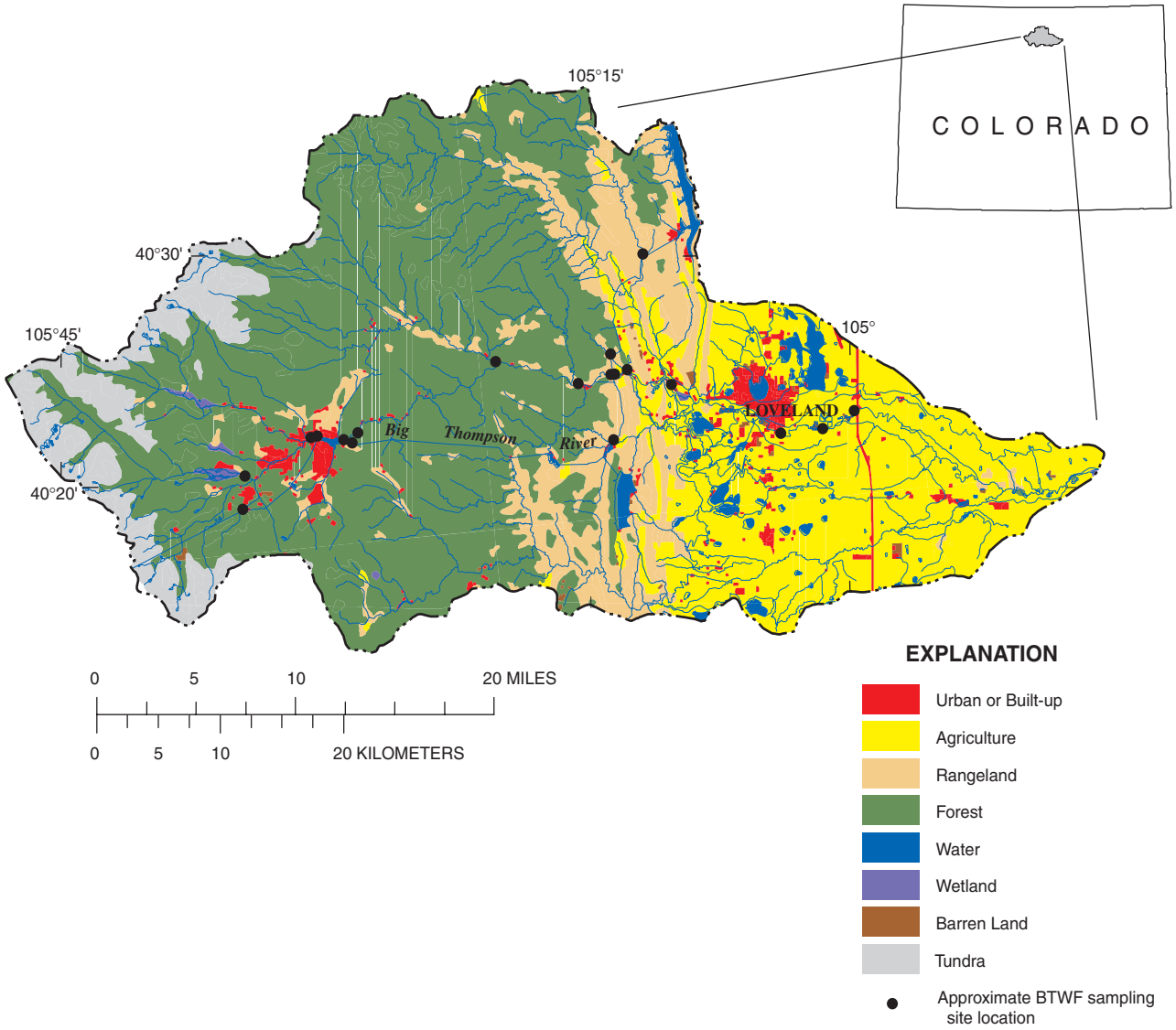

Figure 4. Location and land use in the Big Thompson River watershed. [BTWF, Big Thompson Watershed Forum; land use based on GIRAS (Geographic Information Retrieval and Analysis System) land-use data from the 1970's (Fegeas and others, 1983), and refined with 1990 population data (Hitt, 1995).]

USGS is able to ensure consistent data quality and therefore incorporate data from the laboratories into its National Water Information System (NWIS) database. Each laboratory is responsible for a specific subset of the parameter list. For example, none of the laboratories currently being used by BTWF members could detect nutrient concentrations low enough to meet the monitoring objectives. Therefore, the USGS National Water Quality Laboratory is being used to do the nutrient analysis. The City of Fort Collins laboratory is conducting analysis for a limited number of organic compounds. The City of Loveland is performing bacterial analysis, and Acculabs, a private laboratory, is conducting the trace-element and major chemistry analyses for all sampling sites. Dividing up the parameter list ensures that the analysis for an individual parameter is consistent among all sampling sites within the watershed.

The BTWF wanted active involvement in collecting samples. A graduate student from CSU was funded as a representative of the BTWF to participate in the sample collection. The student's salary became a part of the cooperative agreement between the USGS and the BTWF. In addition to sampling responsibilities, the student has filled the role of liaison between the USGS and the BTWF.

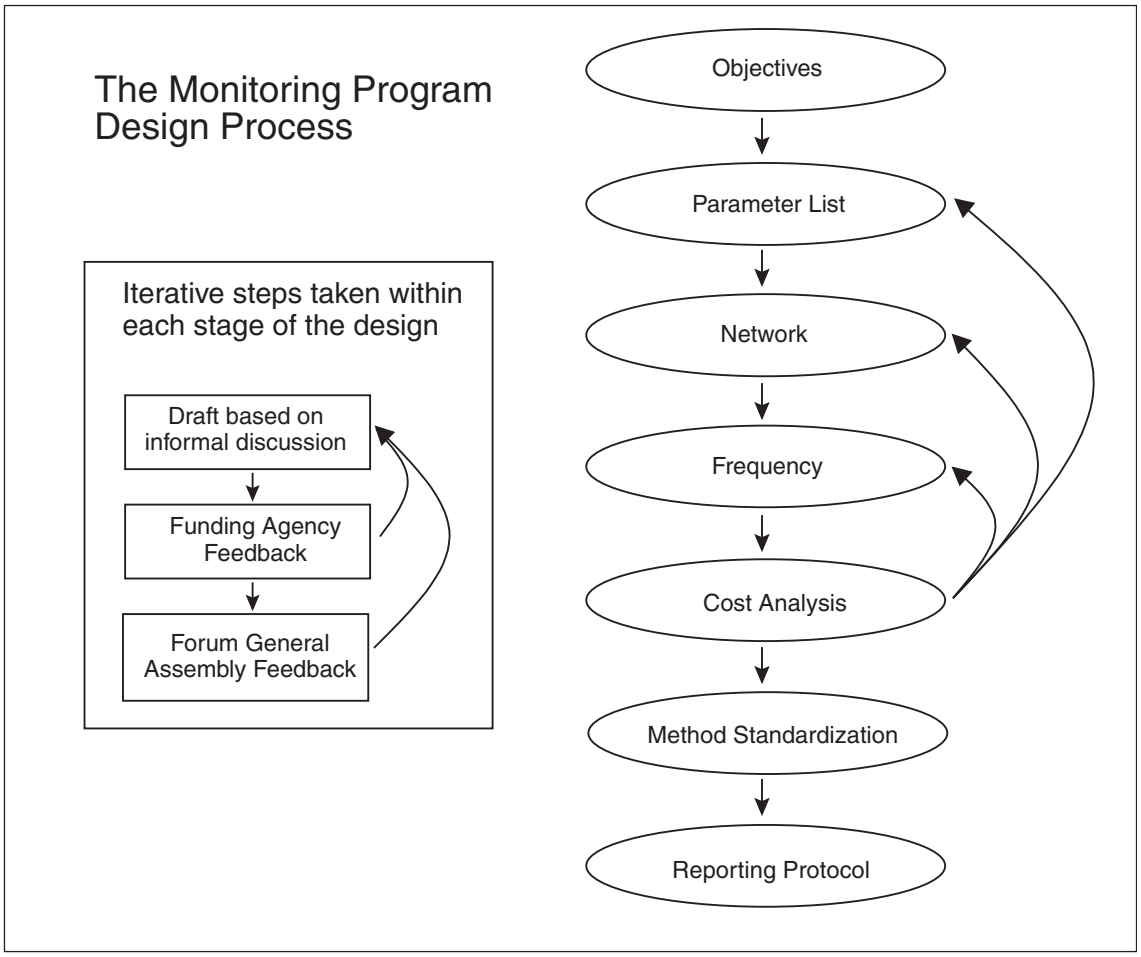

Figure 5. The Big Thompson Watershed Forum monitoring program design process (Greve, 1999).

Quality Control for the Big Thompson Watershed Forum

The BTWF is implementing the monitoring program in steps. Monitoring began at 14 sites in August 2000. Six more sites were added in January 2001 (fig. 4). As a preliminary QC plan, the monitoring program began collecting one field blank and one field replicate during each sampling run (15 sampling runs a year). These samples formed the basis of a QC data set until the formal data-quality plan was completed as part of this report.

Types of Basic Quality-Control Samples to be Collected

A system of field blanks and field replicates is used in the Big Thompson watershed. Four volatile organic carbon compounds—benzene, toluene, ethybenzene, and xylene (BTEX)—are included on the BTWF parameter list. These compounds may be subject to matrix interference or degradation; therefore, field-spike samples also are included in the QC samples. Laboratory-matrix spikes also provide information on matrix interaction with a lower possibility of field contamination. An extra sample of water is submitted to the laboratory where it is spiked. If degradation is of low concern for these four compounds, the laboratory-matrix spikes will meet BTWF needs.

Types of Topical Quality-Control Samples to be Collected

In order for the data reported by the three laboratories outside the USGS to be included in the National Water Information System (NWIS) database, the laboratories are required to participate in the BQS certification program. A part of this program is a system of Standard Reference Samples (SRS). The program requires two samples a year. The BTWF

will analyze SRS samples quarterly for the first 2 years of the monitoring program at all four laboratories: USGS National Water Quality Laboratory, City of Fort Collins Laboratory, the City of Loveland, and Acculabs. This will provide a baseline description of laboratory performance for data users. Participating in the certification program is an important step because data from some of these laboratories are not widely used, and potential data users may not be familiar with them.

At the beginning of each sampling year, an equipment blank is collected from each set of equipment. Currently in the BTWF program, there is one set of equipment. If equipment is used consistently and field-blank samples continue to show no contamination, collection of this equipment blank can be dropped.

Inference Space

The BTWF/USGS surface-water sample collection is conducted by a single sampling team with uniform methods. All analysis for a given constituent is done by a single laboratory. The potential sources of error are not expected to differ either spatially or temporally. Therefore, the entire moving-water (streams, tunnels, and canals) monitoring network is considered part of the same inference space. If at a later time, due to changes in water chemistry, staff, methods, or environmental conditions, the assumption of a single inference is not valid, inference spaces will be redefined and the QC sample design modified accordingly.

The BTWF intends to add a reservoir-monitoring program, independent of the cooperative USGS surface-water program, as well as a volunteer monitoring program. These two monitoring efforts likely will involve different sampling personnel and methods and should be placed in separate inference spaces until the sampling programs can be shown to be comparable.

The Number of Quality-Control Samples to be Collected

The BTWF plans to publish an annual report describing the water quality in the basin and spatial trends and a detailed report on temporal trends every 5 years. Therefore, the first analysis of data will occur 1 year after monitoring began, which means that enough QC samples must be collected to meet the informational objectives of the BTWF during the first year.

- **Field-blank samples.** During the first year, 12 field-blank samples will be collected. Subsequent years will have eight field-blank samples. This setup allows the 83d percentile of potential contamination to be estimated with 90-percent confidence after the first year (see table 2, fig. 1). By the second year, the 89th percentile, with 90-percent confidence, will be estimated. After 5 years, the 90th percentile, with 90-percent confidence, will be estimated. In addition, there will be enough QC samples such that the 90th percentile will not be represented by the highest detected concentration. This means that the presence of an extreme outlier would not strongly affect the estimate of bias.
- **Field-replicate samples.** The number of field replicates collected will match the number of blanks: 12 during the first year and 8 during the subsequent years. The 12 field replicates allow the variability to be estimated within 126 percent of the sample standard deviation with 80-percent confidence after the first year of sample collection (table 4, fig. 2). After 5 years, variability can be estimated within 122 percent of the sample standard deviation with 95-percent confidence.
- **Field-spike or laboratory-matrix spike samples.** During the initial first few months of operation, the BTEX analysis resulted in censored data at all sampling sites. Spike data provide evidence of degradation or matrix interaction. Spike information will allow the censored data to be assessed to determine if the concentrations are low due to low environmental concentrations or if they are low due to the ability of analysis methods to detect the compounds. From a subset of sites representing different parts of the basin and different time periods, nine spike sets will be collected.

The number of QC samples should be evaluated each year. If the magnitude of the estimates of bias and variability are small in relation to the environmental concentrations, the informational needs of the BTWF may be able to be met with fewer QC samples.

Distribution of Quality-Control Samples

Based on the previous discussion of sample size, it is evident that some amount of targeting of samples will take place. Specifically, during the first year of the program, selected sampling sites are targeted in order to identify problems early and allow detailed data analysis following the first year of data

collection. At the sites located in the uppermost parts of the watershed, targeting also will take place on a seasonal or hydrological basis. Replicate samples are most effective when the environmental samples have detectable concentrations. Therefore, the replicate samples should be collected when it is most likely that environmental samples have detectable concentrations. Because there are 20 sites, all sites will have at least one replicate and one blank sample by the completion of the second year. In this situation, in several areas of the watershed, sampling sites are located close to one another. One of the two sampling sites will be left for QC sampling during the following year. In this manner, the first year of QC sampling will include complete spatial coverage of the basin. In addition, the timing of the sampling at each site also will vary. For example, during the summer, QC samples can be collected both in the upper and lower portions of the basin. By the completion of the fifth year of sampling, all sites could have at least two field-replicate and two field-blank samples. These two samples will represent different seasonal or hydrologic conditions.

SUMMARY AND CONCLUSIONS

It is becoming more common for community-based stakeholder groups to implement water-quality monitoring programs. The data generated by these programs are often underutilized due to uncertainty in the quality of the data produced. The process of designing and implementing a data-quality program is time consuming; however, it allows the quality of data to be documented and defended. This process adds credibility to the data and allows it to be much more widely used.

Data-quality measures can be broken into three steps: quality assurance, quality control, and quality assessment. The quality-assurance step attempts to control sources of error that cannot be directly quantified. This step is part of the design phase of a monitoring program and includes clearly defined, quantifiable objectives, sampling sites that meet the objectives, standardized protocols for sample collection, and standardized laboratory methods. It is this step that is often most challenging to stakeholder groups due to the involvement of multiple entities, each with different priorities, monitoring goals, or sampling and analysis protocols. The quality-control (QC) step is the collection of samples to assess the magnitude error in a data set due to sample collection, processing, transport, and analysis. In order to design a system of QC samples, a series of issues needs to be considered: (1) potential sources of error, (2) the type of QC samples, (3) inference space, (4) the number of QC samples, and (5) the distribution of the QC samples within an inference space. Quality assessment is the process of evaluating quality-assurance measures and analyzing the QC data. Quality assessment has two parts: one that is conducted on an ongoing basis as the monitoring program is running, and one that is conducted during the analysis of environmental data. The ongoing quality-assessment measures include a series of checks as data are returned from the laboratories. The analysis of QC data provides an estimate of the magnitude of bias and variability. These estimates can be used in the interpretation of environmental data.

The design of a data-quality program is done in conjunction with the monitoring design. Some of the assumptions about the monitoring network, the quality of water in the basin of interest, and the sampling methods may be incorrect or may change over time. The program can be adjusted to adapt to the changes and ensure the monitoring or data-quality objectives are met.

REFERENCES

Buzan, Dave, Thomas, Alex, Halstead, Pam, and Hawks, Laurie, 1996, Getting started—Four perspectives: The Volunteer Monitor, v. 8, no. 1.

Fegeas, R.G., Claire, R.W., Guptil, S.C., Anderson, K.E., and Hallam, C.A., 1983, U.S. Geological Survey digital cartographic data standards—Land use and land cover digital data: U.S. Geological Survey Circular 895–E, 21 p.

Goeldner, Jo, 1996, Organizational development—Weathering the storms: The Volunteer Monitor, v. 8, no. 1.

Greve, A.I., 1999, Cooperative design of a water quality monitoring system for the Big Thompson River watershed, Colorado: Fort Collins, Colorado State University, Master's thesis, 129 p.

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

Hitt, K.J., 1995, Refining 1970's land-use data with 1990 population data to indicate new residential development: U.S. Geological Survey Water-Resources Investigations Report 94–4250, 15 p.

Kenney, D.S., McAllister, S.T., Caile, W.H., and Peckham, J.S., 2000, The new watershed source book—A directory and review of watershed initiatives in the Western United States: Boulder, University of Colorado Natural Resources Law Center, 461 p.

Mueller, D.K., 1998, Quality of nutrient data from streams and ground water sampled during 1993–95—National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 98–276, 25 p.

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.

National Water Quality Laboratory, 1996, Calculation of matrix spike recovery: accessed August 2, 2001, on the World Wide Web at URL http://wwwnwql.cr.usgs.gov/USGS/SpikeCalc.html

Natural Resources Law Center, 1996, The watershed source book: Boulder, University of Colorado Natural Resources Law Center.

Ott, L.R., 1993, An introduction to statistical methods and data analysis: Belmont, Calif., Duxbury Press, 1051 p.

River Network, 2001, National directory of river and watershed conservation groups: accessed August 2, 2001, on the World Wide Web at URL: http://www.rivernetwork.org/library/libnetdir.cfm

U.S. Environmental Protection Agency, 1996, The volunteer monitor's guide to quality assurance project plans: U.S. Environmental Protection Agency Report 841–B–96–003, Office of Wetlands, Oceans, and Watersheds, 59 p.

———1997, Top 10 watershed lessons learned: Washington, D.C., U.S. Environmental Protection Agency Report 840–F–97–001 Office of Water, Wetlands, Oceans, and Watersheds, 59 p.

U.S. Geological Survey, 1997 to present, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9, 2 v., variously paged. [Also available on-line at http://water.usgs.gov/owq/FieldManual/. Chapters originally were published from 1997–99; updates and revisions are ongoing and are summarized at: http://water.usgs.gov/owq/FieldManual/mastererrata.html]