United States
Environmental Protection
Agency

Office of Acid Deposition,
Environmental Monitoring and
Quality Assurance
Washington DC 20460

EPA/600/4-88/018
April 1988

Research and Development

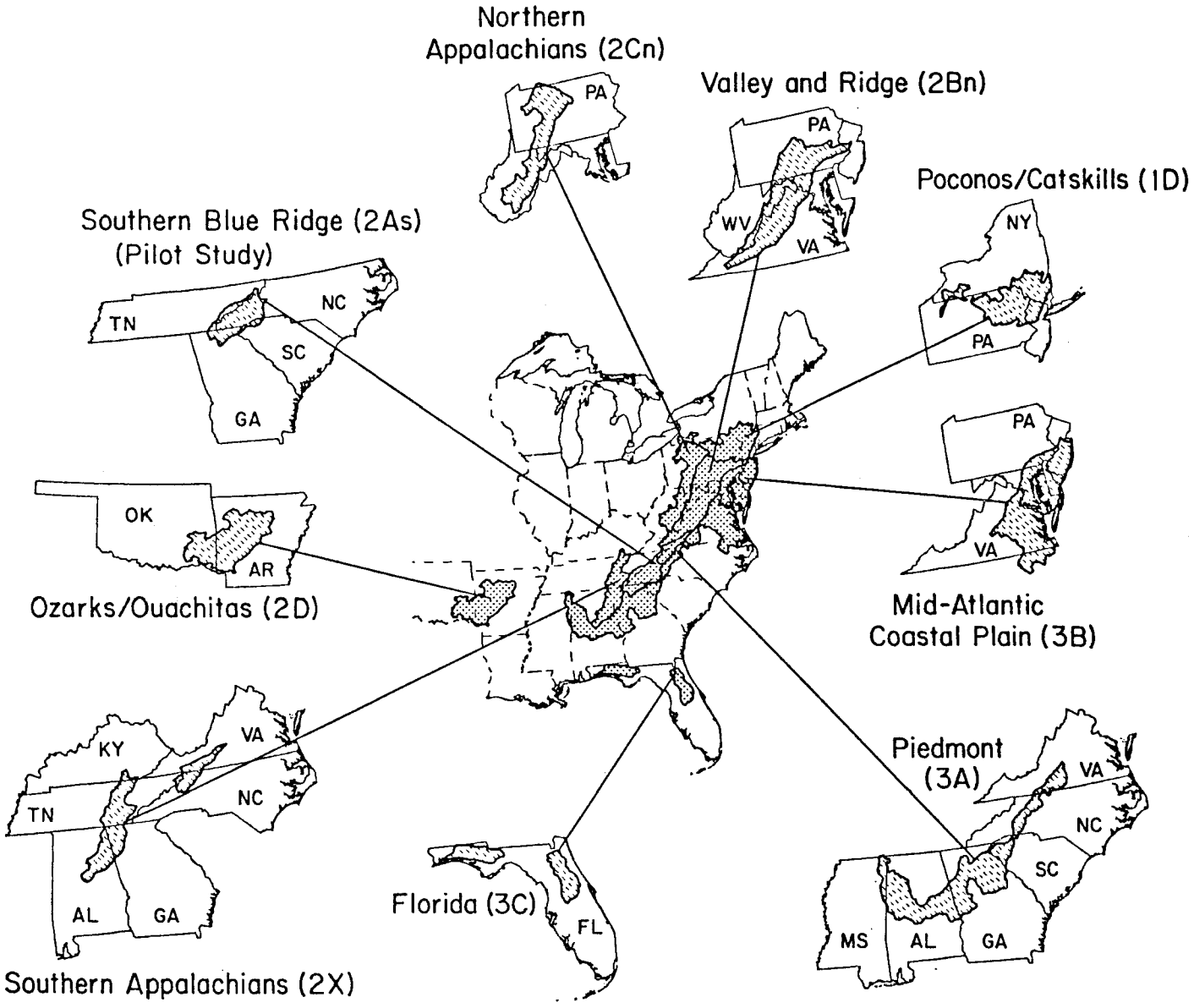
EPA

National Stream Survey - Phase I

Quality Assurance Report

Aquatic Effects Research Program

SUBREGIONS OF THE NATIONAL STREAM SURVEY - PHASE I



EPA/600/4-88/018
April 1988

National Stream Survey

Phase I

Quality Assurance Report

A Contribution to the
National Acid Precipitation Assessment Program

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

Environmental Monitoring Systems Laboratory – Las Vegas, NV 89119
Environmental Research Laboratory – Corvallis, OR 97333

NOTICE

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under contract number 68-03-3249 to Lockheed Engineering and Management Services Company, Inc., and contract number 68-03-3246 to Northrop Services, Inc. It has been subject to the Agency's peer and administrative review, and it has been approved for publication as an EPA document.

The mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This document is one volume of a set which fully describes the National Stream Survey. The complete document set includes the major data report, pilot survey data report, quality assurance plan, analytical methods manual, field operations report, processing laboratory operations report, and quality assurance report. Similar sets are being produced for each Aquatic Effects Research Program component project. Colored covers, artwork, and use of the project name in the document title serve to identify each companion document set.

Proper citation of this document is:

Cougan, K. A., D. W. Sutton, D. V. Peck, V. J. Miller, and J. E. Pollard. 1988. National Stream Survey - Phase I: Quality Assurance Report. EPA/600/4-88/018. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

ABSTRACT

The National Stream Survey - Phase I, conducted during the spring of 1986, was designed to assess quantitatively the present chemical status of streams in regions of the eastern United States where aquatic resources are potentially at risk as a result of acidic deposition. A quality assurance program was implemented to ensure consistency in the collection and analysis of water samples and to verify the reported results. In addition, the quality assurance program provides data users with quantitative and qualitative documentation of the quality of the data base in terms of representativeness, completeness, and comparability and the quality of the analytical results in terms of detectability, accuracy, and precision. This quality assurance report describes the major design and operational aspects of the quality assurance program and the final assessment of the quality of the National Stream Survey data base. This report also describes sampling and analytical problems that occurred during the survey and the corrective actions implemented.

The survey data base is sufficiently representative and complete so that population estimates based on chemical characteristics can be computed and interpreted. The results of the survey can be compared to the results of the Phase I Pilot Survey, to other data bases of the National Surface Water Survey, and to other existing or future water quality data bases with similar design, methodology, reporting units, and quality assurance.

There are only a few cases in which data interpretation may be limited by data quality in terms of detectability, accuracy, and precision. In most of these cases, the limitations affect only interpretation of measurements at low concentrations. A model-based approach to evaluating systematic errors is presented as an appendix to this report. Suggestions for future surveys include performing on-site inspections of all operations earlier in the survey so that most potential problems can be identified before they affect data quality and modifying the procedures for preparation of synthetic audit samples to facilitate improved estimates of accuracy.

This report is submitted in partial fulfillment of contract number 68-03-3249 by Lockheed Engineering and Management Services Company, Inc., under the sponsorship of the U.S. Environmental Protection Agency. This report covers a field work period from February 1986 through May 1986; data verification was completed in March 1987 and data evaluation was completed in March 1988.

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

FIGURES

TABLES

TABLES (continued)

TABLES (continued)

Abbreviations and Acronyms

Acroyms

AERP	Aquatic Effects Research Program
AQUARIUS	Aquatics Quality Assurance Review, Interactive Users' System
AQUARIUS II	modification of the AQUARIUS system developed for the National Stream Survey - Phase I
CLP	Contract Laboratory Program
DQO	data quality objective
ELS-I	Eastern Lake Survey - Phase I
EMSL-LV	U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada
EPA	U.S. Environmental Protection Agency
ERL-C	U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, Oregon
IFB	invitation for bid
IDL	instrument detection limit
NAPAP	National Acid Precipitation Assessment Program
NSS-I	National Stream Survey - Phase I
NSWS	National Surface Water Survey
NTU	nephelometric turbidity unit
ORNL	Oak Ridge National Laboratory
PCU	platinum cobalt unit
QA	quality assurance
QC	quality control
QCCS	quality control check sample
%RSD	percent relative standard deviation
SAS	Statistical Analysis System
SDL	system decision limit
SOW	statement of work

Variables and Units

Al-ext	aluminum, total extractable
Al-mono	aluminum, total monomeric
Al-nex	aluminum, nonexchangeable monomeric
Al-total	aluminum, total
ANC	acid-neutralizing capacity
BNC	base-neutralizing capacity
Ca	calcium
Cl^-	chloride
CO_3^{2-}	carbonate

Variables and Units (continued)

Cond-in situ	specific conductance measured in the field
Cond-lab	specific conductance measured in the analytical laboratory
Cond-PL	specific conductance measured in the processing laboratory
DIC-closed	dissolved inorganic carbon, closed
DIC-eq	dissolved inorganic carbon, equilibrated
DIC-open	dissolved inorganic carbon, open system
DO	dissolved oxygen
DOC	dissolved organic carbon
F^-	fluoride, total dissolved
Fe	iron
H^+	hydrogen ion
HCO_3^-	bicarbonate
K	potassium
Mg	magnesium
mg/L	milligrams per liter
Mn	manganese
Na	sodium
NH_4^+	ammonium
NO_3^-	nitrate
OH^-	hydroxyl
P	phosphorus, total dissolved
pH	negative logarithm of the hydrogen-ion concentration
pH-ANC	pH, initial (acid titration for ANC)
pH-BNC	pH, initial (acid titration for BNC)
pH-closed	pH, closed system
pH-eq	pH, equilibrated
pH-field	pH, measured in the field
SiO_2	silica
SO_4^{2-}	sulfate
µeq/L	microequivalents per liter
µS/cm	microsiemens per centimeter

ACKNOWLEDGMENTS

G. M. Aubertin and M. P. Aubertin (Southern Illinois University), J. K. Taylor (Quality Assurance Consultant), and K. A. Tonnessen (California Air Resources Board) provided peer reviews of this report. Their thoroughness in reviewing the report under a tight time constraint is gratefully acknowledged. Other individuals who reviewed the report include M. M. Mitch and A. Herlihy (Utah State University), M. J. Sale and J. M. Coe (Oak Ridge National Laboratory), and K. N. Eshleman (Northrop Services, Inc., Corvallis, Oregon).

A number of employees of Lockheed Engineering and Management Services Company, Inc., Las Vegas, Nevada, assisted in preparing this report. The authors greatly appreciate the writing contributions of D. J. Chaloud, J. L. Engels, R. Hoenicke, and C. M. Monaco. J. E. Teberg provided statistical support for the data analyses. The authors also acknowledge the following individuals for their assistance and constructive input: J. K. Bartz, R. Corse, S. K. Drousé, R. E. Enwall, M. L. Faber, F. C. Garner, C. A. Hagley, M. L. Hoppus, D. C. Hillman, J. D. Hunter, J. I. Lau, L. K. Marks, M. J. Miah, T. E. Mitchell-Hall, S. L. Pierett, P. F. Showers, M. E. Silverstein, L. A. Stanley, M. A. Stapanian, and A. D. Tansey.

Individuals associated with other organizations also provided assistance and advice: S. Paulsen (University of Nevada, Las Vegas); S. D. Edland, T. J. Permutt, and T. S. Stocking (Systems Applications, Inc., San Rafael, California); E. Canelli (New York State Department of Health, Albany, New York); and L. S. Shepard and R. D. McCracken (Global Geochemistry Corporation, Canoga Park, California).

Finally, guidance in developing, implementing, and administering the quality assurance program for the National Stream Survey - Phase I was provided by R. A. Linthurst, Program Director, and D. H. McKenzie, Acting Program Director, Aquatic Effects Research Program (AERP); J. J. Messer and P. Kaufmann (Utah State University), Technical Directors for the National Stream Survey; and R. D. Schonbrod and D. T. Heggem (U.S. Environmental Protection Agency, Las Vegas, Nevada), who served as Technical Monitors for the National Stream Survey - Phase I quality assurance program.

Section 1

Introduction

The National Stream Survey - Phase I (NSS-I) was designed to determine the present chemical status of streams in regions of the eastern United States where aquatic resources are potentially at risk as a result of acidic deposition. This report describes the quality assurance (QA) program employed during the NSS-I. The QA program was designed to ensure consistency in the collection and analysis of samples, to verify the reported results, and to inform data users of the quality and potential limitations of the resultant data base. This document evaluates the QA program itself as well as the quality of the analytical data base.

Section 2 presents conclusions about NSS-I data quality and recommendations regarding the QA program. Section 3 describes the design of the QA program and Section 4 describes the QA operations. Section 5 discusses the results of the operational aspects of the QA program and Section 6 assesses NSS-I data quality.

Background

The National Stream Survey is one of a series of surveys conducted as part of the National Acid Precipitation Assessment Program (NAPAP). This program is an interagency research, monitoring, and assessment effort initiated to address a growing concern about the possible effects of acidic deposition on the natural resources of the United States and neighboring countries. Congress established the NAPAP as part of the Acid Precipitation Act of 1980 to provide policymakers with technical information concerning the extent and the severity of the effects of acidic deposition.

The NAPAP is composed of seven task groups. Task Group VI oversees the Aquatic Effects Research Program (AERP), which is administrated by the U.S. Environmental Protection Agency (EPA) through its Office of Acid Deposition, Environmental Monitoring, and Quality Assurance. One objective of the AERP is to identify subpopulations of surface waters and the associated biota at risk from acidic deposition. The AERP consists of five large-scale projects that address chronic (long-term) and acute (short-term) exposure of aquatic systems to acidic deposition.

The National Surface Water Survey (NSWS), one of the AERP projects, consists of two components: the National Lake Survey and the National Stream Survey. Figure 1 shows the relationship of the regional surveys and monitoring projects that make up the NSWS. Each component of the NSWS began with a synoptic survey designed to characterize and quantify the chemistry of lakes and streams throughout the United States. The focus was on areas expected to contain the majority of low-alkalinity waters. The National Lake Survey was initiated with a pilot survey in 1983. Lake surveys took place in 1984 in the eastern United States (Linthurst et al., 1986) and in 1985 in the western United States (Landers et al., 1987). The National Stream Survey was initiated with a pilot survey in 1985 in the southern Appalachian region (Messer et al., 1986). The full-scale synoptic survey was conducted in the eastern United States in the spring of 1986.

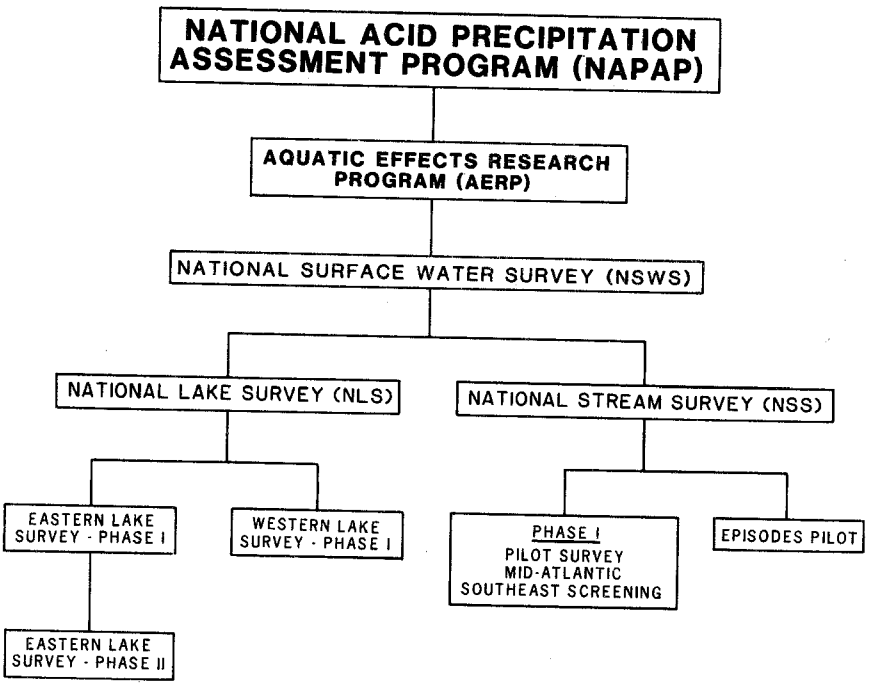


Figure 1. Organization of the National Surface Water Survey, showing the two major components, the National Lake Survey and the National Stream Survey.

All AERP surveys are designed to yield data bases of known quality through the standardized collection of data from regionally typical study sites. Each AERP project includes an extensive QA program. Such a program is required of every EPA-funded monitoring and measurement effort (Stanley and Verner, 1985).

National Stream Survey

The major goals of the NSS-I were to describe and classify streams in the eastern United States target population. Figure 2 shows the regions studied during the NSS-I. The NSS-I activities were initiated during a pilot study conducted in the Southern Blue Ridge province of the United States (Messer et al., 1986; Drousé, 1987). The purpose of the Phase I pilot survey was to evaluate the adequacy of the logistics plan, the statistical sampling design, and the methods proposed for the full-scale Phase I study as well as to finalize QA and quality control (QC) guidelines and data quality objectives. As a result of the pilot study, which was conducted from mid-March to mid-July of 1985, some changes (see Section 3) were made in the design and operations for the full-scale survey conducted in 1986.

The major collection efforts of the NSS-I were conducted in the mid-Atlantic region of the eastern United States, where survey personnel collected more than 1,000 samples from approximately 270 streams. This Mid-Atlantic Survey was designed to estimate the present degree of acidity of streams in areas that are characterized by low surface-water alkalinity, high rates of acidic deposition, and few lakes. In addition, the survey was designed to determine for future study which streams are representative of stream subpopulations. The Mid-Atlantic Survey

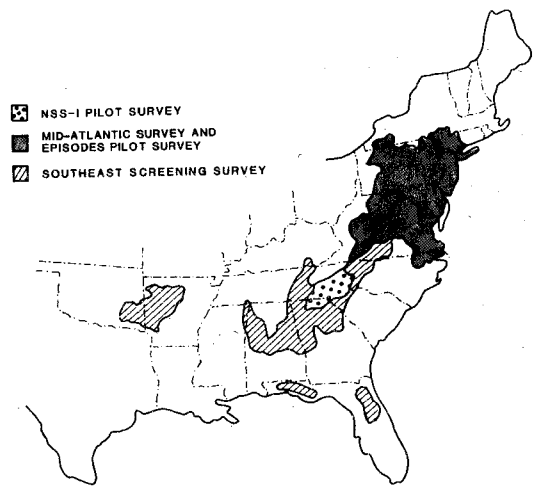


Figure 2. National Stream Survey study areas.

covered stream reaches in an area bounded by the Catskill and Pocono Mountains to the north, the North Carolina-Virginia boundary to the south, the approximate western boundaries of Pennsylvania and West Virginia to the west, and the Atlantic Ocean to the east. Each stream reach (segment of the stream network between two tributary confluences) was sampled twice during spring baseflow conditions, March 15 through May 15. Two sampling points on each of these reaches were located just above the downstream point of confluence and just below the upstream point of confluence.

A less intensive collection effort was conducted concurrently in the southeastern United States. The Southeast Screening Survey was designed to evaluate specific areas for intensive study in the future. The screening survey was conducted in parts of Virginia, North Carolina, South Carolina, Kentucky, Tennessee, Mississippi, Alabama, Georgia, Oklahoma, Arkansas, and Florida (Figure 2) that were identified by the National Lake Survey as having a large number of acidic lakes (Linthurst et al., 1986). One sample was collected from the upstream and downstream ends of 180 stream reaches.

A small-scale episodes pilot survey for a proposed study of episodic events in streams (related to weather conditions that produced snowmelt and rainfall) was conducted in conjunction with the mid-Atlantic field sampling effort. This survey was designed to test the feasibility of using a probability-based sampling design to assess the extent, magnitude, duration, and frequency of acidic episodes on a regional scale. This study also tested specific physical and chemical sampling protocols proposed for the full-scale Episodic Response Project, another NAPAP project.

Results of the episodes pilot study indicated that a synoptic approach to sampling streams during episodes would not be logistically feasible (Hagley et al., in press). Although collection of 30 sets of episode samples was anticipated, dry weather allowed collection of only 2 complete sets and 7 partial sets of samples. Based on the results of the episodes pilot survey, the Episodic Response Project will use a model-based approach to assess the regional importance of episodes to stream chemistry and biota (Eshleman, 1988). The results of the episodes pilot survey will not be discussed further in this report.

NSS-I sampling activities included locating stream sites and collecting water samples and associated data on the physical and chemical characteristics of the streams. After collection, the samples were sent to a processing laboratory where they were organized into sample batches, analyzed for selected chemical and physical variables, split into aliquots, preserved, packed, and shipped to the analytical laboratories. After the samples were analyzed, the analytical laboratories prepared a report on the analytical data produced. Copies of this report were distributed by overnight courier to the data management staff for entry of the information into the NSS-I data base and to the QA staff for verification of the reported results.

Survey Participants

A number of organizations were involved in various aspects of the NSS-I. The National

Stream Survey was funded and administered by the EPA Office of Acid Deposition, Environmental Monitoring, and Quality Assurance in Washington, D.C. The EPA Environmental Research Laboratory in Corvallis, Oregon, was responsible for coordinating the activities of the survey and for project design, site selection, data validation, and data interpretation. Utah State University and Northrop Services, Inc., provided technical services to the Corvallis laboratory. The EPA Environmental Monitoring Systems Laboratory in Las Vegas, Nevada, was responsible for QA and QC activities, sampling and logistical operations, communications coordination, and analytical support. The Las Vegas laboratory received assistance in these areas from Lockheed Engineering and Management Services Company, Inc. The U.S. Soil Conservation Service and other federal and state agencies helped to determine land ownership and to obtain access to field sites. Global Geochemistry Corporation (Canoga Park, California) and the New York State Department of Health (Albany, New York) provided analytical laboratory services. Radian Corporation (Austin, Texas), the support laboratory, provided performance audit samples. The Oak Ridge National Laboratory in Oak Ridge, Tennessee, was responsible for developing and managing the data base for the survey. Personnel at the Oak Ridge laboratory also participated in data interpretation and provided statistical programming, mapping, and other geographical analyses. Systems Applications, Inc. (San Rafael, California), provided support in analysis of analytical laboratory bias and audited the data base. The EPA Sample Management Office in Alexandria, Virginia, was responsible for sample tracking and assessment of analytical laboratory performance to determine financial compensation.

National Stream Survey Documents

This QA report is one of a number of publications that describe the NSS-I. The QA plan for the NSS-I is documented in Drousé et al. (1986a). Messer et al. (1986) describe findings of the NSS-I pilot survey conducted in 1985. Drousé (1987) summarizes the QA data results for the NSS-I pilot survey. Field and processing laboratory operations during the NSS-I are described by Hagley et al. (in press) and Arent et al. (in preparation). The analysis protocols employed in the processing and analytical laboratories are contained in Hillman et al. (1987). The summary report for the NSS-I (Kaufmann et al., in press) describes the design of the survey and presents the major findings.

Section 2

Conclusions and Recommendations

This section presents conclusions and recommendations drawn from the description and evaluation of the NSS-I quality assurance program found in Sections 3 through 6 of this report. For an explanation of the premises on which the following statements are made, as well as additional detail, the reader should refer to the appropriate section.

Conclusions

The success of the quality assurance program depends on how well the data generated by the survey met the data quality objectives. Overall, the program was able to assure that the quality of the NSS-I data was known and acceptable and that the data quality issues were documented. The following general conclusions can be drawn:

- The representativeness, completeness, and comparability of the data are adequate for project objectives.

- In a few cases, data interpretation may be limited by considerations of data quality in terms of precision, accuracy, and detectability. Table 31 defines the status of overall results for each analysis in terms of each data quality objective (DQO). Table 13 lists the DQOs.

- Checks of the internal consistency of results for each sample generally indicate excellent agreement, although some unmeasured ions or noncarbonate protolytes are apparently present in some of the streams sampled.

The following subsections list specific conclusions regarding the three primary DQOs.

Detectability

- For most variables, instrumental and methodological performance and background levels of analyte did not produce any serious problems with data quality.

- Method-level limits of detection for all measurements met the DQOs, except for phosphorus and silica measurements from one analytical laboratory and specific conductance measurements from the processing laboratory for the first half of the survey.

- No DQOs were established for system-level detectability. However, comparison of system-level limits of detection to the method-level DQOs showed that results for acid-neutralizing capacity (ANC), specific conductance, magnesium, potassium, sodium, and sulfate met or nearly met the method-level DQO; results for extractable aluminum, calcium, fluoride, and manganese were less than twice the method-level DQO; and results for total aluminum, equilibrated and initial dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), and nitrate exceeded twice the method-level DQO.

- Samples for which background levels of nonexchangeable monomeric aluminum exceed total monomeric aluminum have an increased level of uncertainty associated

with their values for exchangeable monomeric aluminum, especially at low concentrations.

Accuracy

- For all variables except base-neutralizing capacity (BNC) at low concentrations (less than 30 μeq/L), specific conductance in dilute samples (less than 25 μS/cm), and total dissolved phosphorus during the latter half of the survey, accuracy estimates for measurements at Laboratory 1 were within the DQOs. Measurements of several variables at Laboratory 2 showed potential systematic error; of these, only DOC at low concentrations, silica, and BNC exhibited a degree of potential error that might affect data interpretation.

- The synthetic audit samples provided a reliable means to assess the accuracy of most measurements. However, the potential for loss of aluminum and iron as well as the dependency of ANC, BNC, DIC, and pH values on dissolved carbon dioxide concentration and on the correct addition of other analytes to the formulation could affect the composition of the synthetic audit sample. Therefore, in some cases, accuracy estimates based on synthetic audit samples may not reflect the true quality of the data for those analytes.

- During the last half of the survey, phosphorus data from Laboratory 1 may be affected by a low-level negative calibration bias.

Precision

- Random errors occurring during sample preparation and analysis contribute only a small proportion to the overall measurement error.

- For all variables except total aluminum, DOC, iron, and ammonium, variability among batches contributes more to overall measurement error than does sample collection; therefore, among-batch precision estimates should be used to evaluate measurement uncertainty.

- For total aluminum, DOC, iron, and ammonium, sample-to-sample variability or collection effects are more important than day-to-day or among-laboratory variability in determining overall measurement error; therefore, system-level precision estimates should be used to evaluate measurement uncertainty.

Recommendations

Field and Processing Laboratory Activities

- For the specific conductance measurement made at the processing laboratory, use a water bath to maintain sample temperature at 25 °C rather than calculating the temperature-corrected measurement.

- Test all instrumentation and protocols used for analytical measurements before the survey begins to be certain they will perform as anticipated.

- Minimize contamination of non-acid-washed apparatus by shielding it from the acid washing of filtration equipment.

- Provide longer and more comprehensive training programs for all processing laboratory personnel and include an emphasis on proper completion of data forms.

- Designate an assistant to the base coordinator for the purpose of reviewing and correcting all field data forms before shipment.

- Develop efficient filtration procedures for samples containing large quantities of particulate material for future studies of stream chemistry.

Analytical Laboratory Activities

- Conduct on-site evaluations at the analytical laboratories early in the course of analyses, as directed by the QA plan. Make follow-up evaluations if possible.

- Make confirmation and reanalysis of questionable analytical values within a specific time frame a contractual requirement.

- Specify clearly the required number of decimal places to which analytical results must be reported throughout the survey.

- Use identical software at each laboratory to calculate ANC and BNC in future surveys.

Data Evaluation

- Consider using the closed-system measurements of DIC and pH in the analysis of the NSS-I data because they provide a better estimate of in situ conditions at the time of sampling than do open-system measurements.

- Use analytical measurements of specific conductance because measurements made at the processing laboratory during the first half of the NSS-I may be subject to systematic error.

Design of Quality Assurance Programs

- Perform preliminary statistical evaluation of pooled raw data early enough to identify problems and to allow reanalyses within holding time limitations.

- Raise the detection limit objective for equilibrated and initial DIC measurements if the detection limit is to be estimated from blank samples that are exposed to the atmosphere.

- Clearly define the approach for determining the detection limits and specify the approach in the QA plan. Low-level QCCS or audit samples may be of more use in assessing laboratory performance in terms of detectability than laboratory blank samples. Delineate approaches to assessing data quality in the QA plan.

- Whenever more than one laboratory is involved in analyses and interlaboratory bias is a concern, consider more stringent within-laboratory control limits and use audit samples (or collected samples) representing a wide range of concentrations to monitor, assess, and possibly correct for any biases that occur. This approach will also allow for a more rigorous assessment of accuracy within a laboratory.

- Allow the synthetic audit samples to equilibrate for a period of time before use; subject both synthetic and natural audit samples to rigorous verification measurements against certified standards so their composition is known with a high level of certainty.

- Consider preparing synthetic samples on an analyte-by-analyte basis or as aliquots that include chemically compatible variables.

- Select audit sample compositions that bracket the expected concentrations of analytes in the stream samples.

- Provide the analytical laboratories with known performance standards from a single source so that all laboratories can calibrate their measurement systems to a given target value to reduce interlaboratory bias.

- Consider using a series of split samples, prepared from a composite bulk routine sample, rather than duplicate samples to assess precision and identify components of error more discretely.

- Consider taking chemically well-characterized natural audit samples into

the field and processing them through the sampling device to allow estimates of the total uncertainty due to sampling and measurement.

- If reliable BNC data is required to differentiate weak and strong acid concentrations in natural water samples, modify the analytical methodology so that titration is conducted under an inert atmosphere free of carbon dioxide.

Section 3

Design of the Quality Assurance Program for the National Stream Survey - Phase I

An important design criterion of the NSS-I was that the data collected must be scientifically sound and of known quality. To meet these requirements, standardized collection of data was implemented and a rigorous QA program was established. This program has two separate but integrated components that cover operations and data management. The operations component included QA and QC procedures to ensure that all samples were collected and analyzed consistently and to estimate the accuracy and precision of the reported values with a known degree of confidence. The data management component established a program that stored and tracked the data; identified and corrected entry, reporting, and analytical errors; and kept a record of such changes. These procedures produced documented files that contained data of known quality and that are accessible to project scientists and extramural users. The NSS-I QA plan (Drousé et al., 1986a) defines the activities needed to meet the requirements of the QA program and to guide the operations and data management components. The plan also presents QA protocols for collecting, processing, shipping, and analyzing samples as well as for reporting and verifying analytical results.

Stream Characteristics and Data Quality Objectives

One of the first steps in the design of the NSS-I was to identify the variables to be measured and to define the analytical data quality objectives (DQOs) for measuring each variable. Twenty-seven chemical and physical characteristics of stream water were selected for in-situ or laboratory measurement. Table 1 lists these characteristics along with abbreviations used in this report and analytical methods. These variables were selected because measurements of their concentration in stream waters should provide sufficient information to determine the chemical and physical quality of the streams with respect to fish habitat and the geochemical nature of the waters with respect to past and future susceptibility to acidic deposition. Some variables are of primary interest with respect to these survey objectives (e.g., pH and acid-neutralizing capacity). Other variables are important in interpreting the primary variable data (e.g., dissolved organic carbon (DOC), color, and fluoride are useful in understanding the speciation of aluminum). Variables such as nitrate, sulfate, and DOC are needed to describe the ionic composition of waters, and some may be useful indicators of nonatmospheric pollution (e.g., chloride, total dissolved phosphorus, and ammonium). Finally, some variables may provide clues to the geochemical processes controlling water chemistry in a region and may also be useful in classification of stream reaches for further study (e.g., silica, sodium, potassium, and calcium). Complete chemical analysis for all major ions is needed to conduct verification checks on the accuracy of chemical analyses on the basis of cation/anion balances and specific conductance checks. Messer et al. (1986) and Hillman et al. (1987) give brief descriptions of each variable.

Table 1 lists the instrument or method used in the field and in the laboratory to measure each variable. Some variables (dissolved inorganic carbon, pH, and specific conductance) were measured more than once in each sample, either with different methods or at different locations (field and laboratory),

Table 1. Chemical and Physical Variables Measured During the National Stream Survey - Phase I

Variable (units)	Abbreviation[a]	Instrument or analytical method[b]
	FIELD SITE	
pH, field (pH units)	pH-field	Portable pH meter (Beckman pHI-21); glass combination electrode (Orion-Ross Model 8104)
Specific conductance (μS/cm)	cond-in situ	Portable conductivity meter (YSI Model 33 S-C-T) with probe (YSI Model 3310)
Dissolved oxygen (mg/L)	DO	Portable dissolved oxygen meter (YSI Model 54A); pressure-compensating oxygen-temperature probe (YSI 5739)
Temperature	--	Portable conductivity meter (YSI Model S-C-T) with probe (YSI Model 3310)
	PROCESSING LABORATORY	
Aluminum (mg/L)		
Total monomeric	Al-mono	Colorimetry (complexation with pyrocatechol violet, automated flow injection analyzer), La Chat Quick Chem System IV Colorimeter
Nonexchangeable monomeric	Al-nex	Same as total monomeric
Specific conductance (μS/cm)	Cond-PL	Conductivity meter (YSI Model 32); probes (YSI Model 3417 and Model 3401); NBS thermometer
pH, closed system (pH units)	pH-closed	pH meter (Orion-Ross Model 611), and glass combination electrode (Orion-Ross Model 8104)
Dissolved inorganic carbon, closed system (mg/L)	DIC-closed	Infrared spectrophotometry (carbon analyzer) (Dohrmann Model DC-80)
True color (PCU)	--	Comparator (Hach Model CO-1)
Turbidity (NTU)	--	Nephelometer (Monitek Model 21)
	ANALYTICAL LABORATORY	
Acid-neutralizing capacity (μeq/L)	ANC	Acidimetric titration, modified Gran analysis
Aluminum, (mg/L)		
Total extractable	Al-ext	Atomic absorption spectroscopy (furnace) on methyl isobutyl ketone extract
Total	Al-total	Atomic absorption spectroscopy (furnace)

(Continued)

Table 1. (Continued)

Variable (units)	Abbreviation[a]	Instrument or analytical method[b]
	ANALYTICAL LABORATORY (Continued)	
Ammonium (mg/L)	NH_4^+	Colorimetry (phenate, automated)
Base-neutralizing capacity (μeq/L)	BNC	Alkalimetric titration, modified Gran analysis
Calcium (mg/L)	Ca	Atomic absorption spectroscopy (flame)
Chloride (mg/L)	Cl^-	Ion chromatography
Specific conductance (μS/cm)	Cond-lab	Conductivity cell and meter
Dissolved inorganic carbon (mg/L)		
Open system	DIC-open	Infrared spectrophotometry
Equilibrated	DIC-eq	Infrared spectrophotometry
Dissolved organic carbon (mg/L)	DOC	Infrared spectrophotometry
Fluoride, total dissolved (mg/L)	F^-	Ion-specific electrode
Iron (mg/L)	Fe	Atomic absorption spectroscopy (flame)
Magnesium (mg/L)	Mg	Atomic absorption spectroscopy (flame)
Manganese (mg/L)	Mn	Atomic absorption spectroscopy (flame)
Nitrate (mg/L)	NO_3^-	Ion chromotography
pH (pH units)		
Equilibrated	pH-eq	pH electrode and meter; sample equilibrated with 300 ppm CO_2 in air
Initial (acid titration for ANC)	pH-ANC	pH electrode and meter
Initial (base titration for BNC)	pH-BNC	pH electrode and meter
Phosphorus, total dissolved (mg/L)	P	Automated colorimetry (phosphomolybdate or modification)

(Continued)

Table 1. (Continued)

Variable (units)	Abbreviation[a]	Instrument or analytical method[b]
Potassium (mg/L)	K	Atomic absorption spectroscopy (flame)
Silica (mg/L)	SiO_2	Automated colorimetry (molybdate blue)
Sodium (mg/L)	Na	Atomic absorption spectroscopy (flame)
Sulfate (mg/L)	SO_4^{2-}	Ion chromatography

[a] These abbreviations for variables will be used throughout this report.

[b] Methods and instruments are described in Hagley et al. (in press) for field activities and in Hillman et al. (1987) for processing and analytical laboratory analyses. The analytical laboratories met the instrument requirements as defined in the Statement of Work.

for a total of 35 measurements. The analytical laboratories made 24 of these measurements. Three operationally defined aluminum fractions were measured: total monomeric aluminum, nonexchangeable monomeric aluminum, and extractable aluminum. Nonexchangeable monomeric aluminum was determined after passing the sample through a cation exchange column (Hillman et al., 1987). Closed-system measurements of pH and DIC were made on samples collected in sealed syringes without exposure to atmospheric carbon dioxide. Equilibriated measurements of pH and DIC were conducted after sparging the samples with 300 ppm carbon dioxide in air mixture (Hillman et al., 1987).

Other stream characteristics that were measured or estimated at sampling sites included watershed disturbances, land use, bank vegetative cover, stream substrate, and stream width, depth, and flow velocity. The site information recorded at each sampling location was intended to assist in the initial interpretation of physical ahd chemical data from each site and to aid in locating the site in future studies. This site information was not subjected to the full scope of QA activities and, although it is recorded in the data base, it should not be used to draw quantitative inferences about the other chemical or physical data. These data are not discussed further in this report.

Researchers involved in any monitoring or measurement study funded by the EPA must establish DQOs based on the proposed end use of the data. These objectives are set before the research begins. The expected range of sample concentrations and the objectives for detection limits, precision, and accuracy were developed for each parameter by using data from the published literature, from statistical error simulation, and from the results of Phase I of the Eastern and Western Lake Surveys. Equipment, sampling protocols, and analytical methodologies were selected and were standardized in order to achieve the DQOs. These objectives were also applied to the statistical assessment of sampling, processing laboratory, and analytical laboratory performance. The objectives set criteria for detectability, accuracy, precision, representativeness, completeness, and comparability.

Measures of detectability, accuracy, and precision are estimated by analyzing data from QA and QC samples. Detectability is the ability of an instrument or method to determine a measured value for an analyte above background levels with a specified degree of confidence. Accuracy describes the closeness of a measured value to the true (or index) value of the variable concentration in the sample. Precision describes the closeness of values derived by repeated measurements of the same quantity under specified conditions. The

values and ranges that were established for these three DQOs are given in Section 6, Table 10. For most of the 35 variables measured, the analytical results were evaluated to determine if they met these analytical DQOs. In addition to these six DQOs, relative interlaboratory bias, operationally defined as a systematic difference in analytical performance between laboratories, is evaluated in this report.

The requirements for survey data to be representative, complete, and comparable were addressed by the NSS-I statistical sampling design (Kaufmann et al., in press) and the QA plan (Drousé et al., 1986a). Completeness is a measure of data quality that is the quantity of acceptable data actually collected relative to the total quantity that was expected to be collected. Comparability expresses the confidence with which one data set can be compared to another. Representativeness is a measure of the degree to which sample data accurately and precisely reflect the characteristics of a population. Representativeness also relates to the degree to which QA and QC samples represent routine stream samples.

As the survey progressed, measurements of the QA and QC samples that were made at the stream sites, processing laboratory, and analytical laboratories were compared to the DQOs and concentration ranges. These comparisons provided a mechanism to identify and correct sampling, analytical, and reporting errors before data quality was affected.

Statistical Design of the National Stream Survey

To characterize stream chemistry and associated physiographic attributes accurately and confidently, a statistically based scheme was developed to ensure that the stream reaches sampled would be representative of the target population (i.e., those streams of interest with respect to the primary objectives of the Aquatic Effects Research Program and the National Acid Precipitation Assessment Program). The detailed rationale behind stream selection and sampling during Phase I is described by Blick et al. (1987), Overton (1985, 1987), and Kaufmann et al. (in press).

Sample Collection and Analyses--Quality Assurance and Quality Control

For the NSS-I, a routine sample was collected from the stream in a 3.8-L container and four syringes. In addition, QA and QC samples, described in the QA plan (Drousé et al., 1986a) and in Table 2, were employed in the field, in the processing laboratory, and at the analytical laboratories to maintain the quality of the survey data and to ensure that data quality could be characterized. Stringent requirements for instrument calibration also helped to provide reliable measurements.

Figure 3 shows the relationship of the different QA and QC samples to the collection and analysis process. The results from analyses of QA samples were used to evaluate the performance of sampling methods, laboratory analyses, and overall data quality for the survey. Analyses of the QC samples allowed field samplers and laboratory personnel (in both the processing and analytical laboratories) to identify and correct specific problems such as poor instrument performance or reagent contamination before and during routine sample analyses. Although it was not a requirement of the NSS-I, each laboratory followed its own internal good laboratory practices and measured QC samples that were independent of the survey QA samples.

Quality Assurance Samples

Of the 1,654 NSS-I samples analyzed at the analytical laboratories, 273 (16.5 percent) were QA samples (Table 3). These field and processing laboratory blank, field duplicate, and field and laboratory audit samples (Table 2) were added to a group of routine stream samples either at the stream site or at the processing laboratory. They were analyzed at the processing laboratory (except for laboratory audit samples) and the analytical laboratories. Because analytical laboratory

Table 2. Quality Assurance and Quality Control Samples Used in the National Stream Survey - Phase I

Sample type	Description	Function	Frequency of use[a]
QUALITY ASSURANCE			
Field blank	Reagent-grade deionized water[b] subjected to sample collection, processing, and analysis	To assess detectability and identify possible sample contamination resulting from collection and processing	One per batch
Processing laboratory blank	Reagent-grade deionized water[b] subjected to sample processing and analysis	To estimate background effects due to sample processing and analysis	In lieu of field blank when logistical constraints prevented its collection
Field duplicate	Duplicate sample collected immediately after the routine stream sample	To estimate system precision	One per batch
Performance audit Field	Synthetic or natural lake sample; prepared at support laboratory and processed at processing laboratory	To estimate analytical precision of processing and analytical laboratory measurements; to estimate relative accuracy and relative interlaboratory bias	As scheduled
Laboratory	Synthetic or natural lake sample; prepared and processed at support laboratory	To estimate analytical precision of analytical laboratory measurements; to estimate relative accuracy and relative interlaboratory bias	As scheduled
QUALITY CONTROL			
Calibration blank	Reagent-grade deionized water[b]	To identify signal drift	One per batch for applicable variables
Reagent blank	Reagent-grade deionized water[b] plus reagents for total aluminum and silica analyses	To identify contamination due to reagents	One per batch for total aluminum and silica

(continued)

Table 2. (Continued)

Sample type	Description	Function	Frequency of use[a]
Quality Control (continued)			
Quality control check sample (QCCS)	Standard solution from source other than calibration standard	To determine accuracy and consistency of instrument calibration; to check statistical control of measurement process	Before the first measurement, after the last, and at specified intervals in between for each batch
Detection limit QCCS	Standard solution at 2 to 3 times the required detection limit	To determine precision and accuracy at lower end of linear dynamic range of measurement method; to verify instrument detection limits	One per batch for applicable variables
Processing laboratory duplicate	Split of stream sample	To monitor analytical precision of processing laboratory measurements	One per batch
Analytical laboratory duplicate	Split of sample aliquot	To monitor analytical precision of analytical laboratory measurements	One per batch

[a] Planned frequency for use of QA samples was not always possible due to logistical constraints.
[b] ASTM (1984).

personnel did not know the origin, identity, or chemical composition of the samples, the QA samples were analyzed as if they were routine stream samples. These samples were used to evaluate the overall per- formance of sampling and analytical activities and to estimate data quality. Figure 4 gives a graphic presentation of the numbers and kinds of samples collected during the NSS-I.

Blank Samples--

Field blank samples were prepared at the processing laboratory from deionized water that met American Society for Testing and Materials specifications for Type I reagent-grade water (ASTM, 1984). Sampling crews transported the deionized water to the stream sites and processed the blank sample through sampling equipment as if it were a routine stream sample. Because closed-system DIC and pH analyses were not performed on field blank samples in the processing laboratory, only two syringes of stream water were collected for these samples. These two syringes were used to prepare an aliquot for analysis of extractable aluminum and determination of total monomeric and nonexchangeable monomeric aluminum.

Field blanks were processed along with routine samples at the processing laboratory and were included in the sample batches that

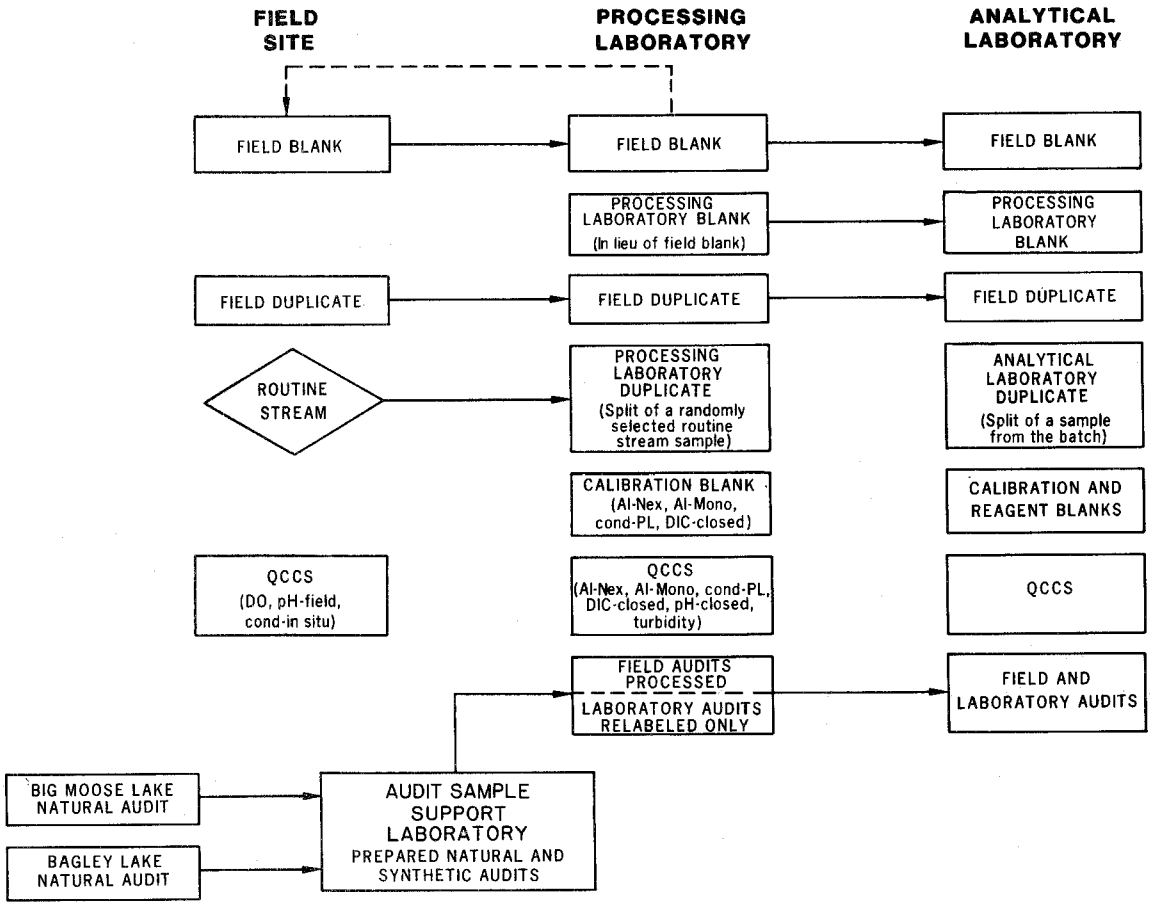


Figure 3. **The relationship of the quality assurance and quality control samples to the collection and analysis process.**

were sent to the analytical laboratories. Analytical data for these QA samples in each batch were used to identify possible contamination problems during sampling and analyses.

Occasionally, due to logistical constraints, field blanks were not available for processing at a stream site on a particular sampling day. In such instances (on five occasions), processing laboratory personnel substituted a deionized water sample for the missing field blank. Although this processing laboratory blank sample was not processed through the sampling equipment, it took the place of the missing field blank sample in the sample batch sent to the analytical laboratory. These processing laboratory blanks were used only to detect contamination and were not used in statistical QA analyses because they did not go through the entire sampling and analysis system from the field through the analytical laboratory.

Field Duplicate Samples--

A field duplicate is a second set of stream water samples collected immediately after the routine sample. The sampling crew used the same procedure to collect both the routine and duplicate samples. Pairs of field routine and duplicate samples were used to assess the precision of the field sampling techniques and the processing and analytical laboratory procedures.

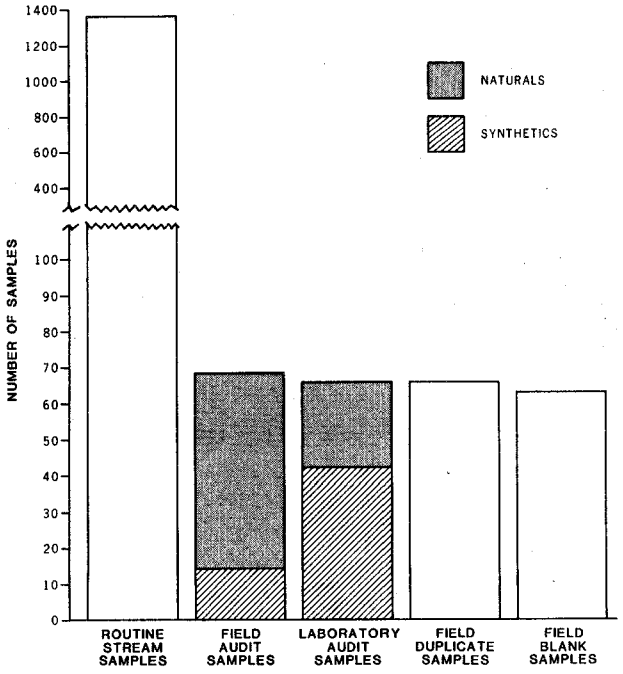


Figure 4. Routine and quality assurance sample collected during the NSS-I.

Performance Audit Samples--

The NSS-I used two types of performance audit samples--natural and low-concentration synthetic. A support laboratory provided and measured both types of audit samples and shipped them to the processing laboratory as needed. Audit samples were added to each batch of samples shipped to the analytical laboratories. These audit samples were added to the batches of routine samples in two ways. Field audit samples were handled and processed at the processing laboratory in the same manner as routine stream samples. Laboratory audit samples were prepared and immediately processed at the support laboratory. The analytical laboratory personnel could not recognize these samples as audit samples and did not know the sample compositions. Field and laboratory audit samples were used to (1) estimate systematic differences in measured values within and between analytical laboratories, (2) indicate the precision of those measurements through repeated analyses of the same sample type in several batches, and (3) provide data to determine if the handling of samples at the processing laboratory affected the analyte concentrations as measured at the analytical laboratories.

Field audit samples were not actually taken to the field and, therefore, were not processed through the sampling equipment in the field. The term "field" was used to identify these audit samples because the processing laboratories were located in the field during Phase I of both the Eastern and Western Lake Surveys rather than at the single Las Vegas location as for the NSS-I. The processing laboratory personnel recognized the field audits as audit samples but did not know the sample compositions. Analytical laboratory personnel neither recognized these samples as field audits nor knew the sample compositions.

Both natural and low-concentration synthetic field audit samples were used during NSS-I to evaluate the overall performance of the processing and analytical laboratories. These natural and synthetic samples represented various concentrations that were expected to be present in stream waters. The field audit samples helped to identify problems that could affect data quality and that could occur during sample processing, shipment, or analysis. When used in conjunction with laboratory audit samples, the analytical results for field audit samples provided data that were used to distinguish processing laboratory problems from analytical laboratory problems.

The field natural audit samples were collected by nonsurvey personnel from two lakes for which chemical data were available. The field natural sample from Big Moose Lake in the Adirondack Mountains of New York represented lakes sensitive to acidic deposition (i.e., waters that have very low buffering capacity or low ANC). The natural sample from Bagley Lake in the North Cascade Mountains of Washington, represented a system that has a higher ANC. Following collection, each natural audit sample was filtered in bulk and was divided into 2-L bottles at the support laboratory. The bottles were

Table 3. Number of Routine and Quality Assurance Samples Collected amd Analyzed During the National Stream Survey - Phase I

Sample type	Number of samples	Percent of total samples collected
Quality assurance samples		
Field blank	63	
Processing laboratory blank	5	
Field duplicate	66[a]	
Laboratory synthetic audit	42	
Field synthetic audit	14	
Laboratory natural audit	24	
Field natural audit	54	
Special studies	5	
Total quality assurance samples	273	16.5
Routine samples		
Mid-Atlantic	1,017	
Screening	343	
Episodes	21	
Total routine samples	1,381	83.5
Total samples collected	1,654	100.0

[a] Includes one sample that was not processed correctly and cannot be used to estimate data quality.

stored at 4 °C to minimize changes in chemical composition.

Field synthetic audit samples, which were prepared at the support laboratory to simulate natural water, included a matrix of analytes at specified theoretical concentrations. The synthetic sample represented surface water with low concentrations of analytes. Because the first lot of synthetic material was exhausted before the end of the survey, the support laboratory prepared a second lot with the same theoretical concentration. Field synthetic audit samples were prepared as concentrates and diluted just before they were sent in 2-L bottles to the processing laboratory. The chemical composition and preparation of the synthetic audit samples is described in Appendix A.

Data obtained from analyses of laboratory audit samples identified problems encountered during the analytical process that may affect data quality. In addition to their use in determining relative interlaboratory bias and the precision of measurements of the same sample type, laboratory audit samples helped to verify the accuracy of analytical procedures. Natural and synthetic laboratory audit samples came from the same sources as did the field audit samples. The support laboratory supplied audit samples already split into seven aliquots to the processing laboratory. The laboratory audit samples were labeled at the processing laboratory in the same manner as routine samples and were indistinguishable from any field sample. However, they were not processed or analyzed at the processing laboratory. They were included in a batch with routine stream samples that were processed and shipped on the same day to an analytical laboratory.

Quality Control Samples

The QC samples (Table 2) were used in the field and at the processing and analytical laboratories. In general, QC samples are used to ensure proper instrument performance and sample analysis. QC samples are defined as control samples for which the analyst knows the true analyte concentration or value. Analytical data for these samples must fall within control limits specified in the QA plan (Drousé et al., 1986a).

Field Quality Control Samples--

Quality control check samples (QCCSs) were used by the field crews to check the calibration of the pH, conductivity, and dissolved oxygen meters before sampling and to check for instrument drift during and after field measurements. Daily QC checks were made before and after sampling. If the

measurement of the QCCS did not fall within the control limits, the meter was recalibrated or checked for proper operation.

Processing Laboratory Quality Control Samples--

Processing laboratory personnel analyzed calibration blank samples, QCCSs, and processing laboratory duplicate QC samples. A calibration blank was analyzed before any samples in the batch to check for baseline drift and for contamination of the carbon analyzer, flow injection analyzer, and conductivity meters. Calibration and drift of the carbon analyzer and of the instruments used to measure pH, turbidity, specific conductance, and the aluminum species were also checked with QCCSs at specified intervals. The QA plan (Drousé et al., 1986a) required observed concentrations to be within the specified control limits. When an unacceptable value was obtained, the instrument was recalibrated and all samples that were analyzed after the last acceptable QC sample were reanalyzed. Each day one routine stream sample was selected randomly as the processing laboratory duplicate; this sample was split and analyzed in duplicate for pH, DIC, true color, turbidity, specific conductance, total monomeric aluminum, and nonexchangeable monomeric aluminum. Immediately after analyses, precision estimates were calculated from these analyses and compared to the DQOs for precision. If the calculated values did not meet the DQOs, then another duplicate sample was analyzed. If the calculated precision estimates from this analysis still did not meet the DQOs, the data were qualified with a tag (Appendix B).

Analytical Laboratory Quality Control Samples--

The analytical laboratories used five types of QC samples--calibration blanks, reagent blanks, detection limit QCCSs, low- and high-concentration QCCSs, and laboratory duplicates. For each analytical procedure, the calibration blank was analyzed after the initial instrument calibration to check for drift in the measurement signal. For silica and total aluminum measurements, the laboratory was required to analyze a reagent blank. The reagent blank, containing all the reagents in the same volumes that were used to prepare a real sample for analysis, was prepared in the same manner as a routine sample. The observed analyte concentration for calibration and reagent blanks could not exceed twice the required detection limit (Section 6, Table 13) for each analyte. If the concentration exceeded this limit, the source of the contamination had to be investigated and eliminated. If the source of contamination could not be identified before reanalysis, the data were qualified.

The QCCSs were either commercially prepared or laboratory-prepared samples that were made from stock solutions independent from those used to prepare calibration standards. The analyst was required to choose a QCCS for a particular variable such that its theoretical concentration fell in the mid-calibration range for that variable. The QCCS was analyzed to verify instrument calibration at the beginning of sample analysis, at specified intervals during sample analyses, and after the final sample in the batch was analyzed. The observed concentrations had to be within the specified control limits (Drousé et al., 1986a). When an unacceptable value for the QCCS was obtained, the instrument was recalibrated and all samples that were analyzed after the last acceptable QCCS were reanalyzed. In addition, the analytical laboratories were required to demonstrate statistical control by plotting the observed concentrations of the QCCS on a QC chart. To ensure continuity of QC charts, QCCSs of the same theoretical concentration were used throughout the plotting process. Both 99 percent and 95 percent confidence intervals were developed and used as control and warning limits, respectively. If the 99 percent control limit differed from the theoretical value by more than the limits given in the QA plan (Drousé et al., 1986a), the laboratories were required to consult the QA staff in Las Vegas regarding corrective action (i.e., sample reanalysis). On a weekly basis, QC charts were updated, cumulative means were calculated, and new warning and control limits (95 percent and 99

percent, respectively) were determined. In addition to QC charts developed with survey data, each laboratory prepared QC charts for the internal QC samples.

A detection limit QCCS is a low-level QC sample that contains the analyte of interest at a concentration of two to three times the required detection limit. A QCCS was analyzed once per batch before routine stream samples were analyzed for specified variables. These QC samples were used to verify the low end of the calibration curve and the values for the low-concentration samples near the detection limits. The concentration of the detection limit QCCS had to be between two and three times the required detection limit and the measured value had to be within 20 percent of the theoretical value. If it was not, the analyst was required to identify and correct the problem before sample analysis.

A duplicate analysis (laboratory duplicate) for each specified variable was performed on one sample in each batch to estimate and monitor analytical precision. If the observed precision did not meet the DQOs established for these variables, then another duplicate sample had to be analyzed (Drousé et al., 1986a). Data for which the precision estimate did not meet the DQOs were qualified with a flag (Appendix B).

Data Base Management

The NSWS data base management system incorporates the results from data collection, evaluation, verification, validation, and enhancement activities. This system assembles, stores, and edits data generated during the NSS-I and other NSWS surveys. The system also provides basic reports of the survey results, performs certain statistical analyses, and provides data security. A detailed description of the system is given in Sale (in press).

An important tool in the development of the NSS-I data base was the use of data qualifiers to mark an individual value or even an entire stream as having a particular feature that may be useful in data interpretation. Two types of data qualifiers, tags and flags, are used in the NSS-I data base (Appendix B). A tag is a code that was added to a value at the time of sample collection or analysis to qualify the value. A flag is a qualifier that was assigned during the verification and validation procedures to data that did not meet the established acceptance criteria or that were in some way unusual. These qualifiers alert future data users to values identified as questionable or unusual by the verification and validation process. These qualifiers also provide a method for identifying and removing clearly erroneous data and retaining questionable data with appropriate tags and flags.

The NSS-I data base was subjected to four levels of QA evaluation to ensure that the data collected during the survey are representative of the physical and chemical characteristics of the samples taken from the streams. Each level of quality assurance produced a new and more refined working data set. These working data sets are defined as: raw (Data Set 1), verified (Data Set 2), validated (Data Set 3), and enhanced (Data Set 4). Data Set 4 is the final product of the refinement process. All data sets are protected from unauthorized or accidental access by individual, system, and file password protection. The development of these working data sets is summarized in Figure 5. The data sets are further described in the following subsections.

Raw Data Set (Data Set 1)

The data from all components of the sampling and analysis process make up the raw data set. The raw data set includes all analytical results and data qualifiers. Appendix B lists the data qualifiers. The data forms used for reporting the raw data can be found in the QA plan (Drousé et al., 1986a). All field and processing laboratory forms on which data were recorded received a preliminary QA review at EMSL-LV before the data were reviewed and entered into the raw data set at ORNL. Data from the analytical laboratory forms were entered into the raw data set before the QA review, which took place during

data verification. To ensure accurate data transfer from field and laboratory reports, the information was entered into two computer files and subjected to automated checking procedures to minimize transcription errors. The raw data set was used to screen the data for problems, perform exploratory data analyses, and evaluate the need for any adjustments in the data analysis plans.

Verified Data Set (Data Set 2)

The objectives of the data verification process were to identify, correct, and flag raw data of questionable or unacceptable quality and to identify data that might need to be corrected during or after data validation. These objectives were met by reviewing the QA and QC data measured and recorded at the sampling site, at the processing laboratory, and at the analytical laboratories and by examining all sample data in terms of chemical charge balance. Verification determines the quality of the analytical data through a rigorous protocol based on known principles of chemistry. It scrutinizes the internal consistency of chemical concentrations as a result of cation/anion balances, conductance balance, or protolyte analysis for each sample.

Computer programs automated much of the verification process and generated reports for evaluating intra- and interlaboratory bias as well as discrepancies in blank, audit, and other QA and QC samples. The Automated Quality Assurance Review, Interactive Users System (AQUARIUS), which was used to process data during the NSS-I Pilot, was modified for use during the remaining stream surveys. The Aquatics Analysis System (AQUARIUS II) generates data changes in the form of transaction records. The records are derived from exception-generating programs that identify or flag analytical results that do not meet the expected QA or QC criteria.

The final product of the verification process is the verified data set in which each sample batch and each sample value has been reviewed individually and all questionable values are either corrected or identified with an appropriate flag. Data verification takes place in two parts: a preliminary evaluation which incorporates the majority of numeric changes and a final evaluation which includes any final numeric changes and the addition of data qualifier flags. The verified data set was used as the basis for data validation.

Validated Data Set (Data Set 3)

While verification procedures evaluated data at the sample and batch level, validation procedures examine the plausibility of sample data in the context of a subregional set of samples. NSS-I subregional boundaries generally group streams of similar geochemistry together. The validation process identified unusual data that would need special attention when used in statistical analysis, particularly in regional estimates concerning the target population of streams. Observations identified as atypical during review of data at subregional levels are considered outliers from the rest of the data. Two components of the validation process are the identification of statistical outliers from subregional distributions of chemistry and the evaluation of possible systematic errors in the measurement process. Such outliers may result from the natural variability of streams in the set of stream reaches, from anthropogenic disturbances in the natural environment, and from errors in the sampling design, as well as from sampling and analytical errors. Conditions that may cause outliers include:

1. Sample collection during an eposidic event for a given reach.

2. Factors other than normal geochemical processes (e.g., pollution or watershed disturbance, including acid mine drainage, brine, or other nonpoint sources).

3. Unusual geochemical properties within a given subregion.

4. Impossible datum, clearly erroneous when reviewing chemistry for that reach.

Although outliers may represent unusual data in comparison with other data, such

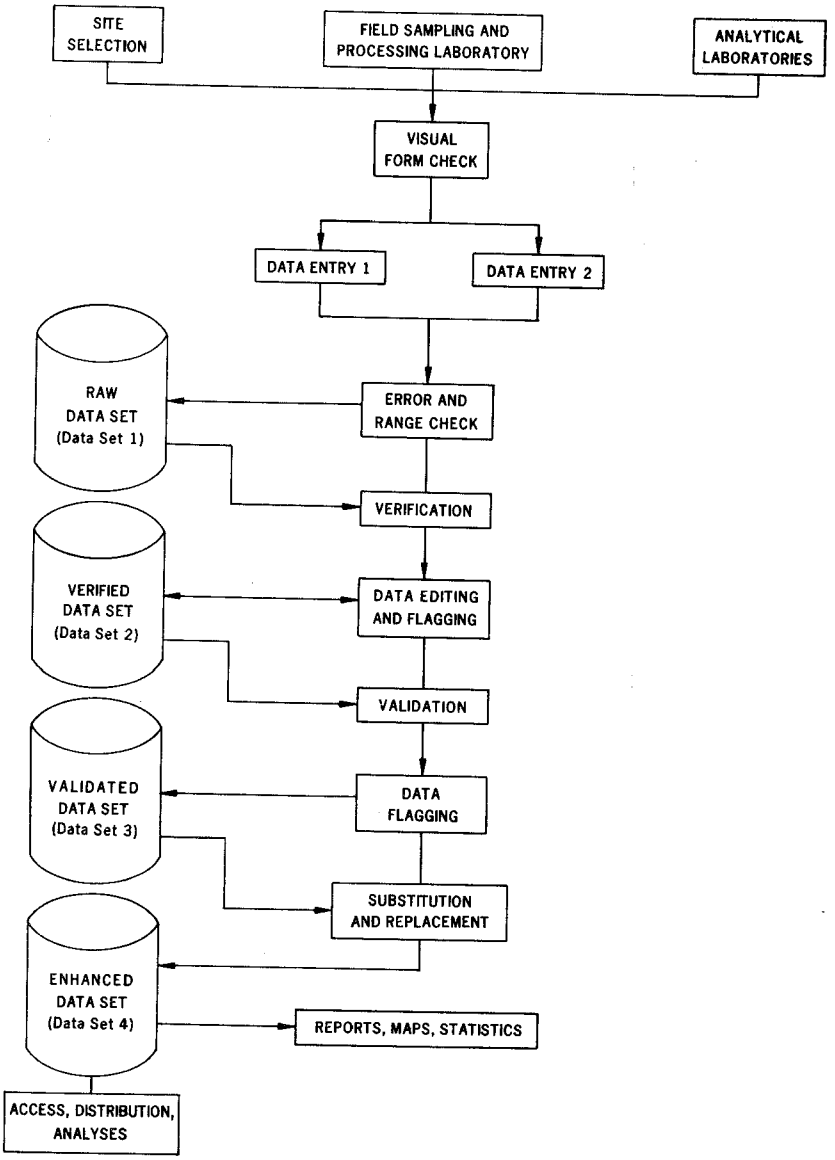


Figure 5. Data base management system.

values are not necessarily inaccurate in their representation of a stream reach. The validation process is, therefore, not meant to be a stringent pass or fail test, but rather a way to search for observations that may represent entry or analytical errors or unusual water chemistry. These unusual observations become apparent when the data are viewed as a set of information using univariate, bivariate, and multivariate analyses. All outliers identified during the validation procedures were investigated further to confirm that they were entered into the data base correctly. Any values determined to be erroneous were corrected in the validated data set. Values identified as unusual as a result of validation analyses were flagged in the validated data set. This data set retains the values for field blank, field duplicate, and performance audit samples. A detailed description of the validation process is given in the QA plan (Drousé et al., 1986a); validation is also described by Kaufmann et al. (in press).

Enhanced Data Set (Data Set 4)

Calculations of population estimates are difficult if values are missing from the data set. To avoid such problems, an enhanced data set was prepared by substituting erroneous or missing values according to specified criteria (Kaufmann et al., in press). Negative concentrations reported by the analytical laboratories were set equal to zero (except for ANC and BNC). An index value for each chemical variable for each sampling site was calculated by averaging the values of routine-duplicate pairs and the values from multiple observations for a sample site. The enhanced data set contains a single value for each variable for each sampling location (i.e., one observation for each upstream and downstream location) and therefore does not include values for QA samples or data qualifiers.

Differences Between the NSS-I and the NSS-I Pilot Survey

A number of changes were made in methods and procedures for the mid-Atlantic and southeast screening surveys as a result of the pilot survey. These changes are summarized in Table 4 and are described in the following subsections.

Processing Laboratory Sample Holding Times

Sample holding times for water samples (the period after sample collection and before aliquot preparation and sample analyses at the processing laboratory) were increased from 12 hours in the pilot survey to 30 hours in the mid-Atlantic and screening surveys. The decision to increase sample holding times was based on the results of two experiments: (1) a laboratory study testing whether or not carbon dioxide can permeate syringes over time (Burke and Hillman, 1987) and (2) a field study of bulk samples held in Cubitainers (Stapanian et al., 1987). The syringe experiments determined that holding times for DIC and pH held in syringes at 4 °C could be increased to 30 hours without a measurable effect on these variables. These experiments did not determine the effects of holding time for aluminum speciation. However, because pH changes that result from changes in dissolved carbon dioxide appear to be the most significant cause of changes in aluminum speciation, it was assumed that syringe aliquots can also be held for at least 30 hours before aluminum extraction. Bulk sample experiments also demonstrated that increasing holding times to as much as 30 hours would have no important effect on analyte concentration. These conclusions may only be applicable to low-ionic-strength natural streamwater samples such as those of the NSS-I and may not be universally applicable to other sample types (e.g., ground water, polluted waters, or industrial wastes).

Processing Laboratory Location

In the pilot survey, mobile processing laboratories were located in the field in order to meet the 12-hour holding time requirements for aliquot preparation, preservation, and preliminary analyses. As a result of the holding time experiments conducted for syringes and bulk samples during the pilot survey, sample processing was centralized in

Table 4. Differences Between the National Stream Survey - Phase I and the NSS-I Pilot Survey

Technique	Pilot	Phase I
Sample holding time	12 hours	30 hours
Processing laboratory location	Decentralized	Centralized
Field pH	Closed-system and open-system	Open-system
Methods of fractionation and determination of aluminum species	8-hydroxyquinoline method	8-hydroxyquinoline method and colorimetric method with pyrocatechol violet
Matrix spike quality assurance samples	Used	Not used
Phosphorus measurement	Total phosphorus (unfiltered)	Dissolved phosphorus (filtered)
Specific conductance in processing laboratory	Not measured	Measured

Las Vegas, resulting in better quality control as well as reduced costs.

Field pH Measurement

During the pilot survey, comparisons were made between two techniques for field pH measurements (Messer et al., 1986). The pH of samples collected in a syringe was measured in a closed system (in a custom-made sample chamber without exposure to the atmosphere) and the pH of samples collected in an open container (beaker) was measured in an open system. Both methods are described in the analytical methods manual (Hillman et al., 1987). When the data resulting from these two measurements were compared, no significant difference ($p = 0.05$) was found between the open-system measurement and the closed-system measurement. Thus, the logistically simple open-system measurement was chosen to determine field pH during the remainder of the NSS-I.

Fractionation and Determination of Aluminum Species

An experimental semiautomated colorimetric method for fractionation and determination of aluminum species by complexation with pyrocatechol violet (Hillman et al., 1987) was used during the NSS-I to measure total monomeric and nonexchangeable monomeric aluminum. This method was expected to be less expensive, less time consuming, and more reproducible than the 8-hydroxyquinoline

method used to measure total extractable aluminum during the pilot survey. Measurement of total monomeric aluminum using the pyrocatechol violet method is expected to yield data similar to data obtained by measurement of total extractable aluminum. The automated method should reduce variability due to different analysts and eliminate problems related to reproducibility and precise timing inherent in the manual 8-hydroxyquinoline method. However, because application of this method on a large scale was in the developmental stages, total extractable aluminum measurements using the 8-hydroxyquinoline method were continued throughout the NSS-I to permit comparison of the two methods.

Matrix Spike Samples

The purpose of matrix spike samples is to establish a matrix that is similar to the matrix of the samples collected and that can be used to verify the accuracy of an analysis. The analyst adds a known quantity of an analyte to a sample of known concentration and then analyzes the spiked sample. The percentage of spiked analyte recovered (percent recovery) determines whether or not there was a matrix effect on the analysis of the original sample. During the pilot survey, the limits for spike recovery were met for every batch and no matrix interferences were observed (Drousé, 1987). The matrix spike samples were not included in the 1986 NSS-I surveys because they did not provide any additional information about the quality of the data. Elimination of these samples also reduced costs.

Phosphorus Measurements

According to recent studies (e.g., Young et al., 1985), particulate-bound phosphorus tends to have a wide range of bioavailability, depending on its source. Measurement of total dissolved (filtered) phosphorus was selected for the NSS-I rather than the measurement of total (unfiltered) phosphorus made in the pilot and previous NSWS surveys because measurement of total dissolved phosphorus provides a better estimate of biologically active phosphorus.

Specific Conductance Measurements

During the pilot survey and the NSS-I, specific conductance was measured in the field and by the analytical laboratories. An additional conductance measurement was made in the processing laboratory during the NSS-I to provide another comparison for the data user. Comparison of these measurements was also used in data verification and validation.

THIS PAGE INTENTIONALLY LEFT BLANK

Section 4

Operations of the Quality Assurance Program

Quality assurance was an integral component of all aspects of the NSS-I including (1) selecting laboratories to analyze the samples; (2) providing QA-related information for training field and processing laboratory personnel; (3) collecting, processing, and shipping the samples; (4) analyzing the samples; (5) managing the data base; and (6) monitoring sample collection and analyses.

Selection of Analytical Laboratories

The Contract Laboratory Program (CLP), established to support the EPA hazardous waste monitoring activities, provided the mechanism for choosing the analytical laboratories. Under the CLP, an invitation for bid (IFB) is advertised. The IFB includes a statement of work (SOW) that defines analytical and QA and QC requirements in a contractual format. Each laboratory submitting a bid in response to the IFB is appraised on the basis of the analysis of performance evaluation samples and an on-site evaluation. The laboratory analyses had to be conducted according to handling, analytical, and QA protocols detailed in the SOW and published in the methods manual (Hillman et al., 1986) and in the QA plan (Drousé et al., 1986a).

The NSS-I analyses were performed under three separate SOWs. Laboratory 1 and Laboratory 2 previously had been awarded SOWs to analyze samples for the Eastern Lake Survey, Phase I (ELS-I), and the NSS-I Pilot, respectively. Analyses during the ELS-I and the NSS-I Pilot did not exhaust the bid lots (600 samples analyzed for each bid lot) that had been awarded to the laboratories, and it was decided to use up these bid lots during the NSS-I activities. The SOWs for the two surveys were basically identical: each required the laboratories to analyze up to 30 samples per day. The SOWs for the ELS-I and the NSS-I Pilot were modified for use in the NSS-I by eliminating analyses of matrix spike samples (see Section 3).

Because the remaining samples in the bid lots for those two laboratories were not sufficient to complete the NSS-I survey, a revised SOW was advertised before the survey began. Laboratory 2 passed the selection process and was awarded three additional bid lots to complete analyses of the NSS-I samples. This SOW became effective when the survey was two-thirds complete (with batch 2147). During the NSS activities, Laboratory 2 analyzed about 66 percent of the total number of samples and Laboratory 1 analyzed the remaining 34 percent. The revised SOW differed from the previous SOWs in the following ways:

1. The laboratory was required to analyze as many as 60 samples per day rather than the previous maximum of 30, thus eliminating the need to use two laboratories in the latter part of the survey.

2. Analyses of detection level QCCSs were required for chloride, sulfate, nitrate, ammonium, and silica in addition to the variables listed in the original SOWs as described in the ELS-I QA plan (Drousé et al., 1986b). This requirement provided an additional check on the low end of the linear dynamic range for these analytes.

3. The method for determining specific conductance was modified to include a step to equilibrate samples at 25 °C before analysis. This additional step uses a constant-temperature water bath. The initial SOW allowed the laboratory to correct the sample measurement to 25 °C after analysis. The modification minimized errors associated with the calculation.

4. The number of decimal places recommended for reporting each measurement was increased by one place for most variables. This change ensured consistent reporting of even low-level concentrations and it minimized rounding errors.

Training of Field, Processing Laboratory, and Quality Assurance Personnel

Training provided to the NSS-I field and processing laboratory personnel and QA staff members ensured consistency in sample collection, processing, and analysis and for QA and data verification.

Field personnel were introduced to the NSS-I project design, given safety training, and issued equipment at the Environmental Monitoring Systems Laboratory in Las Vegas, Nevada (EMSL-LV) (Hagley et al., in press). Training that covered NSS-I logistics and operations, instrumentation, stream sample collection and measurement techniques, QA and QC procedures, and proper data recording continued at the Oak Ridge National Laboratory (ORNL). Training was completed at the Nantahala Outdoor Center in Bryson City, North Carolina, where map reading, outdoor skills, and safety were emphasized and where a dry run to practice sample collection and stream measurement techniques was conducted.

Laboratory supervisors gave individual training to processing laboratory personnel for as long as 10 days in Las Vegas, Nevada (Arent et al., in preparation). This training covered all technical aspects of laboratory operations, including QA and safety procedures.

QA auditors received a week-long training session in Las Vegas, Nevada. Training covered all aspects of QA and QC as described in the QA plan (Drousé et al. 1986a). Auditors worked under close supervision of the QA supervisory staff throughout the verification process.

Field Sampling Operations

There were two separate sample collection operations for both the mid-Atlantic and the southeast screening regions. For each mid-Atlantic operation, five teams composed of two samplers each collected samples and associated field data. For each screening operation there were two teams of samplers. Each group of teams was supervised by a base coordinator who was assisted by a logistics coordinator. The field operations report (Hagley et al., in press) gives an in-depth description of logistics and procedures of sampling.

Each group of teams for an assigned sampling area operated from base sites selected on the basis of their proximity to sampling sites and the availability of required shipping and support services. Each base site occupied as many as 8 to 15 locations in a sampling area. The teams obtained access information for each stream reach before field operations began. Stream sites were reached by vehicle or by foot.

Each team sampled one or two reaches (at upstream and downstream sites) per day between mid-March and mid-May of 1986. The samplers calibrated the pH and dissolved oxygen field meters each morning at the base site. The pH, dissolved oxygen, and conductivity meters were checked with QCCSs before leaving the base site. Samplers also calibrated the dissolved oxygen meter at each stream site and checked the pH and conductivity meters with QCCSs before and after measurements were made. Activities of the field teams and measurement techniques are described in the field operations report

(Hagley et al., in press). Figure 6 shows the flow of field activities. At each sampling site, the samplers recorded watershed disturbances and substrate characteristics on standardized forms. They also made in-situ measurements of specific conductance, temperature, and dissolved oxygen and determined stream pH at streamside on an aliquot (beaker) of water collected by using Tygon tubing and a portable peristaltic pump. Hydrological data collected at downstream sites included stream width, depth, velocity, and discharge.

The sampling team collected a streamwater sample (routine sample) from each stream by pumping water through 1/4-inch Tygon tubing. The water samples were pumped from the midchannel of the stream into a 3.8-L polyethylene Cubitainer by a portable, battery-driven peristaltic pump. The samplers also filled four gastight 60-mL syringes for the analyses performed in the processing laboratory (i.e., pH, DIC, and total monomeric and nonexchangeable aluminum) and for the preparation of the extractable aluminum aliquot. More detailed discussions of these techniques are available in the field operations reports (Knapp et al., 1987; Hagley et al., in press).

Two types of QA samples were collected. Each day, one team at each of two base sites collected a field blank sample at the first site visited. The reagent-grade water for this field blank sample was carried from the processing laboratory to the sample site and pumped through all sampling equipment and into clean sample containers. In addition, using identical techniques, one team at each of the two remaining base sites collected a field duplicate sample, a second set of sample containers (Cubitainer and syringes) filled with stream water from the pump immediately after the routine sample was collected.

All sample containers were transported to the team vehicle in portable soft coolers that contained frozen-gel packs. Team members transferred the samples and associated data forms to insulated shipping containers with frozen-gel packs. The temperatures of the coolers were checked by inserting a thermometer whenever the samples were transferred from one container to another. The containers were shipped on the same day by overnight courier to ensure their arrival at the processing laboratory in Las Vegas, Nevada, on the morning after collection. Because it was necessary to meet overnight courier deadlines, only the stream data form (Form 4) was enclosed with the samples. Other field forms were sent to the EMSL-LV QA staff as soon as the forms were reviewed by the base coordinators. If the review identified any changes necessary on Stream Data Form 4, the coordinator notified the EMSL-LV QA staff by telephone and provided paper documentation with the next form shipment to Las Vegas. Titles of all field forms can be found in Figure 7 and the forms are reproduced in the QA plan (Drousé et al., 1986a).

Processing Laboratory Operations

The processing laboratory provided a controlled environment in which to process and preserve water samples and to measure variables that tend to become unstable over time. Processing laboratory personnel included a laboratory coordinator, laboratory supervisor, and as many as 20 analysts. The processing laboratory personnel:

1. randomly selected and organized the stream, blank, and audit samples into batches;
2. divided the samples into aliquots;
3. prepared the sample aliquots for subsequent analytical laboratory analysis;
4. prepared and shipped the sample batches to the analytical laboratories;
5. measured seven variables (pH, total monomeric aluminum, nonexchangeable monomeric aluminum, specific conductance, dissolved inorganic carbon, turbidity, and true color);

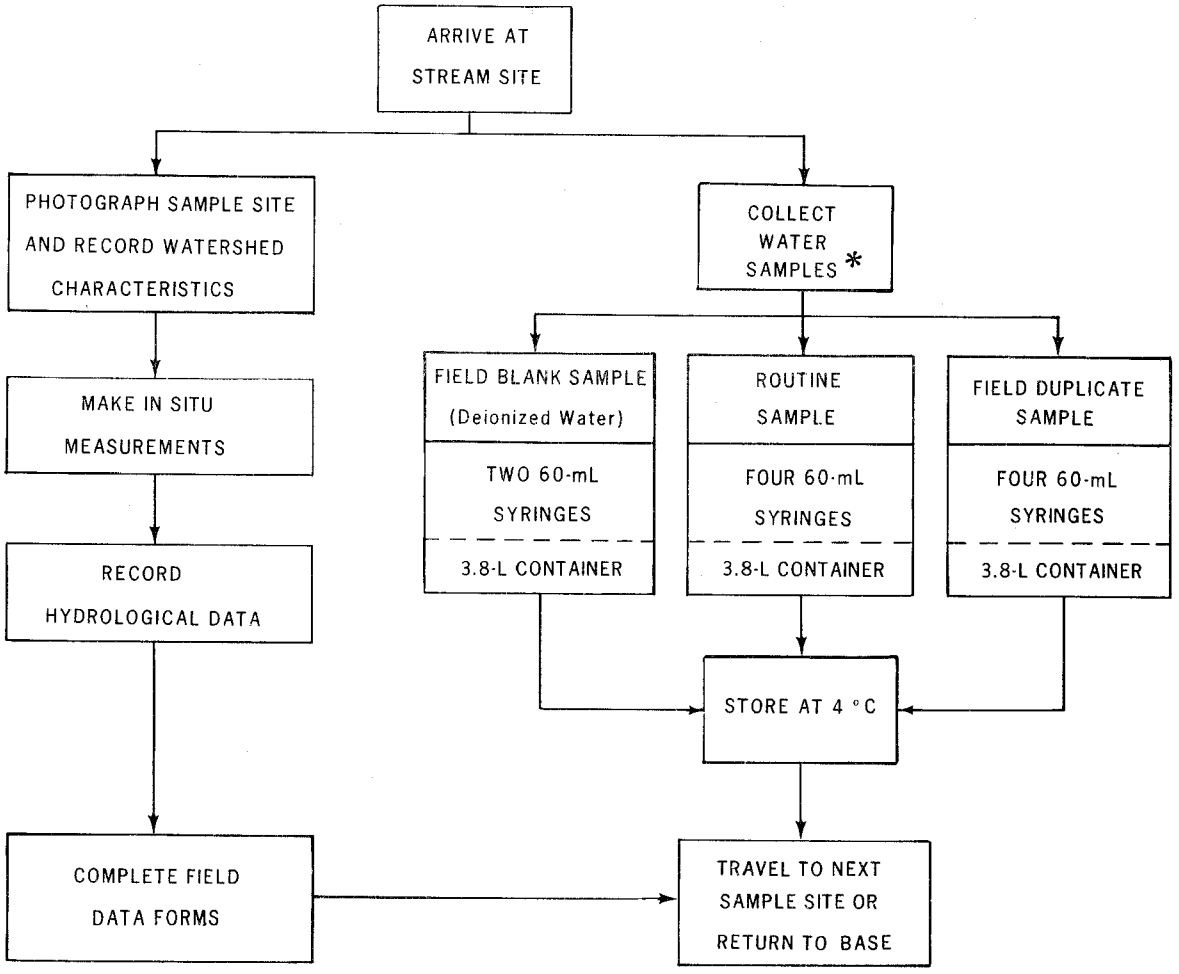
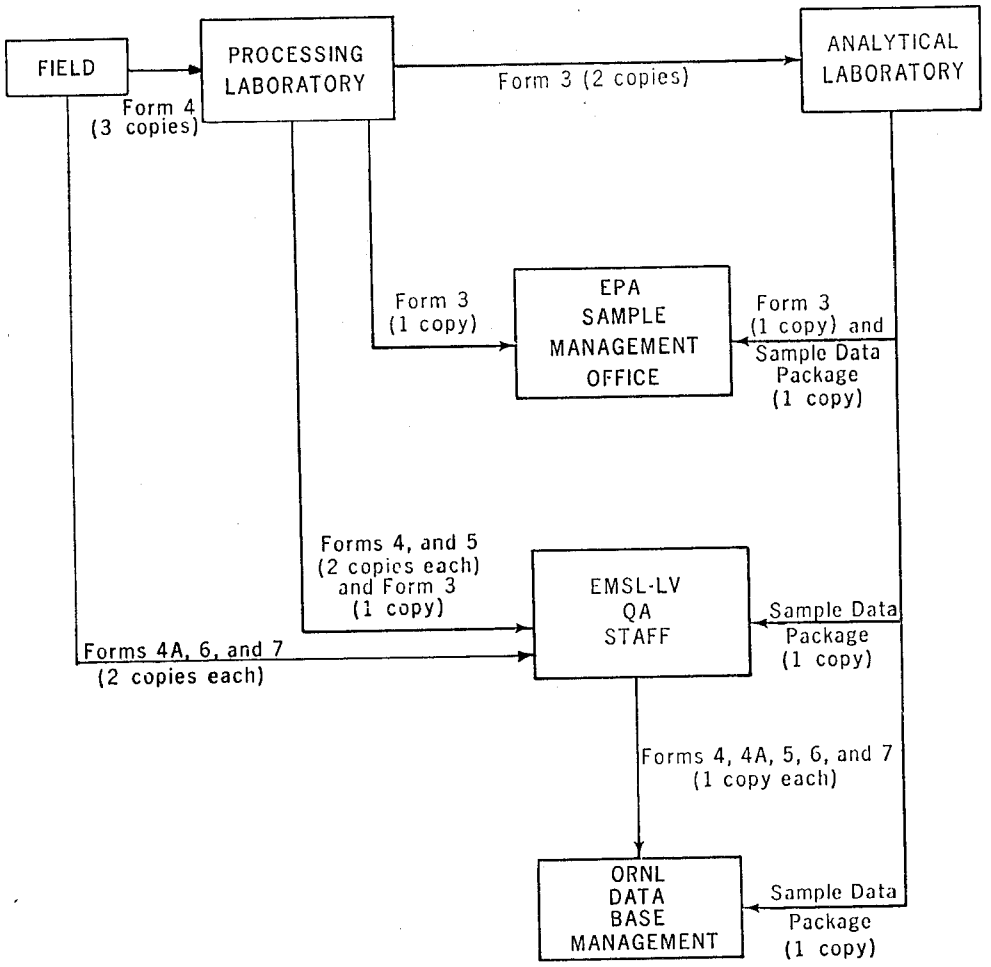


Figure 6. Field sampling activities for the National Stream Survey - Phase I.

6. checked data forms before transfer to the EMSL-LV QA staff; and

7. prepared and shipped reagents and supplies to the field base sites.

Arent et al. (in preparation) give a detailed discussion of processing laboratory protocols for NSS-I. Figure 8 shows the flow of samples and data from the field through the processing laboratory, and a brief description of processing laboratory activities follows. Samples were processed on the same day they were received. When the shipment arrived, the analysts inspected the samples for proper identification and for shipping damage, and noted comments concerning the samples on the sample log-in sheet. Each sample was assigned a unique batch and sample number combination to distinguish it from any other sample in the survey. There were 68 batches of samples (numbered from 2100 to 2167) analyzed during the NSS-I. Each batch of samples contained routine samples, one field (or processing laboratory) blank, one field duplicate, and at least one audit sample. Each



NSWS FORMS

Field and Processing Laboratory

Form	Description
3	Shipping
4	Stream Data
4A	Hydrologic Data
5	Batch/QC Processing Laboratory Data
6	Stream Episode Data
7	Watershed Characteristics

Analytical Laboratory Sample Data Package

Form	Description
11	Summary of Sample Results
13	ANC and BNC Analyses Results
14[a]	QC Data for ANC and BNC Analyses
15[a]	Specific Conductance (Measured and Calculated)
16[a]	Anion-Cation Balance Calculations
17	Ion Chromatography Resolution Test
18	Detection Limits
19	Sample Holding Time Summary
20	Blanks and QCCS Results
21	Dilution Factors
22	Duplicates Results

[a] Form not required to be submitted with data package but recommended for internal QC requirements.

EMSL-LV - U.S. EPA, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada
ORNL - Oak Ridge National Laboratory, Oak Ridge, Tennessee

Figure 7. Data form flow, National Stream Survey - Phase I.

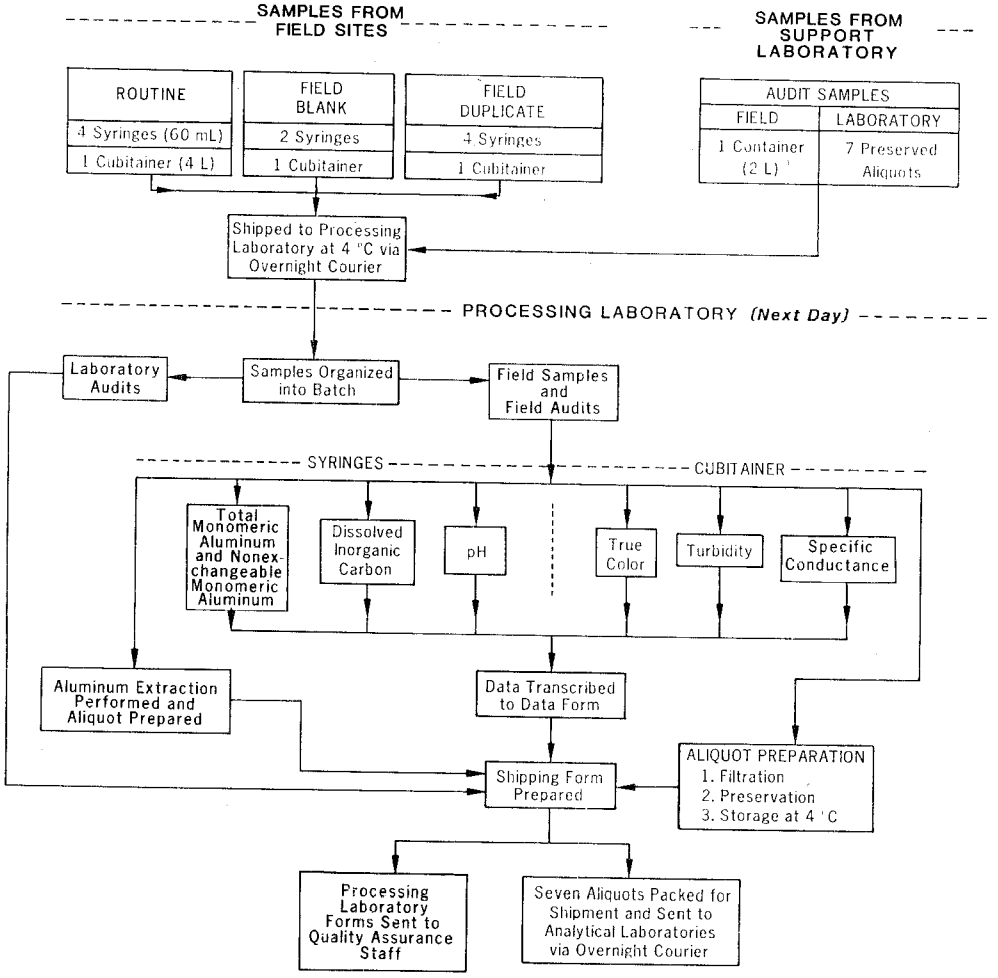


Figure 8. Flow of samples and data from the field through the processing laboratory.

routine, blank, duplicate, and audit sample was randomly numbered within the batch. Each batch was sent as a unit to a specific analytical laboratory. A batch contained up to 40 samples; the fewest number of samples in a batch was 8.

Generally, samples from the mid-Atlantic and southeast screening sites were grouped in the same batch. However, if the total number of incoming samples (including duplicates, audits, and blanks) exceeded the number of sample analyses required of a laboratory in the SOW, then separate batches were prepared for mid-Atlantic and southeast screening samples. In this case, each batch contained a blank, a duplicate, and an audit sample. The communications center personnel at EMSL-LV informed the base site coordinators whenever it was necessary to collect more than one field blank and duplicate to accomodate the large sample load. Each of the two batches were sent to different analytical laboratories. After the new contract with Laboratory 2 became effective, both batches were sent to this laboratory.

After sorting the samples into batches, the four syringes collected in the field were distributed to analysts to measure pH, DIC,

and total and nonexchangeable monomeric aluminum species and to prepare the total extractable aluminum aliquot. The samples collected in sealed syringes allowed measurements at the processing laboratory, within a short holding time, of some variables (pH, DIC, and the aluminum species) that tend to become unstable over time. The sealed syringes minimized chemical changes before analysis. The contents of the Cubitainers were divided into six additional aliquots and subsamples from each Cubitainer were used to obtain specific conductance, turbidity, and true color measurements. Figure 9 shows the preparation and preservation procedures for each aliquot.

The instruments and methods used for analyses are listed in Table 1. Processing laboratory analytical methods are described by Hillman et al. (1987). Quality control check samples used in the processing laboratory were measured as specified in the QA plan (Drousé et al., 1986a).

At the processing laboratory the procedures for preparing and preserving each of the seven aliquots taken from each Cubitainer sample were specific for the variable to be measured at the analytical laboratories. The aliquots were stabilized by using filtration, acid preservation, refrigeration, or some combination of these procedures. Filtration removed suspended material in order to reduce biological activity and to eliminate surfaces that could adsorb or release dissolved chemical species. Acid was added to some aliquots to prevent loss of dissolved analytes through precipitation, chemical reaction, or biological activity. All aliquots were stored and shipped at 4°C to inhibit biological activity and, in the case of total extractable aluminum aliquots, to reduce volatilization of solvent.

Once the samples were preserved, the aliquots were prepared and packed in a shipping container with frozen gel packs and sent by overnight courier to the analytical laboratories. Extractable aluminum aliquots were separated from the other aliquots. These aliquots were inserted into a Styrofoam rack and packed in a separate shipping container that contained frozen gel packs.

A shipping form, Form 3, was completed and copies were sent with the aliquots to the analytical laboratories and to the EPA Sample Management Office (Figure 7). As soon as shipping activities were completed, the processing laboratory personnel notified the EMSL-LV communications center which tracked custody of the samples from the field to the processing laboratory to the analytical laboratories.

Analytical data, QC data, and comments pertinent to sample analyses were recorded in bound laboratory logbooks and then on the batch/QC Form 5 (Figure 7). All logbook data and forms were reviewed by the processing laboratory supervisor or coordinator to ensure that calibration and QC checks were within the required limits and that all comments and qualifiers were complete and understandable. All forms were then sent to the QA staff in Las Vegas for review of data consistency before transmittal to ORNL for data entry.

Analytical Laboratory Operations

Analytical laboratory personnel were responsible for inspecting the samples received from the processing laboratory for damage, logging in the sample batches, analyzing the samples according to procedures described in the statement of work and published in the NSS-I analytical methods manual (Hillman et al., 1987), and preparing and distributing data packages (Figure 7) containing the analytical results. For each shipment, laboratory personnel recorded all notes concerning sample condition on the shipping form and sent a copy of the annotated form to the EPA Sample Management Office.

As part of the contract requirements, the analytical laboratories agreed to follow good laboratory practices related to laboratory cleanliness and the use and storage of reagents, solvents, and gases. For standard guidelines regarding general laboratory

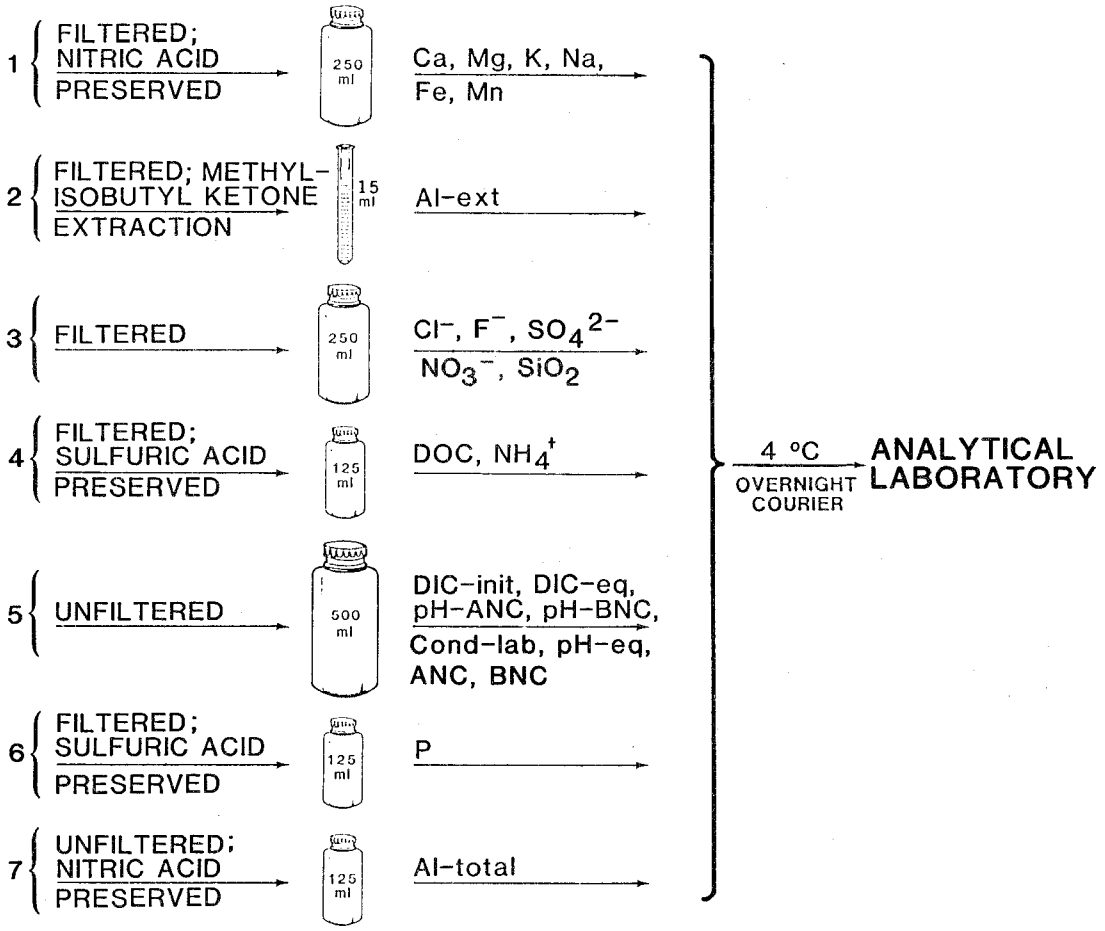


Figure 9. **Preparation and preservation procedures for each aliquot at the preparation laboratory for the National Stream Survey - Phase I.**

practices, the analytical laboratories were directed to follow procedures in the Handbook for Analytical Quality Control in Water and Wastewater Laboratories (U.S. EPA, 1979). The analytical laboratories also were required to operate according to a uniform set of internal QC procedures, as described in the QA plan (Drousé et al., 1986a), to check data consistency, and to document method performance. Table 1 lists the analytical instrument or method used for each variable.

A maximum sample holding time, determined from the time of sample preservation to sample analysis, was established for each variable measured in the analytical laboratories (Table 5). These holding times were based upon information from the literature, the best scientific judgment related to the defined needs, and the logistical demands and limitations of the NSS-I. After all initial analyses were completed, the analytical laboratories refrigerated the samples at a temperature of 4 °C in case reanalyses were necessary. The samples remained at the laboratories for approximately 6 to 12 months or until notice was received from the EMSL-LV QA manager to dispose of the samples or ship them to EMSL-LV for storage.

Each data package prepared at the analytical laboratories included a set of NSWS forms (Drousé et al., 1986a) containing the following information:

1. Measured sample concentration in the appropriate units for each variable.

2. Titrant concentration and titration data points for each sample for ANC and BNC.

3. Percent conductance difference calculation for each sample (optional; this calculation is an initial check made in the analytical laboratory to ensure data consistency, but it is also performed during data verification under the direction of the EMSL-LV QA manager).

4. Percent ion balance difference calculation for each sample (optional; this calculation is an initial check made in the analytical laboratory to ensure data consistency, but it is also performed during data verification under the direction of the EMSL-LV QA manager).

5. Ion chromatograph specifications.

6. Instrument detection limits.

7. Date of sample analysis and sample holding time.

8. Calibration and reagent blank values and QCCS values.

9. Internal (laboratory) duplicate precision calculated as percent relative standard deviation.

Each data package included a cover letter from the analytical laboratory manager to the QA group at EMSL-LV. The letter specified the batch ID number and the number of samples analyzed, identified all problems associated with the analyses, described any deviations from protocol, and contained other information that the laboratory manager considered pertinent to a particular sample or to the entire batch. Copies of the completed data package were sent to the EMSL-LV QA staff for review, to ORNL for data entry, and to the EPA Sample Management Office for sample tracking.

Table 5. Maximum Holding Time Requirements[a] Before Sample Analysis at Analytical Laboratories, National Stream Survey - Phase I

Variable	Holding time
Nitrate, total extractable aluminum[b]	7 days
ANC, BNC, specific conductance, DIC, DOC, pH[c]	14 days
Phosphorus, ammonium, chloride, sulfate, fluoride, silica	28 days
Calcium, iron, potassium, magnesium, manganese, sodium, total monomeric aluminum	28 days[d]

[a] Number of days between sample preservation and sample analysis.

[b] Although the EPA (U.S. EPA, 1983) recommends that nitrate in unpreserved samples (unacidified) be determined within 48 hours of collection, evidence exists (Peden, 1981, and APHA et al., 1985) that nitrate is stable for 2 to 4 weeks if stored in the dark at 4 °C.

[c] Although the EPA (U.S. EPA, 1983) recommends that pH be measured immediately after sample collection, evidence exists (McQuaker et al., 1983) that pH is stable for as long as 15 days if the sample is stored at 4 °C and sealed from the atmosphere. The pH is also measured in a sealed sample at the processing laboratory the day after sample collection.

[d] Although the EPA (U.S. EPA, 1983) recommends a 6-month holding time for these metals, the NSS-I required that all of the metals be determined within 28 days. This requirement ensured that significant changes would not occur and that data would be obtained in a timely manner.

Monitoring

Communications

Monitoring QA activities required continuous communication among the many groups responsible for data collection, verification, and validation. These communications were centralized through the QA staff and a communications center at EMSL-LV. Staff members at the communications center monitored all field activities including sample shipment, the number of streams sampled, weather, sampling projections, supply requests, and miscellaneous problems. The center served as a point of contact for all technical and logistical questions, provided a backup contact for sampling teams when base site and logistics coordinators were unavailable, coordinated the assignment of duplicate and blank samples to the base sites, and served as a liaison between the field operations and the Las Vegas processing laboratory.

The communications center personnel were responsible for tracking sample shipments from the field to the processing laboratory to the analytical laboratories and for locating lost shipments, if necessary. They were also responsible for ordering audit samples and communicating with the audit sample support laboratory. Any appropriate information from the field, processing laboratory, analytical laboratories, or support laboratory was relayed to the QA staff.

Members of the QA staff communicated directly with the field crews and the processing and analytical laboratory personnel on a daily basis. These daily telephone calls were necessary to discuss sampling, processing, and analytical issues related to logistics, methods, QA, and QC so that problems could be resolved quickly and efficiently, and to obtain current sample data and QA and QC information. The QA staff also communicated periodically with:

1. analytical methods experts at EMSL-LV to resolve issues related to analytical methodology,
2. the EPA Sample Management Office concerning sample tracking and analytical laboratory compliance with contractual requirements,
3. the support laboratory personnel to address questions relating to sample preparation, and
4. the data base management group at ORNL to clarify the meaning of comments, to decipher illegible data for data entry, and to discuss data base design and data entry progress.

During the NSS-I, conference calls were held regularly (either weekly or every two weeks, depending on need) for all survey participants to aid in efficient exchange of information, problem solving, and improvements. Discussions during these calls covered survey progress, protocol changes, and issues related to sample collection, sample load and analyses, raw data set development, resolutions of problems relating to QA issues, data evaluation, and progress of report writing.

On-Site Inspections

On-site inspections of field and laboratory activities were conducted to ensure that sampling and analytical procedures were being performed according to the survey protocol. The two mid-Atlantic field base sites, the processing laboratory, and Laboratory 2 were evaluated during the NSS-I. QA and QC sample data were reviewed thoroughly and used in conjunction with on-site evaluations to confirm proper operations and to identify any necessary changes in protocol or the need for reanalysis. The findings from these evaluations were documented in on-site inspection reports. Because of budget constraints at the time of sample analyses, it was not possible to evaluate Laboratory 1 on-site during the analyses of NSS-I samples. However, an on-site inspection performed at this laboratory during the ELS-I determined that all analytical operations followed correct protocol.

Data Base Management and Data Verification

The creation of the four NSS-I data sets (raw, verified, validated, and enhanced) involved numerous operational steps as well as several NSS-I participants. To create a raw data set, all data were entered into two separate data sets by two different operators at Oak Ridge National Laboratory (ORNL). A custom program developed using the Statistical Analysis System (SAS Institute, Inc., 1985) compared the two data sets and identified any inconsistencies in numeric and alphabetic variables. Any errors were corrected by referring to the original forms. All NSS-I data sets were created and maintained at ORNL by using the Statistical Analysis System. When the data sets were complete, they were transferred via magnetic tape to the National Computer Center at Research Triangle Park, North Carolina. There, scientists at the Las Vegas and Corvallis laboratories could gain access to the data sets. The QA staff members in Las Vegas were primarily responsible for data verification and personnel at Corvallis were responsible for data validation.

In order to meet the objectives of the data verification process and to identify raw data of questionable or unacceptable quality that might need to be corrected during or after data validation, the QA auditors examined the data for internal consistency and reviewed the QA and QC data measured and recorded at the sampling sites, the processing laboratory, and the analytical laboratories. Geographical data were verified and validated at the Corvallis laboratory; analytical data were verified at the Las Vegas laboratory. Computer programs provided a mechanism to automate much of the verification procedure. Redundant values, calculated by computer and measured at more than one place, were compared. The verification process evaluated data at both the sample and batch levels. Data verification took place in two parts: initial verification of the numerical changes and final verification that involved the final numerical changes as well as the addition of data qualifier flags. These verification activities are identified in figures 10 and 11.

Review of Field and Processing Laboratory Data Forms--

Verification began when the data forms from the field and processing laboratory (Forms 4, 4A, 5, 6, and 7; Figure 7) were received at EMSL-LV. A QA auditor reviewed the data forms for completeness, for agreement of stream identification codes given on the field and processing laboratory forms, and for proper assignment of sample identification codes and data qualifier tags. Specific conductance and pH measurements recorded on field and processing laboratory forms were compared in order to identify possible measurement or reporting errors. The auditors calculated precision estimates using field and processing laboratory routine-duplicate data and evaluated the estimates using the data quality objective for each variable (see Section 6) as a reference. Measurements for field audit samples were evaluated using measurement data provided by the support laboratory as a reference. Measurements for the audit samples were compared with results from previous NSWS surveys for the same audit types.

Data anomalies were reported to the field base site and processing laboratory coordinators for corrective action, and data reporting errors were corrected before the data were entered into the raw data set. After reviewing the information on all field and processing laboratory data forms for completeness and accuracy, the QA staff sent the forms to the Oak Ridge laboratory by overnight courier.

At times, data errors were identified by communications between the QA staff, field and processing laboratory personnel, and the ORNL staff after the data forms were sent to Oak Ridge. If the data in question were not yet entered into the raw data set, the QA staff sent documentation with instructions to make changes to the erroneous data. Most transcription errors were identified and corrected

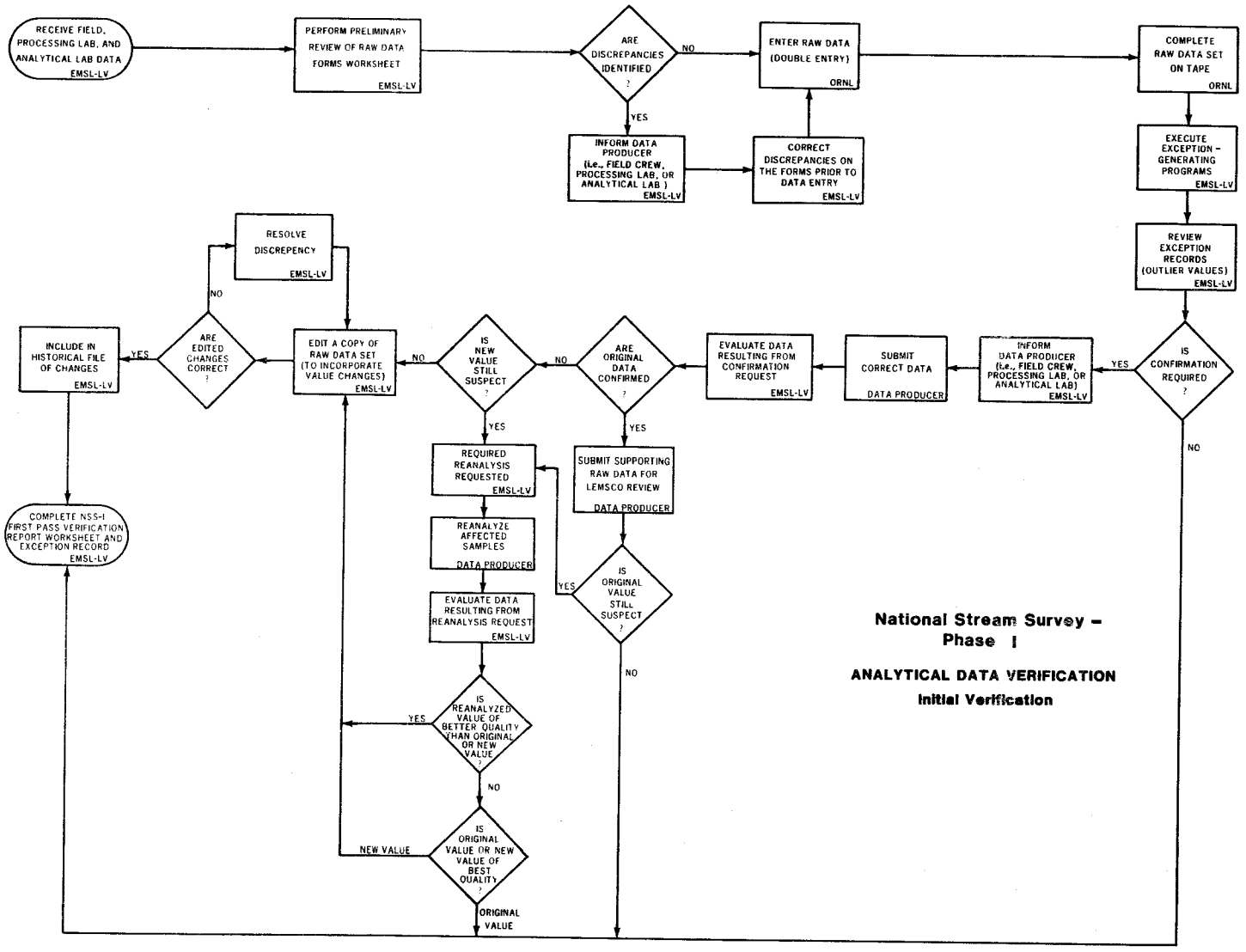


Figure 10. Analytical data verification, initial verification for the National Stream Survey - Phase I.

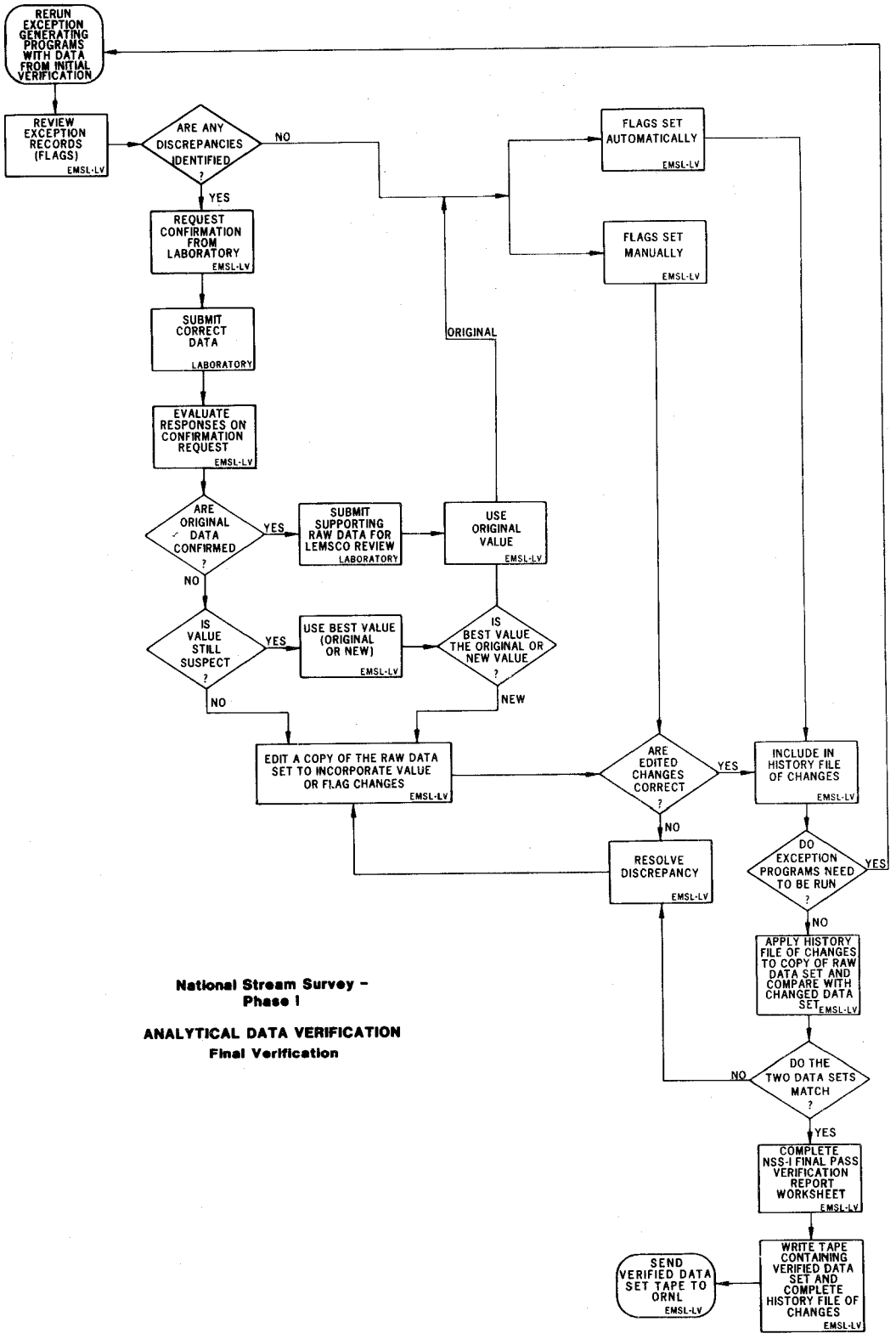


Figure 11. Analytical data verification, final verification for the National Stream Survey - Phase I.

before data entry. All changes made at Oak Ridge were initialed and dated by Oak Ridge personnel. A copy of this documentation was sent to the QA staff as confirmation that the changes had been made. Any data errors identified after the original data were entered into the raw data set were corrected during verification and were incorporated into the verified data set.

Review of Preliminary Results from the Analytical Laboratories--

In response to inquiries from the EMSL-LV QA staff, personnel at the analytical laboratories reported preliminary results of analyses on a daily basis, either by telephone or by electronic data transfer, depending on available resources. When requesting information, the QA staff members would inquire about selected QA samples and a group of randomly chosen routine samples to minimize the possibility that the laboratory personnel would identify the QA samples. These preliminary results were evaluated to ensure that the acceptance criteria (Drousé et al., 1986a) for the QA samples were met and that data analyses were performed according to protocol. Whenever any problems were noted, the QA auditor conferred with laboratory personnel to determine the source of the problem and to implement corrective action. The primary objective of this preliminary evaluation was to identify and resolve issues quickly before they affected data quality or interfered with the completion of the survey.

Initial Review of Analytical Laboratory Sample Data Package--

The analytical laboratories sent the analytical data packages by overnight courier to Oak Ridge and to Las Vegas for simultaneous QA review. While personnel at Oak Ridge were entering the data packages into the raw data set, the QA auditors reviewed the analytical laboratory data packages manually. They reviewed the sample data packages for completeness, internal QC compliance, and appropriate use of data qualifiers. Each auditor used a checklist to ensure consistency in this data review process (Drousé et al., 1986a).

At the Las Vegas laboratory, the QA sample data for each batch were tabulated for each laboratory in a computer file before the raw data set was available. The field blank data were evaluated to determine if the values fell within expected criteria (Appendix C). Because the sampling protocol for obtaining blank samples for the mid-Atlantic and southeast screening surveys was identical to that used during the NSS-I Pilot, the concentrations found in the blank samples from the pilot survey were used to calculate control limits for the NSS-I blank samples. These control limits were used to (1) check for evidence of sample contamination, (2) determine the necessity of data confirmation or reanalysis, and (3) generate data qualifier flags to indicate potential contamination by batch or by sample. The 95th percentile of the NSS-I Pilot field blanks was chosen for the upper control limit, except for a few variables for which the value of the required detection limit (see Section 6) was used. The negative of the value for the required detection limit was used for the lower control limit to account for background noise and minor fluctuations in instrument performance. Concentration values below this limit would indicate possible negative bias. Histograms were also developed with the field blank data for each laboratory to aid in detecting unacceptable data for each variable.

The values for the variables measured in a field routine-duplicate pair were considered an exception when the concentration of both samples in a pair exceeded the required detection limit by a factor of 10 and the precision estimate exceeded the acceptance criteria (Appendix C) developed by the QA staff. Precision was calculated as percent relative standard deviation. The acceptance criteria were established to meet the precision DQOs (see Section 6) although some flexibility was allowed to compensate for the variance due to sample handling.

Bar histograms were developed by the QA staff with the audit sample data for each

variable. Auditors compared the results from each laboratory for each audit sample in order to detect any problems inherent within a laboratory. Again, reported concentrations from the audit samples analyzed in the support laboratory or historical NSWS data were used as a reference before the raw data set was available in order to identify and correct unacceptable data early in the survey. Quality control charts were developed with the audit sample data in order to detect suspect data points. The auditor requested that the analytical laboratory manager confirm all data points outside control limits.

Any repeated measurements made in the field, processing laboratory, and analytical laboratory such as for specific conductance, dissolved inorganic carbon, and pH were compared. Any discrepancies were reported to the field base sites, processing laboratory, or analytical laboratory for corrective action.

Automated Review of NSS-I Data Base--

Once the development of the raw data set was completed at Oak Ridge and the data set was available to the EMSL-LV QA staff, the data were reviewed by using the AQUARIUS II exception-generating and data review programs (Table 6). These computer programs were used to identify or flag results that were exceptions, i.e., results that did not meet the expected QA and QC limits. The auditors used the output from these programs, along with original data and field notebooks, to complete the NSWS verification report specified in the QA plan (Drousé et al., 1986a). The verification report is a worksheet designed to guide the auditor systematically through the verification process by listing the steps that lead to identification of QA exceptions, explaining how to flag data, tracking data resubmissions and requests for confirmations and reanalyses, and summarizing any required modifications to the raw data set (i.e., preparing records of numerical and flag changes required to create the verified data set). Flags and their definitions are given in Appendix B.

Each sample value was verified individually and by analytical batch. All samples had to meet internal consistency checks for percent ion balance difference and for percent conductance difference. Percent ion balance difference (%IBD) is calculated as follows:

$$\left(\frac{\Sigma \text{ anions} - \Sigma \text{ cations} + \text{ANC}}{\Sigma \text{ anions} + \Sigma \text{ cations} + \text{ANC} + 2\,[H^+]} \right) 100$$

where:

Σ anions = $[Cl^-] + [F^-] + [NO_3^-] + [SO_4^{2-}]$

Σ cations = $[Na^+] + [K^+] + [Ca^{2+}] + [Mg^{2+}] + [NH_4^+]$

ANC $\simeq$ alkalinity (the ANC value is included in the calculation to account for the presence of unmeasured ions such as organic ions)

$$[H^+] = (10^{-pH}) \times 10^6 \ \mu eq/L$$

Note: Brackets indicate concentration of an ion in microequivalents per liter.

A list of factors for converting mg/L to μeq/L for each variable is given in the analytical methods manual (Hillman et al., 1986). After confirmation of the original suspect values, samples which had a poor ion balance were flagged or reanalyzed unless a high DOC measurement accounted for the difference. Table 7 lists the acceptance criteria for the ion balance difference.

The calculation program was modified so that whenever the absolute value of ANC was less than or equal to 10 μeq/L, the value zero was substituted for ANC in the equation. The equation is sensitive to slight variations in ANC for samples that have very low ionic strength.

The percent conductance balance is determined as follows:

Table 6. Exception-Generating Programs Within the Aquarius II Data Review and Verification System

Program	Sample (data) type
Exception-generating programs:	
Audit Sample Summary	Field natural and synthetic and laboratory natural and synthetic
Field Blank Summary	Blank
Field Duplicate Precision Summary	Routine-duplicate pairs
Instrumental Detection Limit Summary	All species
Holding Time Summary	All species
% Conductance Difference Calculations	All species
Anion/Cation Balance Calculations	All species
Internal Laboratory Duplicates	Analytical laboratory duplicate data
Protolyte Analysis	DIC, DOC, pH, ANC, and BNC data evaluation
Reagent/Calibration Blanks and QCCS	All species except pH
Data review programs:	
Comparison of Total Aluminum and Extractable Aluminum	Total aluminum and extractable aluminum
Raw Data Listing	All field and laboratory data
Comparison of Form 4 and Form 5	pH, DIC, and specific conductance
Comparison of Form 5 and Form 11	pH, DIC, and specific conductance
QA/QC Flag Summary	All exceptions
Modified Gran Analysis Program	ANC and BNC
Audit Sample Window Generation	All species

Table 7. Chemical Reanalysis Criteria for Sample Ion Balance Difference and Percent Specific Conductance Difference

A. Anion-Cation Balance

Total ion strength (μeq/L)	Maximum % ion balance difference[a]
<50	60
≥50 and <100	30
≥100	15

B. Specific Conductance

Measured specific conductance (μS/cm)	Maximum % specific conductance difference[a]
<5	50
≥5 and <30	30
≥30	20

[a] If the absolute value of the percent difference exceeds these values, the sample is reanalyzed. When reanalysis is indicated, the data for each parameter are examined for possible analytical error. Suspect results are then redetermined and the above percent differences are recalculated (Peden, 1981). If the percent differences for reanalyzed samples are still unacceptable or no suspect data are identified, the QA manager must be contacted for guidance.

$$\left(\frac{\text{calculated conductance} - \text{measured conductance}}{\text{measured conductance}}\right)100$$

The ions used to calculate conductance are Ca, Cl^-, CO_3^{2-}, H^+, HCO_3^-, K^+, Mg, Na, NO_3^-, OH^-, and SO_4^{2-}. Calculated conductance is determined by multiplying the concentration of each ion by the appropriate factor given in Table 8. All three measured specific conductance values from the field, processing laboratory, and the analytical laboratory are compared to the calculated value. The acceptance criteria for the differences are listed in Table 7. Any routine stream sample or QA sample that did not fall within the applicable criterion was qualified with a flag or reanalyzed.

On the basis of the analytical results reported for QA and QC samples, the QA staff directed the analytical laboratory to confirm reported values or to reanalyze selected samples or sample batches, if necessary.

Generally, reanalyses were requested when at least three different QA and QC samples generated flags for a particular variable in a particular batch or when incorrect methodology was used during the original analysis. In such cases, sample reanalysis usually was requested for a given variable on a per-batch basis. A tracking form for data confirmation and sample reanalysis provided a standard format for data transfer between the QA staff and the analytical laboratories. Suspect data that were not corrected through confirmation or reanalysis were flagged with an appropriate data qualifier when the exception-generating programs were rerun during final verification. Additional data qualifiers were added to a given variable if the QA samples (field blanks, field duplicates, or audit samples) within the same analytical batch did not meet the acceptance criteria. Acceptance criteria were calculated for the audit samples on two different occasions. The first calculation was with Data Set 1 when it was received from ORNL and before the initial verification. The second calculation was after all the numerical changes were made to the raw data and before final verification. The QA plan (Drousé et al., 1986a) gives a detailed description of the method for calculation of acceptance criteria for audit samples. Every batch that contained an audit sample with an unacceptable analyte concentration was flagged accordingly. Acceptance criteria for the audit samples are found in Appendix C.

Data were also given qualifier flags if internal QC checks (such as calibration and reagent blank analyses, internal duplicate precision, required instrumental detection limit, QCCS central limit criteria, and maximum allowable holding times) were not met. The protolyte analysis program identified discrepancies related to processing and analytical laboratory measurements of pH, DIC, ANC, BNC, and DOC when carbonate equilibria, corrected for organic species, did not have internal agreement. Flags were produced for data that were questionable. The overall process involved:

1. performing redundant alkalinity calculations using three different measurements for both pH (pH-closed, pH-ANC, and pH-eq) and DIC (DIC-closed, DIC-init, and DIC-eq),

2. verifiying measured ANC and BNC, and

3. determining whether the system is carbonate or mixed. Empirical relationships defined by Oliver et al. (1983) were used to estimate the contribution of organic prototypes to the measured ANC. A data qualifier flag was assigned for samples in which the presence of organic species resulted in an ion imbalance for the sample.

Another program compared the extractable aluminum and total aluminum values for each sample. By definition, the extractable aluminum concentration for a sample could not exceed the total aluminum concentration. The program generated a flag when the value for extractable aluminum was at least 0.015 mg/L and when it was higher than the value for total aluminum by more than 0.010 mg/L (twice the required detection limit). This qualification was intended to account for background noise

Table 8. Factors for Determining the Conductances of Ions (μS/cm at 25 °C)[a,b]

Ion	Factor per mg/L	Ion	Factor per mg/L
Calcium	2.60	Magnesium	3.82
Chloride	2.14	Sodium	2.13
Hydrogen (H^+)	3.5×10^5 (per mole/L)	Ammonium	4.13
		Sulfate	1.54
Hydroxyl (OH^-)	1.92×10^5 (per mole/L)	Nitrate	1.15
Bicarbonate (HCO_3^-)	0.715	Potassium	1.84
Carbonate (CO_3^{2-})	2.82		

[a] Taken from APHA et al. (1985) and Weast (1972). Ion concentration is multiplied by the listed factor to obtain the conductance value. The concentrations of the ions that are not measured directly are calculated by means of the following equations:

$$[H^+] = 10^{-pH}$$

where: pH = initial pH measured before BNC titration. (Brackets represent molar concentrations.)

$$[OH^-] = \frac{Kw}{[H^+]}$$

where: $K_w = 1 \times 10^{-13.8}$ moles/L

$$HCO_3^- \text{ (mg/L)} = \frac{5.080\ (DIC(mg/L))\ [H^+]K_1}{[H^+]^2 + [H^+]\ K_1 + K_1K_2}$$

$$CO_3^{2-} \text{ (mg/L)} = \frac{4.996\ (DIC(mg/L))K_1K_2}{[H^+]^2 + [H^+]K_1 + K_1K_2}$$

where: $K_1 = 4.4463 \times 10^{-7}$ moles/L, $K_2 = 4.6881 \times 10^{-11}$ moles/L, and
K_1 = first ionization constants for carbonic acid
K_2 = second ionization constants for carbonic acid
K_w = ionization constant for water

[b] Conductance factors are not given for ionic aluminum, iron, or manganese because these ions are rarely present in concentrations great enough to affect the percent conductance difference.

(especially at low levels) and for minor fluctuations in instrument reading and calibration. In all cases, each flag generated by the AQUARIUS II system was evaluated by an auditor for reasonableness and consistency before it was entered into the verified data set.

Comparison of upstream and downstream data for all streams and comparison of first and second site visits for streams in the mid-Atlantic region provided an additional method to identify anomalies. If the differences in the data appeared suspicious and there were no obvious reasons for the differences such as a point source of pollution (e.g., mine tailings) or a recent rain event noted on the field forms, then confirmation of the suspect data was requested from the analytical laboratories.

Changes and additions to the raw data set resulting from corrections to analytical data or from reanalysis of samples (Form 4 from the field, Form 5 from the processing laboratory, and the sample data package from the analytical laboratory) were made by the QA staff during initial verification. The QA staff also added all appropriate data qualifiers not marked on the original data forms. This verified data set was sent to Oak Ridge National Laboratory on magnetic tape. Personnel at Oak Ridge added changes resulting from corrections to nonanalytical forms (Forms 4A, 6, and 7) provided by the EMSL-LV QA staff on a modification sheet (Drousé et al., 1986a) to create the official verified data set.

Preparation and Delivery of Verification Tapes--

Three separate versions of the verified data were delivered on magnetic tape to ORNL where they were made available to the ERL-C staff for validation. The first tape was delivered after the initial run of the exception-generating programs. All numeric changes identified at this time were incorporated into the changed data set, and the initial verification was complete. An intermediate tape was delivered midway through final verification in order to aid the ERL-C staff in data assessment. The third and final tape was delivered after all the reanalyzed and corrected data were incorporated into the data set, the exception-generating programs were rerun and evaluated, and all the data qualifiers were applied to the data set.

The third and final verified data set was generated by the EMSL-LV QA computer support staff. The tape was sent to Oak Ridge where it was checked for consistency before use during data validation. Oak Ridge was responsible for generating the official verified data set (Data Set 2) and for archiving the tape.

Data Validation and Data Base Management

Data validation took place primarily at ERL-C. The validation process is a way to search for observations that may represent entry or analytical errors or unusual water chemistry. This process incorporates univariate, bivariate, and multivariate analyses (Drousé et al., 1986a).

Validation of NSS-I data began with the raw unverified data. During validation, a matrix for each subregion was constructed that depicted the results of validation checks on each individual water sample and datum. Using this matrix, outliers were identified and sent to the QA staff at EMSL-LV to be checked for possible entry errors. Sites having atypical chemistry when compared to other sites in a subregion (unique multivariate relationships) were identified and evaluated as unusual sites (e.g., sites affected by acid mine drainage, agricultural impact, or tidal influence). Data from unusual sites generally appeared as outliers in several statistical analyses (e.g., regression and univariate statistics).

In addition to identifying outliers, validation identified stream samples that might have been collected during a precipitation or snowmelt episode (or influenced by it) and therefore would not provide an acceptable index of base flow chemistry (Kaufmann et al., 1988). Although great care was taken not to

collect samples during an episode, such conditions might not have been apparent to a sampling team.

Sites considered to be of noninterest for calculation of NSS-I population estimates were identified as those satisfying certain criteria (Kaufmann et al., 1988). When data for an entire reach were considered unacceptable for the intended use--to make population estimates, for example--a disturbance flag code was placed in the data set. Disturbances that might affect stream reach data include acid mine sites, pollution, tidal influence, and watershed disturbances.

Enhanced Data Set

After data validation, the enhanced data set was prepared to resolve problems with erroneous data and missing values in the validated data set. (Kaufmann et al., 1988). In cases where it was deemed necessary, substitutions were performed according to the following criteria:

1. Values from duplicate samples were used whenever possible.

2. If a duplicate measurement was not available, a value from an alternate visit to the site was used.

3. If a duplicate measurement or a measurement from an alternate visit was not available, a substitution value was calculated by means of a linear regression model. This was done by (l) calculating a predicted value based on observed relationships with other chemical variables or (2) predicting a value based on relationships between upstream and downstream observations of the same chemical variable.

4. The last option for identifying a substitution value was to use the subregional mean.

All substitute values were examined for acceptability before they were included in the final data set. In addition to the substitute values that were calculated, negative values for parameters other than ANC and BNC that resulted from analytical calibration bias were set equal to zero. Streams considered to be of noninterest were flagged in a manner by which they could be excluded when using the enhanced data set for making regional estimates of the target population.

An index value for each chemical variable for each NSS-I sampling site was calculated by averaging data from duplicate pairs and multiple sample visits. The resulting data set contains a single value for each variable for each sampling site (i.e., one observation for each upstream and downstream site).

Section 5

Results and Discussion - Assessment of Operations

Field Sampling Operations and Protocol Changes

The EMSL-LV QA staff conducted on-site inspections of field sampling operations at both of the mid-Atlantic survey base sites. Observations of presampling calibration, QC procedures, sampling methods, sample handling, and sample shipment indicated that the proper protocols were followed. Field personnel strictly adhered to QA and QC protocols, accurately documented problems, and took corrective action when necessary. Due to time constraints, the QA staff did not visit and inspect the southeast screening field operations; however, NSS-I management and supervisory personnel did visit these field operations. Their observations of the sampling activities, sample handling, and sample shipments indicated that all required protocols were followed for activities of the southeast screening survey. All relevant findings from the QA inspection at the mid-Atlantic sites were forwarded to the screening operations. Protocol changes and problems encountered during the field sampling operations for the NSS-I are discussed in detail in Hagley et al. (in press). QA issues identified during the on-site evaluations and in the course of the survey are described in Table 9. Further information relating to the data base variable fields mentioned in this section can be found in the data base dictionary (Sale, in press).

Processing Laboratory Operations and Protocol Changes

Two on-site QA inspections were performed at the processing laboratory during the survey. The laboratory staff followed protocols and all activities were generally satisfactory. Some QA related issues were resolved as a result of the inspections. Protocol changes were implemented in response to the sample load, concentration values, and recommendations for improvements from the EMSL-LV methods development group, QA staff, or processing laboratory staff. These issues are listed in Table 10 and the most complex are discussed further in the following paragraphs.

Specific Conductance Measurements

Processing laboratory conductance measurements are not reliable from batch 2100 through batch 2127. The conductance probe was thought to be faulty early in the survey because increasing values for field blank samples were noted with each new batch. Attempts to repair the probe were unsuccessful. An alternative probe was obtained from the EMSL-LV methods laboratory and was put into use with batch 2104. This probe appeared to operate initially, but beginning with batch 2118, high values were again noted for field blanks. The processing laboratory water systems, which were the sources for the field blank samples, were suspected to be the probable cause of the increasing values for the blank samples. After further investigation, an additional new probe was obtained and was put into use beginning with batch 2128. This newer probe provided acceptable field blank measurements and it was determined that there was not a problem with the processing laboratory water system.

Nitrate Contamination

At the beginning of survey operations, preliminary data from the analytical laboratories indicated nitrate contamination in

Table 9. Significant Findings Concerning Field Sampling Operations and Their Effect on Data, National Stream Survey - Phase I

Finding	Corrective action	Effect on NSS-I data
On two occasions, samplers realized on the second visit that they had sampled the wrong stream on the first visit.	Data for the streams that should not have been sampled were qualified with an XO flag (Appendix B) on the sample identification code in the verified and validated data sets.	No apparent effect. First-visit data for these two streams were not available for the creation of the enhanced data set. Second visit data were used.
Since 1986 was an unusually dry year, a number of streams that might normally be flowing during the spring were completely dry or stagnant. A number of streams could only be sampled at one site because more than 90 percent of the reach was dry.	None.	None. Enough water samples were collected from the target stream population so that the DQO for completeness was met (see Section 6).
Temperatures in the shipping coolers sometimes deviated significantly from the recommended 4 °C upon arrival at the processing laboratory.	Numbers and types of frozen-gel packs were adjusted.	No apparent effect. Temperature of the cooler in which the sample was shipped is noted in the variable field, COOLR, in the verified and validated data sets.
Three sample shipments (8 samples total) were misrouted by the overnight courier service.	Extra effort was expended by the field crew to track and recover the samples.	No apparent effect. Samples processed after the 30-hour holding time were qualified on the sample identification code in the raw, verified, and validated data sets.
Some shipping containers were damaged during shipment causing leaks in 31 of 1,512 Cubitainers that were shipped from the field. Of the 5,912 syringes that were shipped, 13 were received with broken syringe tips.	The use of hard plastic containers or the taping of the more fragile Styrofoam containers for reinforcement prevented any further breakage.	No apparent effect. Data for samples from these damaged containers are qualified in the tag field in the raw, verified, and validated data sets.
Measured values for specific conductance QCCSs used in the field were consistently outside acceptance criteria (Hagley et al., in press) in the beginning of the survey.	Processing laboratory personnel changed the original protocol which called for daily preparation using volume dilution techniques to a protocol which called for using weight dilution techniques and preparing the solution in bulk quantities. More careful attention to good laboratory techniques also resolved this problem.	Comparisons of specific conductance measurements made in the field and at the analytical laboratories indicated that field measurements were acceptable. Data for QCCSs that were not acceptable were qualified in the tag field in the raw and verified data sets.

Table 10. Significant Findings Concerning Processing Laboratory Operations and Their Effect on Data, National Stream Survey - Phase I

Finding	Corrective action	Effect on NSS-I data
Underestimation of daily sample loads required that additional processing laboratory analysts be hired during survey operations. The necessity of a shortened training period covering only one or two methods resulted in a loss of four specific conductance values and one turbidity value; these samples were inadvertently discarded before measurement.	Careful monitoring of the newly trained analysts eliminated this problem.	A turbidity measurement for one sample was lost. For specific conductance, the field or analytical laboratory measurements could be used in place of the processing laboratory measurements.
Field duplicate samples were not assigned randomly in 16 batches.	Processing laboratory personnel were notified of this incorrect practice and the problem was resolved.	Apparently, the analytical laboratory personnel did not identify these consecutive samples as duplicates and there was no effect on data quality.
Early in the survey, one field routine-duplicate pair was inadvertently split between two batches and hence between two analytical laboratories.	Processing laboratory personnel were notified. The problem did not occur again.	Batches 2104 and 2105 do not have field routine-duplicate pairs that can be used as QA samples.
An initial comparison between processing laboratory and analytical laboratory specific conductance measurements indicated that the processing laboratory values were not temperature compensated to 25 °C as required.	After the survey was completed, all processing laboratory specific conductance data were corrected to 25 °C by using the temperature data recorded in the analysts' logbooks. These corrected values are included in the verified, validated, and enhanced data sets. The original, uncorrected values are included in the raw data set.	None.
Processing laboratory specific conductance measurements for the first 28 batches often did not agree with the field and analytical laboratory measurements. See text for further discussion.	The use of a new conductance probe provided acceptable data. See text for further discussion.	For batches 2100 through 2127, specific conductance data from the processing laboratory are considered unreliable. See text for further discussion.

(continued)

Table 10. (Continued)

Finding	Corrective action	Effect on NSS-I data
In previous NSWS surveys, only one pH meter was used per batch of samples; more than one was used during the NSS-I to complete daily analyses of the large sample load.	A dilute-buffer check solution was developed and was measured daily to ensure that the meters produced comparable results (Arent et al., in preparation). Field routine-duplicate pairs always were analyzed on the same meter.	None. The meters produced comparable results. The identification number of the meter that was used to analyze each sample is recorded in the raw and verified data sets in the variable field PHMID.
At the beginning of the survey, nitrate contamination was identified in some QA samples. See text for further discussion.	A change in the filtration procedure and more care during the preservation process eliminated the problem. See text for futher discussion.	Eight samples were identified as nitrate contaminated and are identified by a data qualifier flag in the verified and validated data sets. See text for further discussion.
Filtration of the stream samples was time consuming and labor intensive. Membranes had to be changed frequently during filtration.	For future stream sample analyses, a two-stage filtration procedure that employs a coarse filter in addition to the 0.45-μm filter that was used during the NSS-I is recommended.	None.
The protocol for measuring turbidity was originally developed for the Eastern Lake Survey - Phase I and was based on an expected range of 0 to 20 nephelometric turbidity measurments exceeded this range.	The protocol was modified to include a procedure for high-level samples. Modifications included a separate calibration, QCCS, and dilution procedures (Hillman et al., 1987).	None.
In previous NSWS surveys, the buffer used in the extracable aluminum procedure for Aliquot 2 was ammonium chloride/ammonia. It is thought that volatile ammonium chloride fumes generated by this buffer can coat the surface of labware and the laminar flowhood, resulting in chloride and amonium contamination.	The buffer was changed to amonium acetate/ammonia to eliminate the potential ammonium chloride contamination problem.	Possible improvement in data quality.

(Continued)

Table 10. (Continued)

Finding	Corrective action	Effect on NSS-I data
Unexpected colored precipitates (black, brown, purple, yellow, green) developed with the aluminum extraction for some sample including some of the performance audit samples. In all pH and aluminum ranges.	None. The EMSL-LV methods development group analyzed these samples for metals that may have produced such colors, but no identifiable trend was indicated.	None.
The method that uses flow injection analysis for the measurement of total monomeric aluminum and nonexchangeable monomeric aluminum was under development in the processing laboratory at the time of the NSS-I. Numerous hardware and software problems occured. See text for further descussion.	The protocol was mofified with development of the method. See text for further discussion.	Many samples were analyzed for these aluminum species weeks after the batch was processed. See text for further discussion.

some field blank and performance audit samples. Aliquot 3 is the aliquot in which anions, including nitrate, are measured in the analytical laboratory (Figure 9). During an on-site processing laboratory evaluation, it was observed that the non-acid-washed filtration apparatus used for Aliquot 3 was located in the middle of a series of nitric acid-washed filtration apparatus in order to expedite processing. This practice allowed two technicians to filter samples simultaneously so that deadlines for shipment to the analytical laboratories could be met.

It was suspected that the cause of the nitrate contamination was twofold. Contamination of less than 1.00 mg/L was probably due to nitric acid splashing from the acid-washed units during rinsing steps of the filtration procedure. For higher levels of contamination, it was suspected that the analysts mistakenly preserved Aliquot 3 with nitric acid during the sample preservation procedure. More care during the preservation procedure and the installation of a Plexiglas shield around the non-acid-washed filtration apparatus eliminated the contamination problem. Of the 68 field and processing laboratory blank samples, 4 are suspected to be nitrate contaminated. Of the 68 field audit samples, 4 are suspected to be contaminated.

Of the 1,381 routine stream samples, only one is suspected to be contaminated. Contaminated samples are qualified with an X3 data qualifier flag (Appendix B) in the verified and validated data sets.

Total Monomeric and Nonexchangeable Monomeric Aluminum Measurements

The protocol for measuring total monomeric and nonexchangeable monomeric aluminum by flow injection analysis (FIA) for a large-scale operation was under development at the time of the NSS-I. The QA plan (Drousé et al., 1986a) specifies that the QCCS control limits for all aluminum determinations in the NSS-I must be within $\pm$20 percent. However, in the development of the protocol, attempts were made to determine if more stringent limits ($\pm$10 percent) could be established for these two aluminum measurements. With

development of the method, it was clear that the 10 percent limit would be too stringent. The control limits were successfully widened to ±15 percent for the total monomeric fraction and to ±20 percent for the nonexchangeable fraction for the remainder of the survey. The required frequency of QCCS analysis was also decreased from every 5 samples to every 10 samples.

Finally, because the sample concentration for these variables often exceeded the expected range (0 to 1.50 mg/L) during the NSS-I, a calibration procedure for samples that contained high concentrations was included in the protocol. This calibration pro- cedure included analysis of QCCSs appropriate for the concentration range and a requirement for duplicate analysis for each separate calibration. Further detail is provided in Arent et al. (in preparation) and Hillman et al. (1987).

There were numerous hardware and software problems with the FIA in the developmental stage of the protocol. Because of these problems, several batches of samples could not be analyzed on the day of receipt as the QA plan required. A total of 13 batches (285 samples) were refrigerated and were analyzed as long as 4 weeks after sample collection.

The quality of data from the backlogged samples is uncertain for several reasons. The effects of holding time and atmospheric carbon dioxide exposure on aluminum speciation are not fully known. It was necessary to recycle syringe valves back to the field because the ongoing surveys exhausted the manufacturer's supply. Because the syringes were no longer airtight throughout the holding time before analysis and the samples were thus exposed to atmospheric carbon dioxide, a modification to the protocol was implemented to hasten analyses. The samples were filtered through syringe filters into sample cups and then were analyzed by using an open-air autosampler rather than by direct syringe injection. A data qualifier and comment indicating that the backlogged samples were analyzed by modified analytical protocol or out of holding time protocol were applied to the variables (ALDSVL_T, total (monomeric alumi- num tag field, and ALORVL_T, nonexchange- able monomeric aluminum tag field) in the raw, verified, and validated data sets. The associated comments for these tags are found in the variable field COMMO5.

Analytical Laboratory Operations and Protocol Changes

Through preliminary evaluation of the data, on-site evaluations, and data verification, the QA program was instrumental in identifying and resolving several significant problems at the analytical laboratories. Appropriate changes were incorporated in the verified data set. The most significant issues are discussed below.

Effect of Large Sample Loads

The consistent incoming sample load of 40 samples or more each day was a hardship on Laboratory 2, which accommodated this situation by adding a second work shift. Instrument malfunctions, especially for the carbon analyzer, resulted in a backlog of samples which then exceeded sample holding time requirements for dissolved inorganic carbon and dissolved organic carbon measurements by a few days. All measurements for which the analyses exceeded sample holding time allowances are qualified with a tag in the verified and validated data sets. Of the 1,083 samples measured for both initial dissolved inorganic carbon and equilibrated dissolved inorganic carbon at Laboratory 2, 434 samples (40.1 percent) were analyzed outside the holding time requirement. Of the 1,079 samples analyzed for dissolved organic carbon, 398 samples (36.9 percent) were analyzed outside the holding time requirements. Twenty-five of the initial dissolved inorganic carbon measurements and six of the equilibrated dissolved inorganic carbon measurements were identified as questionable during data verification. A data qualifier flag (Appendix B) was applied to these questionable values.

Laboratory 1 measured 567 samples for both initial dissolved organic carbon and equilibrated dissolved inorganic carbon. Twenty-six analyses (4.6 percent) exceeded the holding time requirements for both measurements. This laboratory measured 568 samples for dissolved organic carbon. Fifteen (2.6 percent) of these analyses exceeded holding time requirements. Holding times were exceeded at Laboratory 1 due to an error in holding time calculations and were not due to instrument malfunction. None of the measurements made at Laboratory 1 for these late analyses were identified as questionable in the veri- fication process. No apparent data quality problem exists because of these late analyses.

Centrifuge Tubes for Extractable Aluminum Analyses

Early in the survey, four plastic centrifuge tubes for extractable aluminum analyses (Aliquot 2) were damaged during sample shipment. These tubes were composed of a different material than tubes used in previous NSWS surveys. Tests con- ducted at the processing laboratory indicated the tubes were extremely fragile. By simply placing the tube in a test-tube rack, the tube could crack. The fragility was thought to be caused by the acid-washing procedure used to clean the tubes. Because it was not possible to procure tubes similar to those used in the previous surveys, the processing laboratory staff packed the centrifuge tubes in Styrofoam racks and in shipping containers separate from the other six aliquots to prevent further breakage. For future surveys, attempts should be made to procure less fragile centrifuge tubes.

Laboratory pH Data

One of the QA checks that the auditors performed during the verification process was a comparison of the initial pH values recorded on the ANC and BNC titration data form (Form 13) with the pH values recorded on the analytical data form (Form 11). On the ANC and BNC titration data form, the laboratories were required to report measured pH values and pH values that were calculated as a result of applying electrode calibration factors to the measured values. The pH values reported on the analytical data form and then entered into the data base should be those that were measured and not calculated. However, Laboratory 2 incorrectly reported the calculated pH values on the analytical data form. The QA staff replaced the calculated values with the measured pH values in the verified data set. Because the QA program identified this problem, the data were not affected.

ANC and BNC Recalculations

As allowed by the contracts, both analytical laboratories developed their own software for the calculation of ANC and BNC and used the Gran analysis algorithm described in the statement of work as a guide. The laboratories submitted the values calculated by using their own software, and these values were included as part of the raw data set used for data verification. During the verification process, certain inconsistencies in the values reported by the two laboratories became apparent. Further analysis revealed shortcomings in both calculation methods used by the laboratories; therefore, the QA staff recalculated all ANC and BNC values by using software prepared at EMSL-LV. These recalculations not only corrected the identified shortcomings in the software used by the analytical laboratories, but also eliminated interlaboratory bias that could be attributed to the differences in software.

A new program, GRANNI.EXE, made it possible to do these recalculations. This program is a noninteractive Gran analysis program that includes a consistent point selection routine and uses the algorithm given in the statement of work.

All ANC and BNC values submitted by the laboratories were recalculated. The values originally submitted were replaced with the new values in the verified data set. Almost all of the new values were calculated using GRANNI.EXE. This algorithm did fail in certain cases (poor titration data) and interactive software was used when necessary.

After the verified data set was delivered to ORNL, EMSL-LV developed an improved data point selection algorithm. The values for the NSS-I data set were recalculated by using a new program, GRAN.EXE, and were delivered to ORNL after review by ERL-C. These new values are used in the analysis of the QA results presented in this report and are included in the official verified data set generated by ORNL.

Sample Holding Time and Reanalysis for Metals

The allowed sample holding time for the metal analyses (Aliquot 1) in the NSWS was 28 days. Sample analyses within this holding time allowed data bases to be created within time frames set by the EPA. Because the EPA-recommended holding time for these metals is 6 months (U.S. EPA, 1983), and the samples are considered stable for that period of time, reanalysis during the NSS-I was requested for some metals even though the 28-day limit specified in the QA plan had been exceeded. All results from analyses performed on samples outside sample holding time requirements are qualified with an H in the tag variable field in the verified and validated data sets.

Calcium Reanalysis--

During preliminary data evaluation, histograms and QC charts were developed with the performance audit sample data. The QA staff compared these histograms and QC charts to those developed by laboratories involved in other NSWS programs. A positive bias was identified in the analysis of calcium by Laboratory 2 and was traced to a difference in nitric acid content of standards (calibration and QC) and survey samples. The standards prepared by the analytical laboratory contained 1.25 percent nitric acid and the samples, after preservation in the processing laboratory, contained approximately 0.15 percent. The higher concentration of nitric acid in the standards suppressed the calcium analytical signal resulting in a positive bias of approximately twenty-five percent in the sample results. The bias was not identified because the NSS-I QC standards also were prepared with the incorrect quantity of acid. The acid concentration in the samples was clearly marked on the aliquot bottles. Therefore the analytical laboratory did not follow protocol when preparing its standards. The laboratory was directed to reanalyze all affected samples (batches 2104 through 2147, 917 samples) using calibration and QC standards containing 0.15 percent nitric acid. After the reanalyses, no significant interlaboratory bias was identified (see Section 6). All samples were analyzed within the 6-month holding time recommended by EPA for metals. The identification of this problem by the QA staff demonstrates the success of maintaining a QA program in which the use of QC charts and performance audit samples is standard protocol.

Total Aluminum Reanalysis--

During analyses of NSS-I samples, an on-site inspection was performed at Laboratory 2. During the inspection, the QA staff discovered that laboratory analysts were using the protocol for total recoverable aluminum rather than the protocol for total aluminum that was required for the NSS-I analyses. The method for total recoverable aluminum calls for a less rigorous digestion procedure than the method for total aluminum. Also, the total aluminum results would be biased lower than those obtained by using the correct methodology. The laboratory was directed to use the designated digestion procedure to redigest and reanalyze all samples (42 batches) that were originally analyzed with the incorrect methodology. The laboratory reanalyzed 1,083 samples within the 6-month holding time recommended by EPA for metals. Due to funding restrictions, an on-site evaluation was not performed at Laboratory 2 during the analyses for NSS-I samples until analyses of two-thirds of the samples had been completed. Therefore, the request for reanalyses involved many samples. For future surveys, thorough on-site evaluations should be performed early in a survey; follow-up evaluations are also recommended.

Magnesium Reanalysis--

Fifty-four samples (batches 2113 and 2116) were reanalyzed for magnesium by Laboratory 2 because QC charts indicated control values slightly outside the 95 percent control limits. The new results are slightly higher than the original values, and the negative bias indicated from the original control charts is eliminated.

Reanalyses of Nitrate, Sulfate, and Chloride

Some stream samples contained very high sulfate concentrations that made analyses of low-concentration nitrate samples difficult with ion chromatography. Data evaluation of results from Laboratory 1 established that if the samples were not diluted, these high-level sulfate samples produced chromatogram peaks that overwhelmed the nitrate peaks and sometimes the chloride peaks. The result was off-scale sulfate readings that masked actual nitrate and chloride peaks and yielded 0 mg/L values for these analytes.

This problem was not identified during initial analyses at the laboratory. After the problem was identified, the laboratory was requested to reanalyze the 10 questionable samples for sulfate, nitrate, and chloride after sample dilution, although the sample holding time was exceeded. Because the original analyses yielded such poor results, it was thought that reanalysis outside the holding time requirement would be an improvement over the zero values originally reported and would provide more information to the users. All data resulting from the reanalysis of these samples are qualified by both H and R tags (Appendix B) in the verified and validated data sets.

Total Extractable Aluminum Values Greater than Total Aluminum Values

During the data verification process, both analytical laboratories identified several samples for which the total extractable aluminum concentration was greater than the total aluminum concentration. Samples for which the extractable aluminum concentration was greater than or equal to 0.015 mg/L and for which this value exceeded the total aluminum concentration by 0.010 mg/L were reanalyzed for total aluminum after confirmation of the originally reported values by the laboratory. The samples were not reanalyzed for total extractable aluminum because (1) the seven-day holding time was exceeded and (2) the problem was thought to have occurred in the digestion procedure for total aluminum. If reanalysis did not provide improved results, the values for total extractable aluminum and total aluminum were qualified with an X1 flag in the verified data set.

This practice of requesting reanalyses using the above guidelines is probably too stringent considering that the deviation allowed by the control limits for QCCS analyses was $\pm$20 percent for both aluminum measurements. For future surveys, error bounds of $\pm$20 percent for each measurement may be appropriate criteria to determine if data are unacceptable and reanalysis is required.

Data Reporting Errors

Data reporting errors that were identified during the survey included variable concentrations incorrectly reported as 0 mg/L, reagent blanks subtracted in error, and the incorrect number of decimal places reported.

Total Dissolved Fluoride--

Total dissolved fluoride is determined by the ion-selective electrode technique. The contracts awarded to the laboratories suggest the use of a digital potentiometer with expanded mV scale capable of reading in 0.1 mV increments. The required detection limit for fluoride was 0.005 mg/L.

For the determination of fluoride, Laboratory 2 used an instrument that did not have the capability of measuring low-concentration samples (less than 0.010 mg/L) if calibrated for higher concentration samples (greater than 0.010 mg/L). The laboratory did not recalibrate or use two different

instruments, one for high concentrations and one for low concentrations. Because sample concentrations below 0.010 mg/L could not be detected, all values below this threshold concentration were recorded as 0.000 mg/L.

During QA data analyses, the QA staff observed that Laboratory 2 had reported fluoride values for all field and laboratory blank samples and for stream samples that were less than 0.010 mg/L as zero. Although the laboratory consistently reported a calculated instrument detection limit (IDL) of less than 0.005 mg/L, which is within contract specifications, the laboratory did not follow the contract guidelines in determining the IDL. The laboratory calculated the IDL as three times the standard deviation of ten nonconsecutive low-concentration standards which were greater than 0.010 mg/L, instead of using values for laboratory blank samples for these calculations. In the verified and validated data sets (Data Sets 2 and 3), an M1 flag was applied to all zero values for fluoride reported by Laboratory 2. This flag indicates that the value was not actually measured and may be inaccurate.

Preliminary QA data analyses using the raw data set, prior to data verification, would have indicated problems early enough to permit fluoride reanalyses. For future surveys, preliminary evaluations should be a project priority.

Total Dissolved Phosphorus, Ammonium, and Silica--

During QA data analyses, the QA staff discovered several instances where values were misreported by Laboratory 2. The laboratory had reported the theoretical value, 0 mg/L, for all total dissolved phosphorous, ammonium, and silica calibration blanks rather than the measured value.

The inspection of raw data also revealed that Laboratory 2 reported total dissolved phosphorous, ammonium, and silica concentrations that originally were measured as negative during analyses as 0 mg/L. This error affected field blank and stream sample values. The laboratory submitted corrected values which are included in the verified, validated, and enhanced data sets and are used in the QA data analyses.

Subtraction of Values for Silica and Ammonium Reagent Blanks--

During QA analysis, the QA staff discovered that silica and ammonium reagent blank values originally were subtracted from all sample concentrations measured by Laboratory 2. Because this practice did not follow protocol, values for reagent blanks for both variables were added to the reported values. The corrected values are included in the official verified, validated, and enhanced data sets.

Decimal Place Reporting--

The initial contracts with both laboratories recommended that values be reported to the number of decimal places in the instrument detection limit (IDL). The second contract with Laboratory 2 recommended that values be reported to the IDL, plus one, or to a maximum of four significant figures (Table 11).

Laboratory 1 consistently provided values with more decimal places than recommended in the first contract. Laboratory 2 consistently reported values as they do in their standard laboratory procedure, that is, to the number of significant figures that they considered meaningful for that concentration of analyte (usually to three significant figures). Although data interpretation and population estimates (Kaufmann et al., 1988) were not affected, this inconsistency created difficulty in the statistical analysis of QA data. This inconsistency could be prevented if future contracts "require" rather than "recommend" the number of decimal places to be reported.

Data Verification Activities

The QA staff reviewed field and processing laboratory forms and analytical data packages to identify and to correct data reporting errors, to evaluate data trends, and to identify which samples needed reanalysis.

Table 11. Recommended Number of Decimal Places

Variable	Original contracts for laboratories 1 and 2	New contract for laboratory 2
Acid-neutralizing capacity	1	1
Aluminum, total	3	4
Aluminum, total extractable	3	4
Ammonium	2	3
Base-neutralizing capacity	1	1
Calcium	2	3
Chloride	2	3
Dissolved inorganic carbon, equilibrated	2	3
Dissolved inorganic carbon, initial	2	3
Dissolved organic carbon	1	2
Fluoride	3	4
Iron	2	3
Magnesium	2	3
Manganese	2	3
pH, equilibrated	2	2
pH, initial acid-neutralizing capacity	2	2
pH, initial base-neutralizing capacity	2	2
Phosphorus, total dissolved	3	4
Potassium	2	3
Silica	2	3
Sodium	2	3
Specific conductance, analytical laboratory	1	1
Sulfate	2	3
Nitrate	3	4

All forms used in the NSS-I are given in the QA plan (Drousé et al., 1986a). Any required changes to the data on these forms resulted in changes to the raw data set and were reflected in the verified data set. The types and numbers of changes made to create the verified data set are given in Table 12.

Review of Field Data Forms

The review of field data forms and the subsequent additions or corrections made to them required a great deal of time and effort during the NSS-I. Due to sample shipment deadlines, the base coordinators made only cursory review of the stream data forms (Form 4) before packing and shipping them with the samples. When the field crews shipped the samples from a remote site, it was not possible for the base coordinator to review the forms. The hydrology and site characteristic forms (Forms 4A, 6, and 7) were not sent to the QA staff until the base coordinator reviewed them after sample shipment. In the beginning of the survey, this process took several days.

When the base coordinator found errors on the forms after they had been sent to the QA staff, the changes were submitted over the telephone, and were followed by hard copy documentation. If the ORNL staff had not already entered the original data, the changes were forwarded to ORNL for entry. If the data had been entered into the raw data set, any changes were made to a copy of the raw data set by the EMSL-LV QA staff in order to create the verified data set. The concept of a "raw data set" was that no changes would be applied after the data were entered. All changes after initial data entry are made to subsequent data sets.

There were two significant problems with this system. The hydrology and site description forms were received by the QA staff piecemeal, and there were numerous changes to the forms after they were shipped from the field. In the future, more emphasis on correct data form completion would minimize changes. The number of NSS-I forms that the base coordinators were required to review took more time than was available. Designating a field member to assist the base coordinator in reviewing forms may expedite the review process and at least make it possible for all forms to be shipped by the following day.

Review of Processing Laboratory Forms

The magnitude of the NSS-I sample processing effort, in conjunction with the concurrent sample processing for the Eastern

Table 12. Changes to Sample Numeric Data Incorporated in the Verified Data Set, National Stream Survey - Phase I

Data source (form number)	Number of changes made from raw to verified data sets[a]	Percent of changes to total values from data source	Comments
Field Data Forms (Form 4)[b]	2 (5,618)	<0.1	Most changes to the field forms were not numerical.
Processing Laboratory Forms (Form 5)[b]	1,600 (10,831)	14.8	Most of these changes are the result of temperature corrections for Cond-PL.
Analytical Laboratory Stream and QA Data Forms (Form 11)[b]	9,011 (39,570)	22.8	Most of these changes resulted from corrections to Al-total, ANC, BNC, Ca, pH-ANC, and pH-BNC.
Total	10,613 (56,019)	19.0	

[a] Number in parentheses is the number of numerical observations in the verified data set.
[b] Changes to flags, tags, and QC data are not included in this table.

Lake Survey - Phase II, made daily form completion difficult in the beginning of these surveys. A backlog of forms developed at the processing laboratory before they were available for the QA staff review. After the processing laboratory was operating more efficiently, the forms were delivered daily for QA review.

The processing laboratory analytical data form (Form 5) was completed correctly most of the time. It was less difficult to correct these forms because the processing laboratory and the QA staff were both located in Las Vegas.

Review of Analytical Data Forms and Correction of Data

Review and verification of the sample data packages submitted by the analytical laboratories was a bigger task than review of the field and processing laboratory data. The QA staff always requested confirmation of suspect data before reanalysis was requested. After confirmation was requested, a response usually took two to five weeks. Due to the reanalyses requested for calcium and total aluminum, Laboratory 2 required three months to complete the task. These analyses were performed within the 6-month holding time recommended by EPA. There was no specific requirement for response time in the contracts for NSS-I. The contracts required a response within a "reasonable" amount of time. A prompt response to the reanalysis request was necessary to meet sample holding time requirements. All reanalyzed sample values incorporated in the verified data set are qualified with an R tag (Appendix B) in the variable tag field. All changes to the analytical laboratory data were documented on the Confirmation/Reanalysis Request Form, Form 26 (Drousé et al., 1986a), or in revised data packages submitted by the laboratories.

Changes to Analytical Data Applied at EMSL-LV

All hydrology and site description changes for the verified data set were made at ORNL, and all changes to the analytical data were madeby the EMSL-LV QA staff. These changes originated from all three data sources: field forms, processing laboratory forms, and analytical laboratory forms. The EMSL-LV QA staff made changes using transaction records that were applied to a copy of the raw data set. The ORNL staff made changes by editing directly into a copy of the raw data set using the SAS full-screen edit facility. Each change

went through a series of checks before it was considered final. The changes were entered from the modification sheets (Drousé et al., 1986a) or from the revised analytical data packages by a data entry technician. A different technician checked the values for accuracy before moving them into the changed data set. After the update of the transaction records, the changed data set was checked point by point to confirm that the intended changes were made correctly. The changes consisted of sub-stituting correct values or adding data qualifiers. All changes to the analytical data and flags from the raw to the verified data set are documented in a history file of changes that was included with the verified data set on the magnetic tape sent to ORNL.

Modifications to the Exception-Generating Programs and New Data Qualifier Flags

The data qualifier flags used in the NSS-I (Appendix B) were similar to those used in all the NSWS surveys with a few modifications:

1. For the anion and cation balance check program, an A9 flag was used to indicate a possible analytical error with the ANC measurement.

2. For the conductance balance program, the original NSWS C7 qualifier indicated a conductance imbalance due to unmeasured protolyte anions. For the NSS-I, the definition of the C7 flag was changed to indicate an imbalance due to the influence of other anions and cations that are not included in the conductance balance calculation. In addition, a new flag, F6, was created for the NSS-I to indicate a problem with the processing laboratory specific conductance measurement.

3. For comparison of field and processing laboratory data in the protolyte analysis program, changes were made to the definitions of the original NSWS F flags to reflect the difference in stream field instrumentation from that used during lake surveys.

4. An M1 flag was created for the NSS-I and applied to all fluoride samples measured by Laboratory 2 that were reported as zero. This flag indicates that the value was not actually measured and therefore may be inaccurate.

5. Additional miscellaneous X flags were used in the NSS-I. The flag, X3, indicates a potential gross contamination of the aliquot. The X7 flags were added to the flag field of the sample identification number to indicate a site disturbance, such as a strip mine or sewage treatment plant, in the watershed.

Delivery of Verification Tapes

The original intention of the NSS-I verification and validation process was to deliver only two verification data tapes to ORNL: the first with numerical changes and the final with the data qualifier flags. However, due to the magnitude of the value changes resulting from reanalyses, three data tapes were delivered. It was not possible to include all reanalyzed data in an intermediate tape needed by the ERL-C staff for data assessment and validation issues. Therefore, the value changes submitted by the analytical laboratories up to that time were included in the intermediate data tape, and the remainder of the value changes were included in the final verified data set with the data qualifier flags delivered to ORNL on magnetic tape. At ORNL, the tape was checked for consistency before the changes to the nonanalytical data were made. ORNL then created the official verified data set that was used in data validation.

Data Base Audit

At the conclusion of the verification process, a data base audit was performed by an independent organization. The audit consisted of reviewing the verification records, evaluating for accuracy the results generated by AQUARIUS II and other computer programs, reviewing the procedures used to

substitute for missing values, and determining the error rates associated with each aspect of the verification procedure. No incorrect value changes were detected in the verified data set and all value changes were well documented (Grosser and Pollack, in preparation).

Section 6

Assessment of Data Quality

Introduction

The quality assurance program of the National Stream Survey - Phase I (NSS-I) was successful in reducing to acceptable levels errors associated with the acquisition and subsequent reporting of data. The program was also successful in identifying and correcting potential problems related to data quality that occurred over the course of the NSS-I.

One purpose of the QA program was to determine if any corrective actions (e.g., reanalyses, qualifying unacceptable values) would improve the quality of the analytical data and, if so, to implement those actions. The second purpose of the assessments was to identify possible limitations of the data base that might affect data interpretation. This second purpose was accomplished by viewing the data in terms of representativeness, completeness, comparability, detectability, accuracy, and precision.

The six aspects of NSS-I data quality fall essentially into two groups. Completeness, representativeness, and comparability apply to the sampling design and to the verified data set. Detectability, accuracy, and precision quantify the performance of one or several components of the collection and measurement system. These properties are evaluated by comparing the data acquired from analysis of QA samples to the established data quality objectives. Data quality objectives for detectability, accuracy, and precision are presented in Table 13. The values given are performance targets that the analytical laboratories were expected to meet. Objectives were not established for field measurements, although these measurements were subject to QC protocols (see Section 4). For most variables, within-laboratory precision goals were established only for measured values greater than ten times the value of the detection limit objective (Drousé et al., 1986a). For other variables (e.g., total monomeric aluminum), objectives were set for specified ranges.

Some evaluations of detectability, accuracy, and precision were improved by the elimination of a small number of extreme values that were considered outliers. Whenever outliers were removed for a particular assessment, that fact is included in the appropriate text discussion. The removal of outlying values sometimes resulted in a difference between the number of samples collected or processed and the number of measured values for a particular analyte.

Completeness

The completeness of the NSS-I data base was a critical aspect of data quality. If an insufficient number of streams were sampled or if a large number of analytical results were invalid, the representativeness and comparability of the NSS-I data base could be compromised. The DQO for completeness was established as 90 percent before the start of the NSS-I. That is, 90 percent or more of the streams initially selected for sampling were expected to yield data that would meet QA criteria and that could thus be used for estimating the number of stream reaches with a particular chemical characterization of interest (e.g., measured ANC less than 0 μeq/L).

Completeness of the data base was evaluated based on the overall number of stream reaches from which samples were collected and on the percentage of acceptable data generated from these samples. A total of

Table 13. Analytical Data Quality Objectives For Detectability, Precision, and Accuracy For The National Stream Survey - Phase I

Variable (units)	Detection limit objective (units)	Within-laboratory precision (%RSD)[a]	Within-laboratory accuracy (%)
FIELD SITE			
pH, field (pH units)	--	--	±0.1[b]
Specific conductance (μS/cm)	--	--	5
Dissolved oxygen (mg/L)	--	--	5
Current velocity (m/s)	--	--	--
PROCESSING LABORATORY			
Aluminum (mg/L)			
Total monomeric	0.01	10 (>0.01 mg/L) 20 (≤0.01 mg/L)	10 (>0.01 mg/L) 20 (≤0.01 mg/L)
Nonexchangeable monomeric	0.01	10 (>0.01 mg/L) 20 (≤0.01 mg/L)	10 (>0.01 mg/L) 20 (≤0.01 mg/L)
Specific conductance (μS/cm)	[c]	3	5
pH, closed system (pH units)	--	0.1[b]	±0.1[b]
Dissolved inorganic carbon, closed system (mg/L)	0.05	10	10
True color (PCU)	0	5[b]	--
Turbidity (NTU)	2	10	10
ANALYTICAL LABORATORY			
Acid-neutralizing capacity (μeq/L)	[d]	10	10
Aluminum (mg/L)			
Total extractable	0.005	10 (>0.01 mg/L) 20 (≤0.01 mg/L)	10 (>0.01 mg/L) 20 (≤0.01 mg/L)

(Continued)

Table 13. (Continued)

Variable (units)	Detection limit objective (units)	Within-laboratory precision (%RSD)[a]	Within-laboratory accuracy (%)
Ammonium (mg/L)	0.01	5	10
Base-neutralizing capacity (μeq/L)	[d]	10	10
Calcium (mg/L)	0.01	5	10
Chloride (mg/L)	0.01	5	10
Specific conductance (μS/cm)	[c]	2	5
Dissolved inorganic carbon (mg/L)			
Initial	0.05	10	10
Equilibrated	0.05	10	10
Dissolved organic carbon (mg/L)	0.1	5 (>5.0 mg/L) 10 (≤5.0 mg/L)	10
Fluoride, total dissolved (mg/L)	0.005	5	10
Iron (mg/L)	0.01	10	10
Magnesium (mg/L)	0.01	5	10
Manganese (mg/L)	0.01	10	10
Nitrate (mg/L)	0.005	10	10
pH (pH units)			
Equilibrated	--	0.05[b]	±0.1[b]
Initial ANC	--	0.05[b]	±0.1[b]
Initial BNC	--	0.05[b]	±0.1[b]
Phosphorus, total dissolved (mg/L)	0.002	10 (>0.010 mg/L) 20 (≤0.010 mg/L)	10 (>0.010 mg/L) 20 (≤0.010 mg/L)
Potassium (mg/L)	0.01	5	10

(Continued)

Table 13. (Continued)

Variable (units)	Detection limit objective (units)	Within-laboratory precision (%RSD)[a]	Within-laboratory accuracy (%)
Silica (mg/L)	0.05	5	10
Sodium (mg/L)	0.01	5	10
Sulfate (mg/L)	0.05	5	10

[a] %RSD = percent relative standard deviation. Unless otherwise noted, this is the precision goal at concentrations greater than or equal to 10 times the required detection limit.
[b] Precision or accuracy goal in terms of applicable units.
[c] The mean of six nonconsecutive blank measurements must not exceed 0.9 μS/cm.
[d] The absolute value of each laboratory calibration blank measurement was required to be less than or equal to 10 μeq/L.

479 stream reaches were initially selected for sampling in the mid-Atlantic and southeast screening regions. In addition to those reaches selected for inclusion in the probability sample, the total included a number of reaches on "special interest" streams where research programs independent of the NSS-I were in progress.

Table 14 presents the number of streams by region from which samples were collected. Of all reaches individually identified for sampling, samples were collected from 429 (90 percent) at both upstream and downstream sites on every visit. Of the 1,406 visits scheduled to upstream and downstream sites, water samples were collected from 1,328 sites (95 percent). Only eight of the sites could not be sampled because of access permission difficulties or because of physical inaccessibility. Water samples were not collected from the other 70 sites because they were classified as nontarget reaches as specified in the NSS-I sampling design (e.g., the stream sites were influenced by salt water or no water was present in the streambed).

In the mid-Atlantic region, nine streams were not sampled: two streams were not sampled because of tidal influence, four streams were not sampled because the specific conductance measurement exceeded the 500 μs/cm criteria, and three streams were dry. Five streams were partially sampled: two streams were sampled at only one site because of tidal influence, one stream was sampled at only one site because of the high conductance measurement at one of its sites, and two streams were sampled at only one site because access permission was not obtained for the other site. In the southeast screening region, twenty-three streams were not sampled: 20 streams that might normally be flowing during the spring were completely dry or stagnant, one stream was not sampled because it was inaccessible as a result of hazardous conditions, one stream was not sampled because access permission was not obtained, and one stream was not sampled because it was inundated by a major water project. Thirteen streams were only sampled at one site: twelve streams were sampled at only one site because more than 90 percent of the reach was dry and one stream was sampled at only one site because access permission was denied at the other site.

Reported values are given for all physical and chemical variables for 1,613 (97.7 percent) of the 1,651 stream samples and QA samples listed in the NSS-I verified data set. Of these 1,613 samples, the verification process identified 97 samples (6 percent) with values for one or more variables that should be used only

Table 14. Summary of Streams Visited During the National Stream Survey - Phase I

Region	Total streams targeted (includes special interest)	Total special interest streams	Total streams sampled[a]	Number of streams sampled at only one site[b]	Number of streams not sampled
Mid-Atlantic	276	26	267	5	9
Southeast screening	203	3	180	13	23
Total	479	29	447	18	32

[a] Includes streams sampled at only one site.
[b] Missing upper or lower sites on one or both visits.

with caution. These values were qualified with an M1, X0, X1, or X3 flag (Appendix B) in the verified and validated data sets because of the likelihood of a contami- nation or an analytical method problem. During the data validation process and the creation of the enhanced data set, missing or unacceptable values were replaced with the values from field duplicate samples (if available) or by an estimate computed by one of several approaches described by Kaufmann et al. (in press). Not all values identified as suspect during the verification process were replaced in the enhanced data set because these values may not have been identified as statistical outliers on the subregional level during validation.

Overall, the completeness of the data base exceeded the DQO of 90 percent, and the representativeness and comparability of the NSS-I data base were not affected by incompleteness. The data base is sufficiently complete to provide representative spring base flow chemical indices with which to estimate the chemical status of streams in these areas.

Comparability

For the NSS-I, the confidence in the comparability (or compatibility) of data from samples collected and analyzed by many different individuals and organizations was maximized by the use of standardized protocols for sample collection, processing, and measurement. When the QA staff members identified deviations from the protocols during on-site evaluations or during daily data verification activities (see Section 4), prompt corrective actions helped to improve the comparability of data within the NSS-I data base. The comparability of NSS-I data could also be affected by systematic differences in performance between the participating laboratories (interlaboratory bias). Interlaboratory bias is evaluated as part of the assessments of accuracy and precision.

In addition, the NSS-I data base is comparable to data bases from other AERP programs, a critical objective of the NAPAP. The use and documentation of standard sampling and analytical methodologies and the large volume of QA and QC data present in the verified data set allow quantitative evaluations of data comparability to past and future studies. Data from analyses of performance audit samples from laboratories participating in the NSS-I and other NSWS programs are presented in Appendix A. These data may be useful in evaluating the comparability of the NSS-I data base to those of other NSWS projects. In addition, data from the special interest streams are included in the NSS-I data base. These data will be useful in classifying these sites. The data are also potentially useful in making regional extrapolations based on information gained from intensive study at these sites.

Representativeness

The statistical frame for NSS-I sampling was designed to ensure that analytical results would represent the stream chemical conditions in the subregions sampled. Standardized protocols defined the appropriate weather conditions for sampling activities, the criteria for selecting a sampling site on a reach, and the criteria for selecting a sampling location in the stream (Hagley et al., in press). These protocols helped to ensure that each sample collected was representative of spring base flow chemical conditions existing in the stream at the time of sampling.

Detectability

Two aspects of detectability were assessed for the NSS-I. Laboratory performance was assessed by estimating the minimum limit of detection for each analytical method except for pH, color, and turbidity. This "method-level" limit of detection represented the smallest quantity of a chemical variable that a method (or instrument) could measure reliably. The second aspect of detectability assessed was "background" or the quantity of a chemical variable that was introduced into streamwater samples during their collection, handling, and preparation for analysis. The assessment of background is especially important to data interpretation. Background quantities serve as decision points; they are the lowest concentration of a given chemical variable that can be identified (with specified statistical confidence) as having been present in streamwater samples at the time of collection. The estimation of background represents a "system-level" assessment of detectability.

Assessment of Method-Level Limits of Detection

Numerous operational definitions and computational approaches exist for estimating detection limits (Currie, 1968; Hubaux and Vos, 1970; American Chemical Society, 1980; Glaser et al., 1981; Keith et al., 1983; Long and Winefordner, 1983; Oppenheimer et al., 1983; Clayton et al., 1987). The data quality objectives for detection established for the NSS-I were based on the "limit of detection" advocated by the American Chemical Society (American Chemical Society, 1980; Keith et al., 1983). The limit of detection is defined as $3s_0$, where s_0 represents the standard deviation at the lowest level of measurement (Taylor, 1987), which is usually zero. This expression of the limit of detection does not specify the probability of falsely concluding that a chemical variable is absent (termed a false negative, β, or Type II error; Clayton et al., 1987). Specifying $3s_0$ provides Type I and Type II error rates of approximately 7 percent each. This limitation is not critical when assessing NSS-I data quality, although it may be important in work that tests for the presence (or absence) of a toxic substance.

For the NSS-I, s_0 was estimated from laboratory blank samples (i.e., calibration blanks or reagent blanks) and from field blank samples, rather than from the analyses of low concentration standard solutions. The use of blanks is advocated by Campbell and Scott (1985) and Hunt and Wilson (1986), while the use of low-level standards is advocated by Taylor (1987). When the limit of detection is estimated from the analyses of blank samples, it is operationally similar to an analytical or an instrument detection limit (Keith et al., 1983; Taylor, 1987) because the samples are only prepared for analysis and are not subjected to collection or processing.

During the course of the NSS-I, personnel at the processing and analytical laboratories calculated detection limits weekly and reported them to the QA staff at EMSL in Las Vegas. These limits were based on the analyses of either laboratory blank samples or detection limit quality control check samples (Table 2). At the processing laboratory, detection limits were calculated for total monomeric and non-exchangeable monomeric aluminum and for closed-system DIC. The detection limits reported by the laboratories met the requirement that they be less than the detection limit objectives (Table 13). For this report method-level limits of detection were calculated to confirm the reported detection limits. The results are presented in the following sections.

Limits of Detection Based on Laboratory Blank Samples--

Laboratory 1 and Laboratory 2 analyzed 26 and 42 laboratory blank samples, respectively. Laboratory blanks were not used to calculate limits of detection for ANC, BNC, or pH measurements. Laboratory blank measurement data for ANC and BNC were not reported on the standard reporting form for QC samples in the analytical data package (Form 20) but were included in the titration data files for ANC and BNC. The titration data are not included in the verified data set. The processing laboratory analyzed 68 laboratory blanks for monomeric aluminum, 66 for nonexchangeable monomeric aluminum, 58 for closed-system DIC, and 46 for specific conductance. During the first half of the NSS-I, specific conductance measurements were made on laboratory blank samples for only seven batches. On several occasions, closed-system DIC measurements for more than one batch were made using a single calibration. Therefore, one laboratory blank may have been used in the measurement of more than one batch of samples.

Summary statistics and limits of detection based on laboratory blank sample measurements are presented in Table 15. For measurements made at the analytical laboratories, values qualified with an X flag (Appendix B) were excluded from detection limit calculations. Measured values from the processing laboratory that were identified as statistical outliers (Grubbs' test, $p \leq 0.05$; Grubbs, 1969) were excluded from the detection limit calculations. Laboratory blank measurements made in the processing laboratory for total monomeric aluminum, nonexchangeable monomeric aluminum, specific conductance, and closed-system DIC are not included in the NSS-I data base and thus were not subject to the same verification procedures as analytical laboratory blanks.

Laboratory 1--Limits of detection estimated from laboratory blank samples were within the detection limit objective for all variables (Table 15). Mean values for all variables were at or very near to zero (Table 15). All values for chloride, specific conductance, DIC, sodium, and nitrate were reported as zero, indicating that a low-level QCCS may be more suitable than a laboratory blank sample to calculate an instrumental detection limit for these variables.

Laboratory 2--Limits of detection were less than or near to the detection limit objective for all variables except total aluminum (0.017 mg/L). In addition, the limit of detection for fluoride was estimated as 0.010 mg/L, because laboratory personnel did not calibrate the instrument to measure concentrations less than 0.010 mg/L (see Section 5). Mean values of laboratory blank measurements were at or near zero for all variables except total aluminum (0.010 mg/L) and silica (0.04 mg/L), indicating the possibility of sporadic reagent contamination or calibration bias.

During data verification, the QA staff set control limits for the reported values of laboratory blank measurements. The lower control limit was established as the negative value and the upper control limit as twice the value of the detection limit objective for each variable as required by the contracts with the analytical laboratories. The lower control limit allowed for minor fluctuations in instrument performance. Of the 42 total aluminum measurements, 15 were greater than twice the detection limit objective, which resulted in a large mean value and a large standard deviation.

For silica measurements, only two values were greater than twice the detection limit objective. Altogether the distribution of silica measurements included four values (both positive and negative) that were statistical outliers (Grubbs' test, $p \leq 0.05$; Grubbs, 1969). It appears that silica measurements in a small number of batches may have been affected by low-level reagent contamination or a negative calibration bias.

Processing Laboratory--The limit of detection for closed-system DIC measurements (0.03 mg/L) was less than the detection limit objective, and the mean value (0.02 mg/L) was near zero (Table 15). For total monomeric

Table 15. Estimates of Limits of Detection Based on Analyses of Laboratory Blank Samples, National Steam Survey - Phase I

Variable	Units	Detection limit objective	Analytical laboratories[a]							
			Laboratory 1				Laboratory 2			
			n	Mean	s[b]	Estimated limit of detection[c]	n	Mean	s[b]	Estimated limit of detection[c]
Al-ext	mg/L	0.005	26	<0.001	--	<0.001	42	0.001	0.0007	0.002
Al-total	mg/L	0.005	26	0.002	0.0017	0.005	42	0.010	0.0057	0.017
Ca	mg/L	0.01	26	<0.01	--	0.01	42	<0.01	--	0.02
Cl^-	mg/L	0.01	26	0.00[d]	0.00	0.00[d]	42	<0.01	--	0.01
Cond-lab	µS/cm $\bar{X}$	<0.9	26	0.00[d]	0.00	--[e]	42	0.6	0.20	--[e]
DICf	mg/L	0.05	26	0.00[d]	0.000	0.00[d]	42	<0.01	--	0.07
DOC	mg/L	0.1	26	<0.1	--	<0.1	42	<0.1	--	0.1
F^-	mg/L	0.005	26	0.002	0.0007	0.002	42	0.000[g]	0.0000	0.010[g]
Fe	mg/L	0.01	26	<0.01	--	0.003	42	<0.01	--	<0.01
K	mg/L	0.01	26	<0.01	--	<0.01	42	<0.01	--	0.02
Mg	mg/L	0.01	26	<0.01	--	<0.01	42	<0.01	--	0.01
Mn	mg/L	0.01	26	<0.01	--	<0.01	42	<0.01	--	<0.01
Na	mg/L	0.01	26	0.00[d]	0.000	0.00[d]	42	<0.01	--	0.01
NH_4+	mg/L	0.01	26	<0.01	--	<0.01	42	0.00	0.001	0.003
NO_3^-	mg/L	0.005	26	0.000[d]	0.000	0.000[d]	42	<0.001	--	0.009
P	mg/L	0.002	26	<0.001	--	0.001	42	<0.01	--	0.001
SiO_2	mg/L	0.05	26	<0.01	--	<0.01	38	0.04	0.022	0.07
SO_4^{2-}	mg/L	0.05	26	<0.01	--	<0.01	42	<0.01	--	0.04

			Processing laboratory[a]			
			n	Mean	s[b]	Estimated limit of detection[c]
Al-mono	mg/L	0.010	65	0.009	0.0047	0.012
Al-nex	mg/L	0.010	65	0.014	0.0064	0.019
Cond-PL[h]	µS/cm $\bar{X}$	<0.9	7	5.6	1.48	--[e]
Cond-PL[i]	µS/cm $\bar{X}$	<0.9	39	0.6	0.27	--[e]
DIC-closed	mg/L	0.05	58	0.02	0.011	0.03

[a] n = number of samples; s = standard deviation.
[b] Dashes indicate that the standard deviation is nearly zero.
[c] Estimated limit of detection = 3s.
[d] All measurements reported as zero by laboratory.
[e] Detection limit objective expressed as the mean of the blank sample measurements
[f] DIC detection limits apply to both equilibrated and initial measurements.
[g] Laboratory reported all concentrations less than 0.010 mg/L as zero. Limit of detection set at 0.010.
[h] Values for blanks measured in the first 28 batches.
[i] Values for blanks measured in the remaining 40 batches.

aluminum, three statistical outliers (Grubbs' test, $p \leq 0.05$; Grubbs, 1969) were not included in assessing method-level detectability. All values except for the outliers were less than twice the detection limit objectives. The limit of detection for total monomeric aluminum (0.012 mg/L) measurements was very near the detection limit objective (Table 15). The mean value for total monomeric aluminum measurements (0.009 mg/L) indicates a possibility of reagent contamination or a calibration bias.

For the nonexchangeable monomeric aluminum measurements, one statistical outlier was not included in assessing method-level detectability. Fifteen values were greater than twice the detection limit objective. The limit of detection for nonexchangeable monomeric aluminum (0.019 mg/L) was less than twice the detection limit objective (Table 15). The mean value (0.014 mg/L) suggests that the in-line cation exchange column in one channel of the flow injection system caused a positive calibration bias for this channel.

Because of the malfunctioning specific conductance probe used in the processing laboratory during the first half of the NSS-I (Section 5), detectability was evaluated for each half of the survey. The mean value before the probe was replaced was high (5.6 μS/cm). The mean value after the probe was replaced (0.6 μS/cm) indicated that the problem had been corrected.

Limits of Detection Based on Field Blank Samples--

Field blank samples offered several advantages over laboratory blank samples in assessing method-level detectability. Field blanks were blind samples (except at the processing laboratory) inserted at random into sample batches. The values obtained from measuring field blanks were not subject to control limits at the laboratory, as were laboratory blanks or detection limit quality control check solutions. Field blank samples for the NSS-I were prepared from a single source of reagent water (the processing laboratory) and were thus independent of blank samples used in calibrations (except at the processing laboratory). Finally, field blank measurements were subjected to the same data verification procedures as streamwater samples; this was not always true for laboratory blank measurements.

The limit of detection calculated from measurements of field blank samples should be similar to the instrumental limit of detection calculated from laboratory blank samples. The use of field blank samples to assess limits of detection does have some limitations, but has been recommended for wet precipitation samples (Campbell and Scott, 1985). Conceptually, field blanks are similar to laboratory blanks, except that the sources of variability are different. Field blanks were processed through the entire collection and measurement system of the NSS-I; thus, they were potentially subject to more sources of error than were laboratory blank samples and provide a more representative estimate of a detection limit for the entire collection and measurement system. Unless mean levels of chemical variables measured in field blank samples are substantial, however, the overall variance should not be affected by relationships of variance to concentration; hence, the variance for field blank samples should be similar to the variance expected from a laboratory blank sample or from a low-concentration standard solution.

Before limits of detection were estimated from field blank measurements, all values qualified with an X flag (Appendix B) were eliminated. In addition, all values identified as significant outliers by Grubbs' test ($p \leq 0.05$; Grubbs, 1969) were eliminated. No more than three values were identified as outliers for any variable. Removal of the outliers provided a more representative estimate of variance that was not influenced by occasional cases of possible sample contamination during collection and processing. Summary statistics and limits of detection based on field blank measurements are presented in Table 16.

For ANC and specific conductance, mean values from both analytical laboratories were less than or near the detection limit objectives. For magnesium, potassium, sodium,

Table 16. Estimates of Limits of Detection Based on Analyses of Field Blank Samples, National Stream Survey - Phase I

Variable	Units	Detection limit objective	Analytical laboratories[a]							
			Laboratory 1				Laboratory 2			
			n	Mean	s[b]	Estimated limit of detection[c]	n	Mean	s[b]	Estimated limit of detection[c]
Al-ext	mg/L	0.005	22	0.004	0.0037	0.011	39	0.001	0.0008	0.002
Al-total	mg/L	0.005	21	0.005	0.0053	0.016	39	0.013	0.0063	0.019
ANC	μeq/L	$\lvert\bar{X}\rvert \leq 10$	24	-0.4	0.56	1.7	34	4.4	2.76	8.3
BNC	μeq/L	$\lvert\bar{X}\rvert \leq 10$	24	14.5	2.03	6.1	37	23.3	5.02	15.1
Ca	mg/L	0.01	24	0.01	0.006	0.02	37	0.01	0.004	0.01
Cl^-	mg/L	0.01	23	<0.01	--	0.02	39	0.01	0.009	0.03
Cond-lab	μS/cm	$\bar{X} \leq 0.9$	24	<0.1	--	0.7	39	1.0	0.12	0.4
DIC-eq	mg/L	0.05	24	0.08	0.053	0.16	39	0.13	0.081	0.24
DIC-init	mg/L	0.05	24	0.15	0.048	0.14	37	0.10	0.069	0.27
DOC	mg/L	0.1	24	0.1	0.13	0.4	37	0.2	0.12	0.4
F^-	mg/L	0.005	24	0.003	0.0013	0.004	39	NC[d]	NC[d]	0.010[d]
Fe	mg/L	0.01	24	<0.01	--	0.03	37	<0.01	--	<0.01
K	mg/L	0.01	23	<0.01	--	0.01	36	<0.01	--	0.01
Mg	mg/L	0.01	24	<0.01	--	0.01	38	<0.01	--	<0.01
Mn	mg/L	0.01	23	<0.01	--	0.02	39	<0.01	--	<0.01
Na	mg/L	0.01	24	<0.01	--	0.01	38	<0.01	--	0.01
NH_4+	mg/L	0.01	24	0.01	0.010	0.03	39	<0.01	--	0.01
NO_3^-	mg/L	0.005	23	0.010	0.012	0.036	34	0.004	0.0048	0.014
P	mg/L	0.002	24	<0.001	--	0.011	36	<0.001	--	0.003
SiO_2	mg/L	0.05	24	0.02	0.027	0.08	38	0.03	0.045	0.14
SO_4^{2-}	mg/L	0.05	24	0.01	0.007	0.02	39	0.02	0.016	0.05

Variable	Units	Detection limit objective	Processing laboratory[a]			
			n	Mean	s[b]	Estimated limit of detection[c]
Al-mono	mg/L	0.010	60	0.010	0.0031	0.009
Al-nex	mg/L	0.010	61	0.014	0.0061	0.018
Cond-PL[e]	μS/cm	$\bar{X} \leq 0.9$	46	1.3	0.27	0.8
DIC-closed	mg/L	0.05	58[f]	0.02[f]	0.011[f]	0.03[f]

[a] n = number of samples; s = standard deviation.

[b] Dashes indicate that the standard deviation is nearly zero.

[c] Limit of detection = 3s.

[d] NC = Not calculated. Laboratory reported all values less than 0.010 mg/L as 0 mg/L. Limit of detection estimated as 0.010 mg/L.

[e] Laboratory blank data indicate measurements from batches 2100 through 2127 may be inaccurate as a result of a faulty probe. These data were not included in this estimate.

[f] Field blanks were not measured. Calibration blank measurements were used to estimate the limit of detection.

and sulfate, limit of detection estimates from both analytical laboratories were near or less than the detection limit objectives. The limit of detection for total monomeric aluminum analyses at the processing laboratory was less than the detection limit objective.

From both analytical laboratories, limits of detection were near or less than twice the detection limit objective for extractable aluminum (0.011 and 0.002 mg/L), calcium (0.02 and 0.01 mg/L), fluoride (0.004 and 0.010 mg/L), and manganese (0.01 and <0.01 mg/L) (Table 16). The limit of detection for nonexchangeable monomeric aluminum (0.018 mg/L) was near the detection limit based on laboratory blank samples (Table 15). Measurements of these variables are apparently subject to sources of variability from sample collection and processing, but the field blank measurements for these variables do not indicate a data quality problem related to method-level detectability. Mean values for all of these variables were near zero at all laboratories.

For five variables, limit of detection estimates from both analytical laboratories were greater than twice the detection limit objective. These variables were total aluminum (0.016 mg/L and 0.019 mg/L), equilibrated DIC (0.16 and 0.24 mg/L), initial DIC (0.14 mg/L and 0.27 mg/L), DOC (0.4 and 0.4 mg/L), and nitrate (0.036 and 0.014 mg/L). The detection limits estimated for DIC measurements are derived from a different sample matrix (analyte-free water versus a natural water sample), but provide an indication of the amount of change expected in samples undersaturated with dissolved carbon dioxide. Measurements of DOC appear to have been affected by sporadic low-level contamination during collection or processing, as 5 values from Laboratory 1 and 31 values from Laboratory 2 were greater than twice the detection limit objective. Addition of nitrate to samples also appears to have occurred, especially during the early stages of the NSS-I, as shown by the mean (0.010 mg/L) and limit of detection estimate (0.036 mg/L) from Laboratory 1 (Table 16). This laboratory did not analyze samples in the latter part of the survey. The most likely source of the addition appears to have been the processing laboratory (see Section 5). The additional precautions taken at the processing laboratory appear to have lowered the levels of nitrate in field blank samples during the latter part of the survey as shown by the mean (0.004 mg/L) and limit of detection (0.014 mg/L) from Laboratory 2 (Table 16).

For total aluminum analyses, examination of the measured field blank values for total aluminum from Laboratory 1 did not indicate a data quality problem. Only one value was greater than twice the detection limit objective. At Laboratory 2, 19 values were greater than twice the detection limit objective. Variability among batches that occurred during the digestion procedure and sporadic additions of low concentrations to field blank samples in the field or at the processing laboratory would increase the standard deviation of the blank measurements and thus the limit of detection estimate at both laboratories. However, for Laboratory 2 the analyses of laboratory blank samples (Table 15) suggest that total aluminum measurements may have been affected by reagent contamination.

For Laboratory 1, limits of detection for iron (0.03 mg/L), manganese (0.02 mg/L), ammonium (0.03 mg/L), and phosphorus (0.011 mg/L) were equal to or greater than twice the detection limit objective (Table 16). For iron, manganese,and ammonium, all measurements (except one for ammonium) were within the control limits established for laboratory blanks. No data quality problems are indicated for these variables. For phosphorus, 12 measurements were outside the control limits for laboratory blanks (four were greater than twice the detection limit objective and eight were less than the negative value of the detection limit objective). It is possible that, for a few batches, the phosphorus measurements at Laboratory 1 may be affected by a very low-level negative calibration bias.

For Laboratory 2, limits of detection for chloride (0.03 mg/L) and silica (0.14 mg/L) were greater than twice the detection limit objective (Table 16). In addition, the mean value for BNC measurements (23.3 μeq/L) was more than twice the detection limit objective.

Examination of chloride measurements showed only one value greater than twice the detection limit objective, and no data quality problem is indicated. For silica, two measurements were greater than twice the detection limit objective. However, the mean value (0.03 mg/L, Table 16) indicated contamination caused by a sporadic addition of silica to blank samples. This addition probably occurred at Laboratory 2, based on the analysis of laboratory blank samples (Table 15). For BNC measurements, the BNC of a field blank sample should be due totally to dissolved carbon dioxide. A high background level and considerable variability in BNC can be expected if the sample is not protected from the atmosphere during the base titration. The high variability of DIC in field blank samples at this laboratory (Table 16) would result in a high variability in BNC.

Assessment of System-Level Detectability (Background)

Background quantities of chemical variables were assessed by examining the values of field blank measurements pooled across both analytical laboratories. Values qualified with an X flag (Appendix B) were not included in the assessment, but statistical outliers were not removed as they were for the estimates of detection limits. The two statistics of interest in assessing background are the mean (or median) and the system decision limit. The mean (or median) can provide an average estimate of the amount of background contamination added during collection and processing. The system decision limit represents the lowest measured value of a chemical variable that is distinguishable from field blank measurements at a specified level of confidence. It is a critical value when testing the null hypothesis that a single measured value is not greater than the average of field blank measurements. System decision limits should not be confused with detection limits.

System decision limits were calculated from measurements of field blank samples based on both parametric and nonparametric statistics. For many variables, distributions of field blank measurements may be non-normal, and the use of nonparametric statistics provides a more representative estimate of background that is less sensitive to outlying measurements. A parametric system decision limit (SDL_p) was calculated as follows:

$$SDL_p = \bar{x}_b + 1.65s$$

where $\bar{x}_b$ is the mean of field blank measurements and s is the standard deviation of field blank measurements. The constant 1.65 represents the number of standard deviations from the mean of blank samples within which approximately 95 percent of the measurements would be expected to lie if they belonged to a normal distribution. A nonparametric system decision limit (SDL_{np}) was calculated using the approach of Permutt and Pollack (1986):

$$SDL_{np} = P_{95}$$

where P_{95} is the the 95th percentile of the field blank measurements.

Summary statistics and system decision limits are presented in Table 17. For chemical variables whose field blank measurements are distributed more or less normally, parametric and nonparametric decision limits are approximately equal. For almost all variables, parametric and nonparametric system-level decision limits were nearly equal (Table 17). Data from streamwater samples that contain chemical variables in quantities less than the system-level decision limit should be compared and interpreted cautiously, because the source of the variablility is confounded between what was present in the stream at the time of collection and what may have been added as background during collection and processing. System-level decision limits were not calculated for closed-system DIC or closed-system pH because field blank samples were not measured for these variables.

For nearly all variables the system decision limits did not indicate serious data quality problems related to detection of the analyte or to background levels. The percentage of routine samples collected during the NSS-I with measured concentrations for a variable that were below the system decision limit is

Table 17. Estimates of System Decision Limits Based on Analyses of Field Blank Samples Pooled Across Laboratories, National Stream Survey - Phase I

		Parametric				Nonparametric		
Variable	Units	n	Mean	s[a]	System decision limit (SDL_p)[b]	n	Median	System decision limit (SDL_{np})[c]
Al-ext	mg/L	61	0.002	0.0028	0.007	61	0.001	0.007
Al-total	mg/L	61	0.011	0.0081	0.024	61	0.010	0.027
Al-mono	mg/L	61	0.010	0.0038	0.016	61	0.010	0.015
Al-nex	mg/L	61	0.014	0.0061	0.024	61	0.015	0.023
ANC	μeq/L	58	1.8	4.10	8.6	58	1.5	8.7
BNC	μeq/L	63	20.8	7.99	34.0	63	19.3	34.9
Ca	mg/L	62	0.01	0.006	0.020	62	<0.01	0.02
Cl^-	mg/L	63	0.01	0.010	0.026	63	<0.01	0.03
Cond-PL[d]	μS/cm	26	3.5	2.59	7.8	26	2.0	8.0
Cond-PL[e]	μS/cm	33	1.5	1.08	3.2	33	1.2	2.2
Cond-lab	μS/cm	63	0.6	0.49	1.4	63	0.9	1.2
DIC-eq	mg/L	63	0.11	0.075	0.236	63	0.10	0.23
DIC-init	mg/L	63	0.13	0.085	0.276	63	0.13	0.23
DOC	mg/L	63	0.2	0.25	0.6	63	0.2	0.5
F^-	mg/L	63	NC	NC	0.010[f]	63	NC	0.010[f]
Fe	mg/L	63	<0.01	--	0.01	63	<0.01	0.02
K	mg/L	62	<0.01	--	0.01	62	<0.01	0.02
Mg	mg/L	62	<0.01	--	<0.01	62	<0.01	<0.01
Mn	mg/L	63	<0.01	--	0.010	63	<0.01	0.012
Na	mg/L	62	<0.01	--	0.01	62	<0.01	0.012
NH_4^+	mg/L	63	<0.01	0.008	0.02	63	0.01	0.32
NO_3^-	mg/L	62	0.008	0.0112	0.026	62	0.005	0.02
P	mg/L	63	0.001	0.0029	0.006	63	<0.001	0.006
pH-ANC	pH units	58	5.73	0.155	NC	58	5.69	NC
pH-BNC	pH units	58	5.77	0.131	NC	58	5.74	NC
pH-eq	pH units	63	5.96	0.233	NC	63	5.89	NC
SiO_2	mg/L	63	0.03	0.043	0.10	63	0.02	0.09
SO_4^{2-}	mg/L	63	0.01	0.014	0.033	63	0.01	0.05
True color	PCU	63	6	3.7	12.1	63	5	10
Turbidity	NTU	63	0.3	1.36	2.5	63	0.1	0.2

[a] Dashes indicate that the standard deviation is nearly zero.
[b] SDL_p = mean + 1.65s.
[c] SDL_{np} = 95th percentile of distribution of field blank measurements.
[d] Measurements for first half of survey.
[e] Measurements for second half of survey. NC = Not calculated.
[f] Laboratory 2 reported all measured concentrations less than 0.010 mg/L as zero. System decision limit estimated as 0.010 mg/L.

presented in Figure 12, and the ranges of measured values are presented in Table 18. In some cases, the extreme values (e.g., extremely high specific conductance or extremely low ANC) represent rare cases of extreme conditions that are not representative of the streams of interest. The data shown in Figure 12 indicate that for total aluminum, silica, and fluoride, 90 to 100 percent of the routine samples had measured concentrations that were greater than the respective system decision limits. Thus, the potential problems identified for these three variables will not have a large effect on the interpretation of routine sample data.

Discussion and Summary: Detectability

For most variables, method dectection limits and system decision limits indicated no serious problems with data quality that were related to either instrument performance, methodological performance, or background levels of analytes. Background levels of total aluminum and silica (SDL_{np} = 0.027 mg/L and 0.09 mg/L, Table 17) should not confound data interpretation as nearly all routine samples had measured concentrations above the system decision limit (Figure 12). The error in fluoride measurements at Laboratory 2 (not measuring values less than 0.010 mg/L) also should not confound data interpretation, as nearly all routine samples had measured concentrations greater than the system decision limit.

The observed background levels of total monomeric aluminum (SDL_{np} = 0.015 mg/L) and nonexchangeable monomeric aluminum (SDL_{np} = 0.023 mg/L) are the result of instrument variability or calibration bias rather than contamination. The usefulness of these data may be limited at low concentrations, especially since the variable of interest, exchangeable (or labile) monomeric aluminum, is calculated as the difference between the two measured fractions. As a result of low concentrations, the difference calculated for exchangeable monomeric aluminum sometimes results in negative values.

The background level of BNC (SDL_{np} = 39.14 μeq/L) suggests that there can be considerable variability in dilute streamwater samples with low BNC. The BNC measurement protocol for the NSS-I was designed primarily to assist in verifying the measurements of ANC on a sample-by-sample basis (Hillman et al., 1987). The usefulness of routine stream sample data to determine the presence of weak versus strong acids may be limited, because the samples were not protected from atmospheric carbon dioxide during collection, handling, or titration.

Specific conductance measurements at the processing laboratory were affected by a faulty probe for the first 28 sample batches, and thus we recommend that the analytical laboratory measurement of specific conductance be used for data interpretation. The primary purpose of the processing laboratory measurement of specific conductance was to check on the stability of streamwater samples between the time of collection and the time of processing.

Certain other types of data interpretation activities, not related to acidification or stream classification, may be limited by the background levels introduced into NSS-I samples. For example, examination of nutrient relationships or productivity may be limited by background levels of nitrate and phosphorus. The measurement program for blank samples that was used during the NSS-I was designed to control calibration biases and background contamination, rather than to correct routine sample measurements at trace concentrations (see Taylor, 1984 and 1987). Conceivably, the data from field or laboratory blank measurements presented in this report could be used to develop a correction factor by using the equations presented by Taylor (1984, 1987). However, the uncertainty associated with the blank measurements will be conservative because it will be based on an among-batch rather than a within-batch estimate of measurement variability.

The concept of detection limits may need to be more clearly defined for future AERP

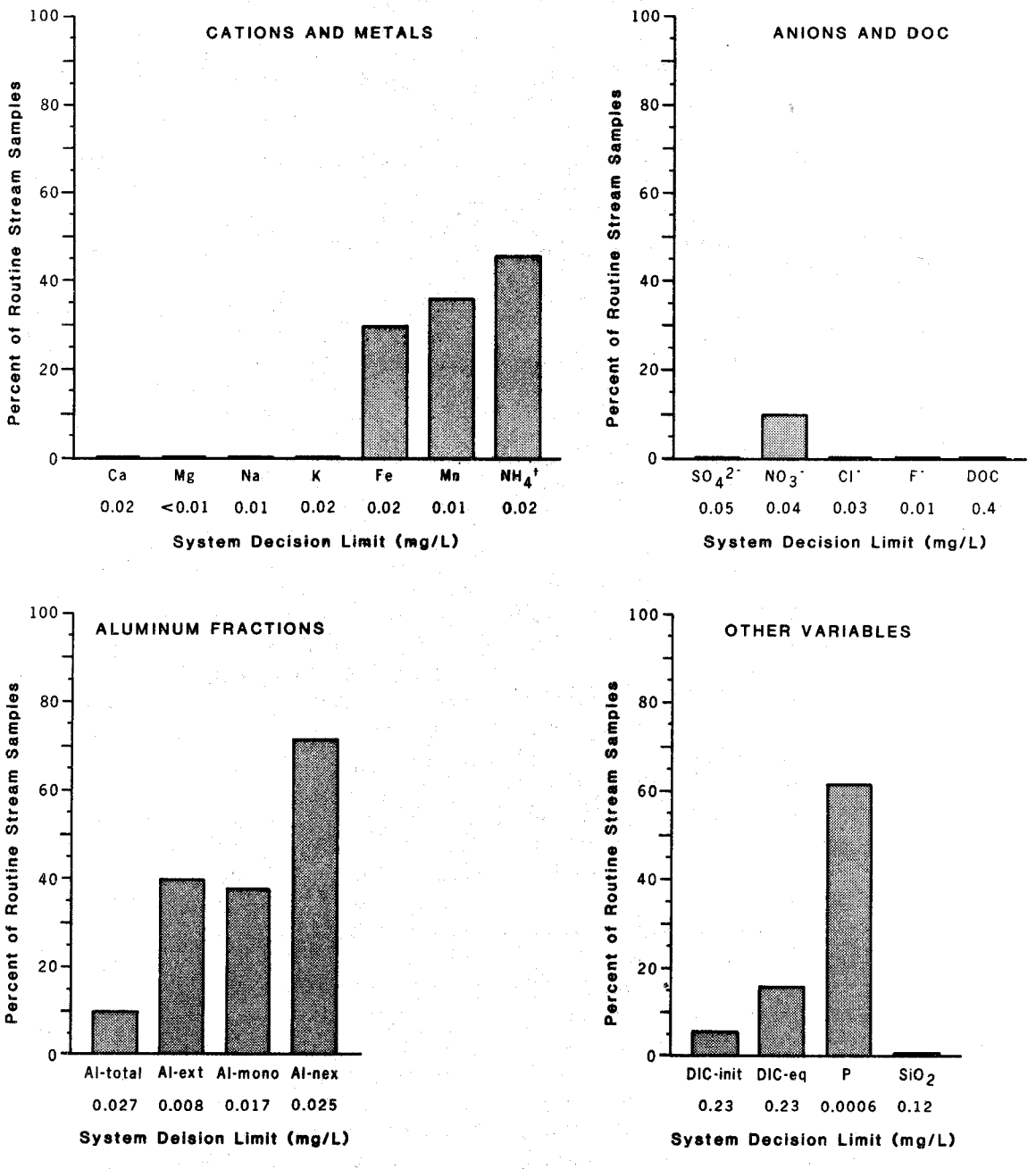


Figure 12. Percentage of NSS-I routine samples with measured concentrations belw the system decison limit.

projects and data quality objectives may need to be modified as appropriate for the particular research questions to be answered. Different procedures for calculating detection limits will yield different estimates. If researchers require information regarding the lowest limit of reliable measurement, detection limits may need to be defined with statistical error rates in mind, as has been suggested by Clayton et al. (1987). Both laboratory and field blank samples have limitations when they are used to estimate limits of detection. Other samples

Table 18. Range and Central Tendancy Statistics for Analyte Concentrations in Routine Stream Samples, National Stream Survey - Phase I

Variable[a]	Number	Minimum value[b]	Median value	Mean value	Maximum value
Al-ext	1,345	-0.002	0.009	0.080	10.100
Al-total	1,343	-0.006	0.111	0.345	37.100
Al-mono	1,353	0.001	0.017	0.098	12.223
Al-nex	1,349	0.000	0.018	0.026	0.375
ANC (μeq/L)	1,378	-1,750.600	176.550	448.039	7,602.800
BNC (μeq/L)	1,378	-85.400	59.250	95.011	2,421.400
Ca	1,375	0.063	4.850	10.365	96.623
Cl^-	1,375	0.075	2.980	6.725	380.000
Cond-PL (μS/cm)	1,366	10.500	65.700	111.708	1,376.300
Cond-lab (μS/cm)	1,378	10.500	63.750	109.612	1,294.000
DIC-closed	1,378	0.036	3.204	6.261	92.440
DIC-eq	1,378	-0.102	1.878	5.191	71.200
DIC-init	1,378	0.059	2.358	5.720	92.706
DOC	1,374	0.000	1.610	3.338	171.000
F^-	1,375	0.000	0.040	0.050	0.520
Fe	1,376	-0.010	0.041	0.262	32.800
K	1,376	0.001	0.940	1.185	8.842
Mg	1,376	0.098	1.591	3.341	37.345
Mn	1,376	-0.010	0.023	0.136	12.100
Na	1,376	0.129	2.401	4.263	185.000
NH_4^+	1,374	-0.011	0.023	0.051	3.035
NO_3^-	1,373	-0.001	0.842	3.147	70.000
P	1,374	-0.008	0.004	0.016	1.420
pH-closed (pH units)	1,378	3.270	6.840	6.633	9.360
pH-ANC (pH units)	1,378	3.000	6.850	6.631	8.890
pH-BNC (pH units)	1,378	3.000	6.865	6.644	8.920
pH-eq (pH units)	1,378	3.050	7.340	7.077	8.860
SiO_2	1,375	-0.006	5.300	6.506	34.125
SO_4^{2-}	1,375	0.046	8.126	15.610	340.000
True color (PCU)	1,367	0.000	15.000	31.137	900.000
Turbidity (NTU)	1,364	0.030	2.200	5.776	650.000

[a] Concentrations are in mg/L unless otherwise indicated.

[b] Negative values are a result of analytical laboratory instrument calibration bias.

(such as audit samples) may be of more use when assessing laboratory performance in terms of detectability.

Accuracy

Accuracy within the processing laboratory and each analytical laboratory involved in the NSS-I was evaluated by examining the results from analyses of performance audit samples. During the data verification process, audit sample results were compared to acceptance criteria (ranges) calculated from performance audit data from previous NSWS projects. Further detail about the calculation of acceptance criteria is found in Drousé et al. (1987). The control limits for the acceptance criteria are presented in Appendix C. Batches

of samples containing audit sample values that were outside the critiera were qualified with an N flag (Appendix B) for more intensive review.

Accuracy was evaluated primarily on the basis of the results of analyses of the synthetic audit samples. However, other AERP studies that have used the synthetic audit samples (Drousé, 1987; Silverstein et al., 1987) have reported that the concentrations of several variables are affected either by changes in dissolved carbon dioxide concentration between the time of preparation and measurement (e.g., BNC, DIC, and pH) or by concentrations of other variables (e.g., ANC). For these variables, theoretical concentrations are inaccurate or not available. For these selected variables, data from the analysis of natural audit samples are presented in this section to assist in assessing accuracy. The natural audit samples appear to be less sensitive to changes in dissolved carbon dioxide, and their chemical composition is not affected by preparation errors. Data for all variables from measurements of natural audit samples are presented in Appendix A.

The accuracy estimates presented in this section provide an indication of the presence of systematic errors in measurement. An accuracy estimate for an analyte derived from a theoretical value for a synthetic audit sample can be used as an estimate of absolute analytical method bias only if preparation errors are assumed to be negligible. This assumption is not valid for all variables, because measured values from the support laboratory did not always agree with theoretical values (see Appendix A). Thus, accuracy estimates from synthetic audit samples are probably conservative. Accuracy estimates for variables not having defined theoretical concentrations in the synthetic audit samples, or estimates based on natural audit samples, only provide an estimate of relative bias, because the index value (the value measured by the analytical laboratories) represents a measured value obtained by a specific methodology. In addition, the audit sample concentrations do not bracket the range of sample concentrations or many variables. For these reasons, accuracy estimates should not be used as quantitative estimates of systematic measurement errors. Estimates of among-batch precision presented later in this report include the effects of both systematic and random measurement errors and thus can provide estimates of measurement uncertainty, subject to the limitations mentioned above.

The following equation provides an estimate of percent accuracy for each laboratory for all analytes except pH:

$$\text{Percent accuracy} = [(\bar{x} - R) \div R]100$$

where $\bar{x}$ is the mean of measured values and R is the theoretical value or an index value based on measurements from one or more laboratories. For pH, accuracy was expressed as the difference between the mean measured value and the theoretical or index value.

For the synthetic audit samples, index values were developed for all analytes based on verification measurements made at the support laboratory immediately after preparation of a sample lot. Six replicate measurements were made for each variable for each lot. All replicates for each lot were measured in a single batch, providing an estimate of within-batch variability. Index values for ANC, BNC, DIC, and pH are presented in this section, while values calculated for all variables are presented in Appendix A. For all variables except BNC, standard solutions at two concentrations were obtained from the EPA Environmental Monitoring Systems Laboratory in Cincinnati and were analyzed with each batch of verification samples. The measurement of these standards served to validate the analytical measurements made at the support laboratory. Data from these standards were obtained for six different batches of synthetic audit verification samples to provide an estimate of the among-batch variability of the support laboratory measurements. These data, based on measurements of the EPA standard solutions, are presented in Table 19 for ANC, pH, and DIC. Data for these standards for all other variables except BNC are presented in Appendix A.

Table 19. **Summary Statistics for Selected Variables for EPA Reference Samples Measured at the Support Laboratory, National Stream Survey - Phase I**

Variable	True value of standard	Number	$\bar{x}$	s	$\bar{x}$ - True value	Accuracy percent	Precision (%RSD)
ANC (μeq/L)	68.8	5	67.1	0.77	-1.7	-2.5	1.1
DIC (mg/L)	1.13	6	1.20	0.048	0.07	6.2	4.0
pH (pH units)	7.8	6	7.76	0.042	-0.04	-	-

For ANC, the within-batch standard deviation estimates for Lot 14 (2.90 μeq/L) and Lot 15 (2.93 μeq/L), presented in Appendix A, were greater than the among-batch standard deviations estimated from the standard solutions (0.77 μeq/L; Table 19). For BNC, within-batch standard deviations (Appendix A) for Lot 14 (4.37 μeq/L) and Lot 15 (3.85 μeq/L) were the only estimates available. Index values for ANC and BNC were estimated as the mean measured values, with the 95 percent confidence intervals calculated by using the within-batch standard deviation estimate. Index values for DIC and pH also were estimated as the mean measured values from the support laboratory, but the 95 percent confidence intervals were calculated from the among-batch standard deviations estimated from the EPA standards (Table 19).

Index values for the natural audit samples were developed from measurements made at different laboratories. The Bagley Lake audit sample used during the NSS-I was also analyzed at two other laboratories during the Western Lake Survey-Phase I (Silverstein et al., 1987) in late 1985. The Big Moose Lake audit sample was analyzed by two other laboratories during Phase II of the Eastern Lake Survey, which was conducted during and after the NSS-I. Data from all these laboratories (including the two laboratories involved in the NSS-I) were used to estimate index values. Index values were calculated as a grand mean, based on weighted mean values from each laboratory, by the following equation (Taylor, 1987):

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{L} w_i \bar{x}_i}{\sum_{i=1}^{L} w_i}$$

$$s = \frac{1}{\sum_{i=1}^{L} w_i}$$

where $w_i = \frac{n_i}{s_i^2}$

$\bar{\bar{x}}$ = grand mean
L = number of laboratories
w_i = weighting factor for laboratory i
s = standard deviation of weighted means
$\bar{x}_i$ = mean value from laboratory i
s_i^2 = variance from laboratory i
n_i = number of measurements from laboratory i

Ninety-five percent confidence intervals about the index value were also calculated. Theoretical and index values for synthetic and natural audit samples are presented in Table 20.

Accuracy estimates were calculated for each laboratory and for each lot of synthetic audit samples. Synthetic audit samples from Lot 14 were used to evaluate accuracy during the initial part of the NSS-I, while samples from Lot 15 were used during the latter part.

For those analytes that were measured at the analytical laboratories, systematic

Table 20. Theoretical and Index Values for Analyses of Synthetic and Natural Audit Samples, National Stream Survey - Phase I

Theoretical values[a] of synthetic audit samples

Variable	Units	Theoretical value	Variable	Units	Theoretical value
Al[b]	mg/L	0.020	Mg	mg/L	0.447
Ca	mg/L	0.194	Mn	mg/L	0.098
Cl^-	mg/L	0.343	Na	mg/L	2.75
Cond-lab	µS/cm	17.5	NH_4^+	mg/L	0.168
DIC[c,d]	mg/L	0.959[d]	NO_3^-	mg/L	0.467
DOC	mg/L	1.0	P-	mg/L	0.0273
F-	mg/L	0.042	SiO_2	mg/L	1.070
Fe	mg/L	0.059	SO_4^{2-}	mg/L	2.280
K	mg/L	0.203			

Index values

	Synthetic audit samples[e]							Natural audit samples[f]					
	Lot 14			Lot 15				Bagley Lake			Big Moose Lake		
Variable	n	Mean	CI[g]	n	Mean	CI[g]	Variable	n	Grand mean[h]	CI[g]	n	Grand mean[h]	CI[g]
ANC	6	101.7	3.04	6	109.2	3.07	Al-ext	--	--	--	4	0.197	0.0052
BNC	6	32.0	4.58	6	22.8	4.04	Al-total	4	0.016	0.0006	4	0.272	0.0045
DIC[c]	6	1.14	0.050	6	1.26	0.050	Al-mono	--	--	--	2	0.193	0.0174
pH[i]	6	7.22	0.044	6	7.29	0.044	Al-nex	--	--	--	2	0.053	0.0178
							ANC	4	121.0	0.16	4	-3.1	0.45
							BNC	4	29.7	0.52	4	72.9	1.36
							DIC-closed	2	1.67	0.068	2	0.55	0.061
							DIC-eq	4	1.52	0.015	4	0.11	0.009
							DIC-initial	4	1.52	0.012	4	0.36	0.009
							Fe	--	--	--	4	0.05	0.002
							pH-ANC	4	7.06	(0.014)	4	5.10	0.010
							pH-BNC	4	7.06	(0.014)	4	5.15	0.013
							pH-eq	4	7.29	(0.017)	4	5.17	0.006
							pH-closed	2	7.04	(0.066)	2	5.14	0.060

[a] Assuming no preparation error or external effect.
[b] Value applicable to all aluminum fractions.
[c] Value applicable to all DIC measurements.
[d] Value does not include carbon dioxide added during air equilibration procedure.
[e] Index values based on support laboratory measurements.
[f] Index values based on analytical laboratory measurements from NSWS programs, including NSS-I.
[g] One-sided 95% confidence interval.
[h] Grand mean calculated from weighted means from four analytical laboratories or from two processing laboratories.
[i] Value applicable to all pH measurements.

errors may have been introduced at the processing laboratory during the preparation and processing of the several aliquots from bulk streamwater samples (or syringe samples in the case of extractable aluminum). These errors could result from contamination, improper filtration or preservation, or changes in the sample composition between the time of collection and the time of processing.

For synthetic audit samples, results for each laboratory of field audit and laboratory audit sample analyses were compared by using a single-classification analysis of variance. For all variables (except total aluminum, extractable aluminum, and iron), mean values for field audit samples were not significantly different ($p \leq 0.05$) from corresponding laboratory audit samples. For total aluminum, extractable aluminum, and iron, the mean concentrations in field audit samples were substantially lower than in laboratory audit samples at both laboratories. Loss of total aluminum and iron from field audit samples may result from adsorption of these species onto container surfaces. Sample filtration at the processing laboratory also may remove iron as well as the aluminum that would otherwise transfer to the MIBK extract and be measured as total extractable aluminum, especially if iron or aluminum-containing precipitates have formed (Silverstein et al., 1987; Drousé, 1987). Therefore, for these three variables, only laboratory audit samples were used to estimate accuracy. For all other variables, measured values of field and laboratory audit samples were combined for each laboratory when both types of samples were measured. Laboratory 1 did not measure any field audit samples from Lot 15; accuracy estimates for all variables for this lot are based on laboratory audit samples only. For each variable, outlying values were identified by using Grubbs' test ($p = 0.005$; Grubbs, 1969) and were not included in the analyses. No more than one outlier was identified and removed for any single variable. Removing outliers served to improve the precision estimates for the values and thus the confidence in the estimated mean value. Removal of outliers did not necessarily improve the estimate of percent accuracy.

Percent Accuracy Estimates for Laboratory 1

Summary statistics and percent accuracy estimates for synthetic and natural audit samples measured at Laboratory 1 are presented in Tables 21 and 22. For the synthetic audit samples, percent accuracy estimates for eleven variables for which theoretical values were available were within or near the data quality objective for both lots: chloride, DOC, fluoride, potassium, magnesium, manganese, sodium,ammonium, nitrate, silica, and sulfate. The observed percent accuracy for calcium for Lot 14 (16 percent) represents an apparent bias of 0.03 mg/L, is barely significant at the 95 percent level of confidence, and does not indicate a data quality problem.

For specific conductance, percent accuracy estimates for both lots were less than 10 percent (Table 21) and represent an apparent negative bias of less than 2 μS/cm. Conductance measurements in the natural audit samples from Laboratory 1 also indicated the potential for negative bias, compared to index values (see Appendix A).

Percent accuracy estimates for four variables for which theoretical values were available were well outside data quality objectives for one or both lots: extractable aluminum, total aluminum, iron, and phosphorus. For the synthetic audit samples, mean values for extractable and total aluminum were not significantly different from theoretical values because of the large variability (Table 21). Data from the Bagley Lake and Big Moose Lake samples (Table 22), for these two variables were not significantly different from the index values. There is no evidence for systematic error in extractable aluminum and total aluminum measurements.

Accuracy estimates for iron based on the synthetic audit sample may be unreliable because of a change in the sample between the time of sample preparation and the time that preserved aliquots were prepared at the support laboratory. Mean values for both lots (0.03 mg/L, Table 21) were in agreement with verification values from the support laboratory

Table 21. Percent Accuracy Estimates for Laboratory 1 Measurments of Synthetic Audit Samples, National Stream Survey - Phase I

Variable	Units	Accuracy objective (%)	Theoretical value[b]	Lot 14[a] n	Lot 14[a] Mean	Lot 14[a] CI	Lot 14[a] Percent accuracy	Lot 15[a] n	Lot 15[a] Mean	Lot 15[a] CI	Lot 15[a] Percent accuracy
Al-ext	mg/L	10	0.020	4	0.027	0.0254	35 ns	3	0.015	0.0086	-25 ns
Al-total	mg/L	10	0.020	5	0.018	0.0063	-10 ns	3	0.029	0.0226	-21 ns
Ca	mg/L	10	0.19	14	0.22	0.026	16*	4	0.20	0.047	5
Cl^-	mg/L	10	0.34	13	0.33	0.006	-3*	4	0.33	0.012	-3
Cond-lab	µS/cm	5	17.5	14	16.3	0.18	-7*	4	16.0	0.20	-9*
DOC	mg/L	10	1.0	13	1.0	0.06	0	4	1.0	0.42	0
F^-	mg/L	10	0.042	13	0.040	0.0013	-5*	4	0.039	0.0043	-7
Fe	mg/L	10	0.06	9	0.03	0.011	-50*	4	0.03	0.041	-50 ns
K	mg/L	10	0.20	13	0.20	0.001	0	4	0.19	0.004	-5*
Mg	mg/L	10	0.45	14	0.43	0.002	-4	4	0.43	0.008	-4
Mn	mg/L	10	0.10	14	0.09	0.004	-10*	4	0.09	0.007	-10*
Na	mg/L	10	2.75	14	2.82	0.023	2*	4	2.78	0.036	1
NH_4^+	mg/L	10	0.17	13	0.17	0.003	0	4	0.16	0.011	-6
NO_3^-	mg/L	10	0.467	13	0.477	0.0243	2	3	0.487	0.0027	4*
P	mg/L	10	0.027	13	0.021	0.0025	-22*	3	0.015	0.0087	-44*
SiO_2	mg/L	10	1.07	14	1.20	0.015	12 ns	4	1.19	0.036	10 ns
SO_4^{2-}	mg/L	10	2.28	14	2.21	0.029	-3	4	2.25	0.028	-1

	Accuracy objective (%)	Lot 14[a] Index value(CI)[c]	Lot 14[a] n	Lot 14[a] Mean	Lot 14[a] CI	Lot 14[a] Percent accuracy	Lot 15[a] Index value(CI)[c]	Lot 15[a] n	Lot 15[a] Mean	Lot 15[a] CI	Lot 15[a] Percent accuracy
ANC	10	101.7 (3.04)	14	104.7	1.56	3	109.2 (3.07)	3	104.2	1.54	-5*
BNC	10	32.0 (4.58)	13	22.6	2.10	-29*	22.8 (4.04)	4	18.8	2.42	-18 ns
DIC-eq	10	1.14 (0.050)	14	1.33	0.049	17*	1.26 (0.050)	4	1.23	0.081	-2
DIC-init	10	1.14 (0.050)	14	1.41	0.030	24*	1.26 (0.050)	4	1.39	0.099	10
pH-ANC	0.1[d]	7.22 (0.044)	13	7.04	0.035	-0.18 *,e	7.29 (0.044)	4	7.06	0.083	-0.23 *,e
pH-BNC	0.1[d]	7.22 (0.044)	13	7.07	0.035	-0.15 *,e	7.29 (0.044)	4	7.08	0.078	-0.21 *,e
pH-eq	0.1[d]	7.22 (0.044)	14	7.17	0.159	-0.05[e]	7.29 (0.044)	4	7.31	0.166	0.02[e]

[a] n = number of measurements.
CI = one-sided 95 percent confidence intervals.
ns = not significantly different from the theoretical or index value at $p = 0.05$.
* = significantly different from the theoretical or index value at $p \leq 0.05$.

[b] The theoretical value is the expected value of the synthetic audit sample assuming no preparation error and no external effects.

[c] Index value is based on the mean values from the support laboratory measurements.

[d] Objective expressed in pH units.

[e] Accuracy expressed as the difference between the index value and the mean analytical laboratory value.

Table 22. Percent Accuracy Estimates for Laboratory 1 Measurements of Selected Variables in Natural Audit Samples, National Stream Survey - Phase I

Variable	Units	Accuracy objective (%)	Index value (n^a = 4) Grand[b] mean	±CI[c]	Laboratory 1 measured values n^a	Mean	±CI[c]	Percent accuracy[d]
Bagley Lake samples								
Al-total	mg/L	10	0.016	0.0006	14	0.019	0.0152	19[ns]
ANC	µeq/L	10	121.0	0.16	14	120.7	1.83	<1
BNC	µeq/L	10	29.7	0.52	14	23.0	2.42	-22*
DIC-eq	mg/L	10	1.52	0.015	14	1.56	0.032	3
DIC-initial	mg/L	10	1.52	0.012	14	1.68	0.029	0*
pH-ANC	pH units	0.1[e]	7.06	0.014	14	7.08	0.051	0.02[f]
pH-BNC	pH units	0.1[e]	7.06	0.014	14	7.13	0.068	0.07[ns,f]
pH-eq	pH units	0.1[e]	7.29	0.014	14	7.30	0.103	0.01[f]
Big Moose Lake samples								
Al-ext	mg/L	10	0.197	0.0052	13	0.223	0.0296	13[ns]
Al-total	mg/L	10	0.272	0.0045	13	0.259	0.0196	-5
ANC	µeq/L	10	-3.1	0.45	15	-2.9	1.24	-6
BNC	µeq/L	10	72.9	1.36	15	64.4	3.68	12*
DIC-eq	mg/L	10	0.11	0.009	15	0.07	0.026	-36*
DIC-initial	mg/L	10	0.36	0.009	15	0.25	0.022	-31*
Fe	mg/L	10	0.05	0.002	15	0.04	0.010	0
pH-ANC	pH units	0.1[e]	5.10	0.010	15	5.14	0.041	0.04[f]
pH-BNC	pH units	0.1[e]	5.15	0.013	15	5.16	0.052	0.01[f]
pH-eq	pH units	0.1[e]	5.17	0.006	15	5.17	0.008	0

[a] n = number of measurements.
[b] Grand mean based on weighted mean values from four laboratories.
[c] ±CI = 95 percent confidence interval.
[d] For pH, accuracy is expressed as the absolute difference between the index value and the measured value.
* = mean value significantly different from the index value at $p \leq 0.05$.
ns = accuracy estimate outside data quality objective, but mean value is not significantly different from the index value at p = 0.05.
[e] Objective expressed in pH units.
[f] Accuracy expressed as the difference between the index value and the mean analytical laboratory value.

(0.04 mg/L, see Appendix A) and those of Laboratory 2 (0.02 to 0.04 mg/L, discussed in the next subsection). Measurements of iron in the synthetic audit samples were imprecise, with the 95 percent confidence intervals representing from ±30 percent of the mean (0.011 mg/L for Lot 14) to ±136 percent of the mean (0.041 mg/L) for Lot 15 (Table 21). Data from the Big Moose Lake sample (Table 22), which had an iron concentration similar to that of the synthetic audit sample, did not indicate a relative bias with respect to the index value. There is no strong reason to suspect systematic errors in iron measurements within the range of concentrations represented by the audit samples.

Data for phosphorus measurements from lot 14 (Table 21) indicate the potential for negative bias at high phosphorus concentrations (greater than 0.20 mg/L). For Lot 14, the mean value was in agreement with that from Laboratory 2 (0.022 mg/L; discussed in the following subsection), and thus the observed error may represent a loss of phosphorus from the synthetic audit during the preparation of preserved aliquots at the support laboratory. However, the mean value from Lot 15 (0.015 mg/L) is not in agreement with that from Laboratory 2 (0.023 mg/L); see the next subsection), indicating that measurements made at Laboratory 1 during the last half of the survey may be underestimates of the true sample concentrations. The accuracy estimate for Lot 15, however, is based on only three measurements. One outlying value (-0.0050 mg/L) was not included in estimating accuracy.

The accuracy estimates for ANC, DIC-eq, and pH-eq for both lots of synthetic audit samples were within or near the data quality objectives based on comparisons to the index values (Table 21). For DIC-initial, pH-ANC, and pH-BNC measurements, the observed biases (positive for DIC in Lot 14), indicate that the dissolved carbon dioxide concentration in the audit samples generally increased between the time of sample preparation and measurement. This change could result from a decrease in temperature (increasing the solubility of carbon dioxide), a higher ambient atmospheric concentration of carbon dioxide at the analytical laboratory, or a combination of both effects. Equilibrated pH and DIC measurements, although within accuracy objectives, were imprecise, providing additional evidence of the sensitivity of the synthetic audit samples to changes in dissolved carbon dioxide concentrations. Data from the Bagley Lake sample, with index values for ANC, DIC, and pH that were similar to those calculated for the synthetic audit sample, yielded accuracy estimates that were within or near the data quality objectives for ANC, all DIC measurements, and all pH measurements (Table 22). Data from the Big Moose Lake sample (Table 22) had index values for pH that were between 5.10 and 5.20, index values for DIC less than 0.4 mg/L, and an index value for ANC of approximately -3 μeq/L. Accuracy estimates for ANC and pH from Laboratory 1 for the Big Moose Lake sample were within data quality objectives. Accuracy estimates for DIC were outside the data quality objective but the observed differences were small (approximately 0.1 mg/L or less).

For BNC, accuracy estimates for the synthetic audit samples are outside the data quality objective for Lot 14 (Table 21), when compared to the index value based on support laboratory measurements. Data from the Bagley Lake sample (Table 22), which has an index value for BNC similar to that of the synthetic audit samples, suggest that measurements of BNC from Laboratory 1 may be underestimates of sample concentrations, but the observed magnitude of the relative bias is small (5 to 10 μeq/L) and occurs only at very low concentrations of BNC (30 μeq/L). Data from the Big Moose Lake audit sample (Table 22) which had a higher index value for BNC (72.9 μeq/L), indicate that BNC measurements from Laboratory 1 exceeded the DQO of 10 percent by 2 percent relative to the index values.

In conclusion, the only data quality problems observed for Laboratory 1 that are related to accuracy appear to be the potential for underestimating BNC at low concentrations (less than 30 μeq/L), specific conductance in dilute samples (less than 25 μS/cm), and phosphorus during the latter half of the survey. For all other variables, percent accuracy estimates were within or near the data quality objectives.

Percent Accuracy Estimates for Laboratory 2

Accuracy estimates for synthetic and natural audit samples for Laboratory 2 are presented in Tables 23 and 24. For the synthetic audit sample, accuracy estimates for ten variables for which theoretical values were available were within or near the data quality objectives for both lots (Table 23): calcium, chloride, fluoride, potassium, magnesium, manganese, sodium, ammonium, nitrate, and sulfate.

Table 23. Percent Accuracy Estimates for Laboratory 2 Measurements of Synthetic Audit Samples, National Stream Survey - Phase I

Variable	Units	Accuracy objective (%)	Theoretical value[b]	Lot 14[a] n	Lot 14 Mean	Lot 14 CI	Lot 14 Percent accuracy	Lot 15[a] n	Lot 15 Mean	Lot 15 CI	Lot 15 Percent accuracy
Al-ext	mg/L	10	0.020	5	0.021	0.0024	5	23	0.015	0.0008	-25*
Al-total	mg/L	10	0.020	4	0.035	0.0103	75*	23	0.031	0.0032	55*
Ca	mg/L	10	0.19	9	0.19	0.017	0*	29	0.18	0.004	-5*
Cl^-	mg/L	10	0.34	8	0.32	0.018	-6*	28	0.33	0.011	-3
Cond-lab	µS/cm	5	17.5	9	19.5	0.21	11*	28	19.6	0.07	12*
DOC	mg/L	10	1.0	9	1.3	0.20	30*	28	1.1	0.08	10*
F^-	mg/L	10	0.042	9	0.043	0.0012	2	28	0.043	0.0004	2*
Fe	mg/L	10	0.06	5	0.04	0.003	-33*	23	0.02	0.006	-67*
K	mg/L	10	0.20	9	0.20	0.006	0	29	0.20	0.004	0
Mg	mg/L	10	0.45	9	0.44	0.016	-2	28	0.44	0.004	-2*
Mn	mg/L	10	0.10	9	0.11	0.002	10	29	0.10	0.001	0
Na	mg/L	10	2.75	8	2.76	0.062	0	29	2.70	0.038	-2*
NH_4+	mg/L	10	0.17	9	0.19	0.010	12*	28	0.17	0.005	0
$N0_3^-$	mg/L	10	0.467	8	0.465	0.0140	0	28	0.473	0.0132	1
P	mg/L	10	0.027	9	0.022	0.0026	-18	28	0.023	0.0012	15*
SiO_2	mg/L	10	1.07	9	0.92	0.044	-10*	28	0.92	0.024	-14*
$SO_4 2-$	mg/L	10	2.28	9	2.15	0.157	-6	29	2.30	0.046	1

Variable	Accuracy objective (%)	Lot 14 Index value(CI)[c]	Lot 14 n	Lot 14 Mean	Lot 14 CI	Lot 14 Percent accuracy	Lot 15 Index value(CI)[c]	Lot 15 n	Lot 15 Mean	Lot 15 CI	Lot 15 Percent accuracy
ANC	10	101.7 (3.04)	9	114.4	5.73	7*	109.2 (3.07)	29	119.0	2.71	10*
BNC	10	32.0 (4.58)	8	49.1	2.59	53*	22.8 (4.04)	29	58.9	4.74	158*
DIC-eq	10	1.14 (0.050)	9	1.37	0.202	20[ns]	1.26 (0.050)	29	1.32	0.082	5
DIC-init	10	1.14 (0.050)	9	1.62	0.074	42*	1.26 (0.050)	29	1.51	0.102	20*
pH-ANC	0.1[d]	7.22 (0.044)	9	6.73	0.049	0.49[*e]	7.29 (0.044)	29	6.69	0.048	-0.6[*e]
pH-BNC	0.1[d]	7.22 (0.044)	8	6.71	0.033	0.51[*e]	7.29 (0.044)	29	6.69	0.049	-0.6[*e]
pH-eq	0.1[d]	7.22 (0.044)	9	7.25	0.069	0.03[e]	7.29 (0.044)	29	7.23	0.056	0.06[e]

[a] n = number of measurements.
CI = one-sided 95 percent confidence interval.
ns = not significantly different from the theoretical or index value at p = 0.05.
* = mean value significantly different from the index value at $p \leq 0.05$.

[b] The theoretical value is the expected value of the synthetic audit sample assuming no preparation error and no external effect.

[c] Measured mean values from the support laboratory for Lot 14 and Lot 15.

[d] Objective expressed in pH units.

[e] Accuracy expressed as the difference between the index value and mean analytical laboratory values.

Table 24. Percent Accuracy Estimates for Laboratory 2 Measurements of Selected Variables, in Natural Audit Samples, National Stream Survey - Phase I.

Variable	Units	Accuracy objective (%)	Index value (n^a = 4) Grand[b] mean	±CI[c]	Laboratory 2 measured values (n^a = 24) Mean	±CI[c]	Percent accuracy[d]
Bagley Lake samples							
Al-total	mg/L	10	0.016	0.0006	0.031	0.0191	94^{ns}
ANC	µeq/L	10	121.0	0.16	132.0	5.45	9*
BNC	µeq/L	10	29.7	0.52	43.1	11.38	45*
DIC-eq	mg/L	10	1.52	0.015	1.47	0,105	-3
DIC-init	mg/L	10	1.52	0.012	1.46	0.097	-4
pH-ANC	pH units	0.1[e]	7.06	0.014	6.97	0.036	0.09^{*f}
pH-BNC	pH units	0.1[e]	7.06	0.014	6.98	0.036	0.08^{*f}
pH-eq	pH units	0.1[e]	7.29	0.014	7.31	0.027	0.0^{f}
Big Moose Lake samples							
Al-ext	mg/L	10	0.197	0.0052	0.210	0.0082	7^{ns}
Al-total	mg/L	10	0.272	0.0045	0.281	0.0071	3
ANC	µeq/L	10	-3.1	0.45	-1.4	2.50	-48^{ns}
BNC	µeq/L	10	72.9	1.36	79.8	4.01	8
DIC-eq	mg/L	10	0.11	0.009	0.10	0.038	-9
DIC-init	mg/L	10	0.36	0.009	0.23	0.035	-38*
Fe	mg/L	10	0.05	0.002	0.06	0.005	20*
pH-ANC	pH units	0.1[e]	5.10	0.010	5.22	0.079	0.12^{*f}
pH-BNC	pH units	0.1[e]	5.15	0.013	5.24	0.079	0.09^{*f}
pH-eq	pH units	0.1[e]	5.17	0.006	5.24	0.032	0.07^{*f}

[a] n = number of measurements.
[b] Grand mean based on weighted mean values from four laboratories.
[c] ± CI = 95 percent confidence interval.
[d] For pH, accuracy is expressed as the absolute difference between the index value and the measured value.
* = mean value significantly different from the index value at $p \leq 0.05$.
ns = Accuracy estimate outside data quality objective, but mean value is not significantly different from the index value at p = 0.05.
[e] Objective expressed in pH units.
[f] Accuracy expressed as the difference between the index value and the mean analytical laboratory values.

For three variables (extractable aluminum, specific conductance, and phosphorus), accuracy estimates were outside the data quality objectives for one or both lots, but these estimates do not represent data quality problems. The mean value for extractable aluminum in Lot 15 (0.015 mg/L) is in agreement with the mean value measured at Laboratory 1 (0.015 mg/L; Table 21). The accuracy estimate based on the Big Moose Lake sample (7 percent; Table 24), which has a higher concentration of extractable aluminum, was not statistically significant and is less than the apparent bias observed for Laboratory 1 (13 percent; Table 22). At low concentrations, systematic errors in extractable aluminum measurements will probably be masked by imprecision due to the extraction process.

The apparent bias in specific conductance measurements (11 to 12 percent; Table

23) represents a difference of approximately 2 μS/cm. This difference is about equal in magnitude, but in the opposite direction, to that observed for Laboratory 1 (approximately 16 μS/cm, Table 21). However, measurements of specific conductance from both types of the natural audit samples are in agreement with index values (see Appendix A). Thus, evidence for systematic errors in specific conductance measurements is inconclusive.

For phosphorus measurements, mean values observed in the synthetic audit samples were consistent for both lots (0.022 and 0.023 mg/L, Table 23) and similar to the mean from Laboratory 1 for Lot 14 (0.021 mg/L, Table 20). The magnitude of the bias is small (approximately 0.005 mg/L) and may partially result from an error in sample preparation at the support laboratory. There is no strong evidence to suggest a data quality problem with respect to accuracy for phosphorus measurements.

The mean value for DOC was slightly greater (1.1 to 1.3 mg/L) than the theoretical value in the synthetic audit sample. For the Big Moose Lake sample, which had a higher index value for DOC (approximately 3.6 mg/L; see Appendix A), the mean value for measurements made at Laboratory 2 (n = 24) was 4.1 mg/L (Appendix A). It appears that measured values of DOC from Laboratory 2 may be an overestimate of true sample concentrations. The magnitude of the apparent bias is within the range of background levels measured in field blank samples (Table 17).

For iron, mean values in the synthetic audit sample (0.04 and 0.02 mg/L) were lower than the theoretical value (0.06 mg/L, Table 23) but were in agreement with the mean value from Laboratory 1 (0.030 mg/L, Table 21). Data from the Big Moose Lake sample (Table 24) indicate a potential for a positive bias; the mean value (0.06 mg/L) was larger than the index value (0.05 mg/L). The magnitude of the apparent bias (0.01 mg/L) is small and may be masked by measurement imprecision at low concentrations.

For total aluminum and silica, systematic errors may have an effect on data interpretation. Mean values for total aluminum measure- ments of the synthetic audit sample were much greater than the theoretical for both lots (Table 23). Data from the Bagley Lake sample (Table 24), which had an index value for total aluminum similar to the theoretical concentration of the synthetic audit sample, also indicates a potential for positive bias of approxi- mately 0.01 mg/L. The magnitude of this bias (0.01 to 0.015 mg/L) appears to be consistent over a range of concentrations; the mean value for measurements of the Big Moose Lake sample (0.281 mg/L; Table 24), which has a much higher index value for total aluminum (0.272 mg/L), also indicates a positive bias of approximately 0.01 mg/L. The bias may result from reagent contamination, as was suggested by the analysis of blank samples (Table 15) and should only affect the use of values of total aluminum at low concentrations (less than 0.1 mg/L).

For silica measurements, an apparent negative bias was observed for both lots of synthetic audits (0.11 mg/L to 0.15 mg/L; Table 23). Data from both types of natural audit samples, with index values for silica greater than the theoretical concentration of the synthetic audit sample, also indicated a potential negative bias in silica measurements relative to the index values. For the Big Moose Lake audit sample the index value for silica was 4.48 mg/L, while the mean measured value for Laboratory 2 was 3.95 mg/L (see Appendix A). For the Bagley Lake sample, the index value was 9.48 mg/L, while the mean measured value for Laboratory 2 was 8.90 mg/L (see Appendix A).

Accuracy estimates for ANC and equilibrated pH relative to index values were within or near to the data quality objectives for both lots (Table 21). For equilibrated DIC, the mean value for Lot 14 was not significantly different from the index value, while the accuracy estimate for Lot 15 was within the data quality objective.

Data for DIC and pH indicate that the synthetic audit samples (Table 23) may have increased their dissolved carbon dioxide concentration to a greater degree at Laboratory 2 than at Laboratory 1 (Table 21). Mean values for initial DIC measurements made at Laboratory 2 were generally 0.1 to 0.2 mg/L greater than those from Laboratory 1, while pH-ANC and pH-BNC measurements were approximately 0.3 pH units lower.

Data from the Bagley Lake audit sample (Table 24) indicate that accuracy estimates for initial DIC (-4 percent), equilibrated DIC (-3 per- cent), pH-ANC (-0.09 pH units), pH-BNC (-0.08 pH units) and pH-eq (0.0 pH units) measurements were within the data quality objectives for accuracy with respect to the index values. For the Big Moose Lake sample, mean values for equilibrated DIC (0.10 mg/L; Table 24) were within the accuracy objectives. The mean for initial DIC measurements (0.23 mg/L), although outside the data quality objective with respect to the index value, was in agreement with the mean value for Laboratory 1 (0.26 mg/L; Table 22). Mean values for pH measurements from the Big Moose Lake sample (Table 24) were 0.07 to 0.12 units higher than the index value, although they were still near or within the data quality objectives with respect to the index value.

BNC measurements from Laboratory 2 appear to overestimate actual sample concentrations significantly at low concentrations. For the synthetic audit sample, mean values for both lots were approximately 1.5 and 1.8 times greater than index values based on support laboratory measurements (Table 23). For the Bagley Lake sample, which had an index value for BNC similar to that of the synthetic audit sample, the mean value (43.1 μeq/L) was much greater than the index value (29.7 μeq/L; Table 24). For the Big Moose Lake sample, which had a higher index value for BNC (72.9 μeq/L), the accuracy estimate for Laboratory 2 of 8 percent (Table 24) was within the data quality objective.

In conclusion, systematic errors in measurements of total aluminum, BNC, DOC, iron, and silica may affect the interpretation of analytical data from Laboratory 2. For all of these variables, the suspected systematic errors will be most influential at low concentrations.

Percent Accuracy Estimates for the Processing Laboratory

Accuracy estimates for six variables measured in synthetic and natural audit samples at the processing laboratory are presented in Tables 25 and 26. For the synthetic audit samples, the theoretical concentration of total monomeric and nonexchangeable monomeric aluminum should be equal to the value for extractable aluminum (0.020 mg/L). However, the processing laboratory only analyzed field audit samples, and accuracy estimates for the two aluminum fractions will not be representative, because of the possible loss of aluminum from the field audit samples that was noted in the introduction to the accuracy section. For both lots, mean values for the two aluminum fractions were less than the theoretical values. Mean values for nonexchangeable monomeric aluminum were greater than those for total monomeric aluminum, probably resulting from the high instrument background observed in the analysis of blank samples (Table 15).

Aluminum concentrations in the Bagley Lake sample were below the limit of detection. Data from the Big Moose Lake sample (Table 26) show that the percent accuracy estimates were within the data quality objectives for both aluminum fractions (1 percent for Al-mono; 5 percent for Al-nex). Systematic errors do not appear to be evident at higher concentrations of total monomeric or nonexchangeable monomeric aluminum.

Specific conductance measurements were not accurate during the first half of the NSS-I, as evidenced by the mean value observed for Lot 14 (22.7 μS/cm). The malfunctioning probe that was used during the first half of the NSS-I was the source of this error.

For the closed-system DIC and pH measurements, data from both lots of

Table 25. Percent Accuracy Estimates for Processing Laboratory Measurements of Synthetic Audit Samples, National Stream Survey - Phase I

Variable	Units	Accuracy objectives (%)	Theoretical value[a]	Field synthetic audit samples: Lot 14 n[b]	Mean	CI[c]	%Acc[d]	Lot 15 n[b]	Mean	CI[c]	%Acc[d]
Al-mono	mg/L	10	0.020	7	0.008	0.0031	-60*	5	0.009	0.0065	55*
Al-nex	mg/L	10	0.020	7	0.014	0.0034	-30*	5	0.013	0.0071	-35[ns]
Cond-PL	µS/cm	5	17.5	6	22.7	2.08	30*	4	18.8	1.88	7[ns]
True color	PCU	--	NC	8	5	2	--	4	4	8	--

Variable (Units)	Accuracy objective (%)	Lot 14 Index value[e] (CI)[c] (n = 6)	Lot 14 Mean (CI)[c] (n = 9)	Lot 14 %Acc.[d]	Lot 15 Index value[e] (CI)[c] (n = 6)	Lot 15 Mean (CI)[c] (n = 5)	Lot 15 %Acc[d]
DIC[f](mg/L)	10	1.14 (0.050)	1.38 (0.030)	21*	1.26 (0.050)	1.44 (0.041)	14*,[h]
pH[f](pH units)	±0.1[g]	7.22 (0.044)	6.96 (0.074)	-0.26*	7.29 (0.044)	6.92 (0.078)	-0.37[h]

[a] The theoretical value is the expected value of the synthetic audit sample assuming no preparation error and no external effect. NC = theoretical value not calculated.

[b] n = number of measurements.

[c] One-sided 95% confidence interval.

[d] Percent accuracy.

* = mean value significantly different from the theoretical or index value at $p \leq 0.05$.

ns = percent accuracy outside accuracy objective, but mean value is not significantly different from the theoretical or index value at P = 0.05.

[e] Index value = mean value from support laboratory measurements.

[f] Closed-system measurement.

[g] Accuracy expressed in pH units.

[h] Accuracy expressed as the difference between the index value and the mean analytical laboratory values.

synthetic audit samples show the same trend as was observed for the analytical laboratory measurements of DIC and pH (Tables 21 and 23): a DIC concentration that was greater than the index value and a pH value that was lower than the index value. For DIC and pH, the index values for the Bagley Lake sample were calculated based on measurements collected at the processing laboratory during the NSS-I and the Western Lake Survey-Phase I. For the Big Moose Lake sample, the index value was based on measurements conducted during the NSS-I and four seasonal studies that were conducted during Phase II of the Eastern Lake Survey. Data for closed-system DIC and pH measurements from both natural audit samples do not indicate systematic errors that exceed data quality objectives, based on comparisons to index values.

Interlaboratory Bias

An evaluation by Edland et al. of the relative bias between analytical measurements from the two laboratories involved in the NSS-I is presented in Appendix D. Their report presents four different variations of a linear model to describe possible functional relationships of interlaboratory bias to concentration. The general form of the model is:

Laboratory 1 measurements =
$(1 + \beta)$ (Laboratory 2 measurements) + α.

where α is a constant, analogous to the intercept of a regression equation, and β is a proportionality term, which represents the deviation of the slope of a regression equation from 1.0.

Table 26. Estimates of Percent Accuracy for Analytes Measured at the Processing Laboratory Based on Natural Audit Samples, National Stream Survey - Phase I

			Bagley Lake						
			Index value[a]			Measured values			
Variable	Units	Accuracy objective (%)	n[b]	Grand mean[c]	CI[d]	n[b]	Mean	CI[d]	%Acc[e]
Al-mono	(mg/L)	10	2	--	--	--	--	--	--
Al-nex	(mg/L)	10	2	--	--	--	--	--	--
DIC[f]	(mg/L)	10	2	1.67	0.068	27	1.70	0.024	2
pH[f]	(pH units)	±0.1[g]	2	7.04	0.066	27	6.99	0.019	-0.05*

			Big Moose Lake						
			Index value[a]			Measured values			
Variable	Units	Accuracy objective (%)	n[b]	Grand mean[c]	CI[d]	n[b]	Mean	CI[d]	%Acc[e]
Al-mono	(mg/L)	10	2	0.193	0.0174	24	0.195	0.0049	1
Al-nex	(mg/L)	10	2	0.053	0.0178	24	0.058	0.0054	5
DIC[f]	(mg/L)	10	2	0.55	0.061	27	0.024	0.53	4
pH[f]	(pH units)	±0.1[g]	2	5.14	0.060	27	5.15	0.023	0.01

[a] Index values = mean value from support laboratory measurements.
[b] n = number of measurements.
[c] Grand mean based on weighted means of measurements from the processing lab during two surveys (NSS-I and Phase II of the Eastern Lake Survey).
[d] One-sided 95% confidence interval.
[e] %Acc = percent accuracy expressed in pH units.
* = mean value significantly different from the index value at $p \leq 0.05$.
[f] Closed-system measurements.
[g] Accuracy expressed in pH units.

The model was evaluated using the seven groups of audit samples that were measured by both analytical laboratories (two lots of synthetic samples and two natural audit samples, each prepared both as field and as laboratory samples). Laboratory 1 did not analyze any field audit samples from Lot 15. For the models, α and β were derived by using maximum likelihood estimation techniques. The four variations of the model evaluated were:

1. No bias ($\alpha = 0, \beta = 0$)
2. Constant bias ($\alpha \neq 0, \beta = 0$)
3. Proportional bias ($\alpha = 0, \beta \neq 0$)
4. Bias that includes both a constant and proportionality term ($\alpha \neq 0, \beta \neq 0$).

The assumption of linearity was also evaluated.

The results of these evaluations, presented in Appendix D, indicate that

interlaboratory bias of some type is present for most of the variables; the report provides the information required to transform the data so that the two laboratories are calibrated. However, the authors present several considerations which should be weighed before applying the transformation procedures. In many cases, the audit samples do not bracket a large portion of the range of values measured in routine streamwater samples; thus there may be considerable uncertainty introduced if the model is used to extrapolate findings beyond the range represented by the audit samples. A second consideration is that for a number of variables the assumption of linearity was not confirmed. Finally, any gain in accuracy achieved by transforming the data will be accompanied by a loss in precision because of the uncertainty associated with estimating the model parameters α and β.

Because accuracy estimates for most variables from both laboratories were within the data quality objectives, transformation of the data is not necessary to make population estimates. The estimates of among-batch precision presented later in this report include the effects of interlaboratory bias, and thus they can be used to evaluate uncertainty.

Other data interpretation activities, however, may require laboratory measurements to be intercalibrated. For such activities, Edland et al. (Appendix D) provide the appropriate transformation coefficients and also the deviation of the maximum likelihood estimates. Summary data for audit sample measurements are presented in Appendix A. Estimates for accuracy presented in the preceding sections provide information that may be useful in deciding how to proceed with intercalibration (e.g., whether to correct data from one laboratory to be more equivalent to data from the other, or to correct both data sets to some intermediate value).

Discussion and Summary: Accuracy

For nearly all variables, accuracy estimates for measurements at Laboratory 1 were within the data quality objectives. Phosphorus measurements made by Laboratory 1 during the last half of the NSS-I may be underestimates of the true sample concentrations. It is not known whether this apparent bias occurs at lower concentrations or if it is present only at higher concentrations. In addition, specific conductance measurements throughout the NSS-I may be underestimates of true sample values. The apparent magnitude of the bias in audit samples was approximately -2 μS/cm. It is not known if the bias is constant over the entire range of measured values, because the audit samples all had specific conductances between 10 and 25 μS/cm. Interpretation of data from dilute samples should consider the potential for negative bias.

For Laboratory 2, measurements of several variables in the audit samples had potential systematic errors. In each case, the bias appeared to be constant over the range of values measured in the audit samples; thus, the bias will have the greatest impact in interpreting data at low concentrations. For total aluminum, a positive bias on the order of 0.010 mg/L was observed. This should have little impact on data interpretation, as most of the streamwater samples will have much higher concentrations.

For iron, the observed bias in audit samples was on the order of 0.02 mg/L. At low concentrations, iron is not an important contribution to the overall ion balance, and the potential systematic error does not affect data interpretation. For DOC, the observed bias in audit sample measurements was between 0.1 and 0.5 mg/L and may only need to be considered when comparing groups of samples having low concentrations of DOC. Silica measurements from Laboratory 2 may be underestimates of true sample values, based on audit sample measurements. The observed magnitude of the bias was on the order of 0.15 mg/L at low concentrations (1 mg/L) to 0.5 mg/L at higher concentrations (3 to 10 mg/L). Interpretation of silica data should consider the possibility of systematic errors.

Measurements of BNC in audit samples having low concentrations (less than 30 μeq/L) were subject to systematic errors and also to

poor precision, particularly at Laboratory 2. The usefulness of the BNC data may be limited because of the method of measurement. For the NSS-I, the primary purpose of the BNC measurement was to provide a means of verifying the ANC measurement on a sample-by-sample basis. Because BNC is affected by dissolved carbon dioxide, special precautions must be taken to yield accurate results. Titration with a strong base in a vessel exposed to the atmosphere will cause the sample to continually increase its dissolved carbon dioxide concentration. If future studies require reliable BNC data to differentiate weak and strong acids in natural water samples, the method should be modified so that the titration is conducted under an inert, carbon-dioxide-free atmosphere.

The synthetic audit samples used for the NSS-I provided a reliable means to assess accuracy for most variables. However, this sample, if not filtered and preserved immediately after preparation, was affected by the loss of aluminum and iron. In addition, theoretical values for variables affected by dissolved carbon dioxide, such as ANC, BNC, DIC, and pH, are difficult to calculate and may not be valid because the preparation of the synthetic audit sample composition is not necessarily based on the assumption of carbonate equilibrium. In addition, the concentrations of these four variables can be affected by preparation errors arising from adding other chemical constituents to the synthetic audit sample. It may be desirable to allow the synthetic audit sample to equilibrate for a period of time after preparation, but before use. Following equilibration, the sample would be subjected to rigorous verification measurements that include comparisons to certified standards so that the audit sample composition is known with a high level of certainty for all variables. This kind of verification should also be made for the natural audit samples. Verified natural audit samples would offer a means to assess overall measurement accuracy including systematic errors resulting from sample collection and handling.

As an alternative, synthetic audit samples could be prepared on an analyte-by-analyte basis or as aliquots that include chemically compatible variables. Such samples would provide reliable concentrations for all variables to assess absolute accuracy, but they also would have the disadvantage that they could only be introduced at the point of analysis. In addition, assessments of accuracy for measurements of these kinds of samples would not include errors due to collection and handling.

Finally, if interlaboratory bias is a concern, analytical objectives for accuracy should be developed to control interlaboratory differences. An accuracy objective of $\pm$10 percent allows for interlaboratory differences of up to 20 percent. It may be desirable to establish an accuracy objective at $\pm$5 percent, thus reducing the tolerable interlaboratory difference to $\pm$10 percent. Coincident with such a reduction, the laboratories should be periodically provided with known performance standards from a single source so that all laboratories can calibrate their measurement systems to a given target value, rather than relying on internally developed standards to monitor their performance. Samples of unknown composition would then provide an independent check on laboratory performance.

Precision

The precision (or variation) associated with various components of the collection and measurement system was assessed by using the results from analyses of processing and analytical laboratory duplicate, field duplicate, and audit samples. Table 27 illustrates the potential sources of variation for each kind of sample and those components of variation that are included in an estimate of precision based on each type of QA sample. Processing and analytical laboratory duplicate samples provided an estimate of within-batch analytical precision. The DQOs for precision (Table 13) were based on the expected analytical performance at a single laboratory; laboratory duplicate samples were used to assess precision relative to the DQOs. Field duplicate samples provided estimates of overall within-batch precision, including sample collection, handling, and processing and analytical errors.

Table 27. Components of Variance Included in Precision Estimates From Routine-Duplicate Pairs and Audit Samples, National Stream Survey - Phase I

Sources of measurement error	Type of sample			
	Field duplicate	Laboratory duplicate	Field audit[a]	Laboratory audit[a]
Sampling (system-level)				
Among crews				
Day-to-day				
Among samples	X			
Sampling variance	X			
Processing (aliquot preparation and preservation)				
Among laboratories			X	
Among batches			X	
Within a batch	X			
Subsampling	X		X	
Analysis (method-level)				
Among laboratories			X	X
Among batches			X	X
Within a batch	X	X		
Within a sample	X	X	X	X

[a] Field and laboratory audits also have variance components associated with their preparation.

Audit samples provide estimates of among-batch precision that include the effects of day-to-day differences within a single laboratory, effects from processing and transport, and also the effects due to interlaboratory biases.

None of the samples listed in Table 27 provide an estimate of total overall variability due to the collection and measurement of samples. Precision estimates based on duplicate samples do not account for among-batch variation. Audit samples provide an estimate of among-batch and among-laboratory variation, but this estimate does not include any effects of sample collection. Precision esti-mates from routine-duplicate sample pairs are based on pooled measurements for ranges of concentrations; estimates from audit samples are based on repeated measurements at a single concentration. Nevertheless, qualitative comparisons are possible to discern the major sources of variation in the NSS-I collection and measurement system and to evaluate the effect of imprecision on data interpretation.

Precision Estimates Derived from Field and Laboratory Duplicate Samples

Because of the range of values for chemical variables encountered during the NSS-I and concentration-dependent effects (Mericas and Schonbrod, 1987), a single estimate of precision based on all pairs of routine-duplicate samples may be misleading. This analysis did not use model-based approaches to predict precision as a function of concentration (e.g., Mericas and Schonbrod, 1987) because data from duplicated measurements violated several assumptions of regression

analysis (e.g., concentrations were not measured without error, measured pairs were distributed over the entire measurement range). Because precision varies with concentration, precision was estimated from routine-duplicate measurements for several ranges of values. Examination of scatterplots of the standard deviation versus the mean concentration of sample pairs within each range subset indicated that there are no relation ships between variance and concentration.

A total of 65 field routine-duplicate pairs and 68 laboratory routine-duplicate pairs were analyzed for most variables. Occasionally, a routine-duplicate pair was not included for a batch of samples. On some occasions, an analytical laboratory analyzed more than one sample in duplicate. In the latter cases the first pair was used in precision estimates. Measurements that were qualified with an X flag were not included in precision estimates.

Precision estimates for each range subset were based on a pooled standard deviation that was calculated from the mean and var-iances of the individual sample pairs. The following formula was used to calculate a pooled standard deviation (Taylor, 1987) with the number of replicates in each case equal to two:

$$s_p = \sqrt{\sum_{i=1}^{n} s_i^2/n}$$

where s_p = pooled standard deviation,
n = number of routine-duplicate pairs,
s_i^2 = variance of a routine-duplicate pair.

For ANC and BNC, expressing precision in relative terms can be misleading for values less than 100 μeq/L. Except for ANC, BNC, pH, and true color, precision was also expressed as a relative pooled standard deviation (%RSD_p) by dividing s_p by the grand mean of the sample pairs and multiplying by 100:

$$RSD_p = \left(\frac{s_p}{\sum_{i=1}^{n} \bar{x}_i/n} \right) 100$$

where s_p = pooled standard deviation,
$\bar{x}_i$ = mean of a sample pair,
n = number of sample pairs.

Method-Level Precision Estimates--

Estimates of pooled standard deviations and pooled relative standard deviations for subranges of variables are presented in Table 28. These estimates are for combined measurements from both analytical laboratories. Examination of method-level (within-batch) precision estimates for each laboratory, although not presented in detail here, indicated that for all variables neither laboratory was outside the within-laboratory precision objectives (Table 13). The DQOs for precision were established initially for measured values greater than 10 times the detection limit objective (Table 13). For all subranges greater than the detection limit objective, within-batch precision estimates for the combined measurements (Table 28) were also within the DQOs for precision for all variables.

System-Level Precision Estimates from Field Duplicate Samples--

Because DQOs were not established for system-level precision for the NSS-I, the method-level DQOs were used as a gauge. Measurements of field routine-duplicate sample pairs from both analytical laboratories were combined to estimate system-level (overall within-batch) precision for the NSS-I Table 28). Eighteen variables had s_p or %RSD_p estimates that were within or near the within-laboratory precision goal for all subset ranges except the lowest, which represented values below the system decision limit for most variables. These variables were total monomeric aluminum (largest %RSD_p = 12.4), ANC (largest s_p = 11.22 μeq/L), BNC (largest s_p = 12.25 μeq/L), calcium (largest %RSD_p = 5.6), chloride

Table 28. Method-Level and System-Level Precision Estimates by Concentration Ranges of Variables (Laboratories Pooled), National Stream Survey - Phase I

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]
Al-ext (mg/L)								
<0.007	3	0.004	0.0002	5.3	29	0.003	0.0016	51.5
0.007 to 0.050	35	0.019	0.0009	4.6	25	0.016	0.0059	37.2
0.050 to 0.100	15	0.071	0.0026	3.6	4	0.072	0.0021	2.9
>0.100	15	0.269	0.0055	2.0	7	0.278	0.0166	6.0
All data	68	0.085	0.0029	3.4	65	0.042	0.0067	16.1
Al-total (mg/L)								
<0.027	9	0.014	0.0018	13.1	4	0.019	0.0035	18.7
0.027 to 0.100	6	0.056	0.0034	6.1	16	0.056	0.0079	14.1
0.100 to 0.500	51	0.204	0.0042	2.0	38	0.214	0.0807	37.8
0.500 to 1.000	1	0.791	0.0304	3.8	7	0.767	0.1953	25.5
>1.000	1	1.235	0.0573	4.6	0	--	--	--
All data	68	0.189	0.0087	4.6	65	0.222	0.0905	40.7
Al-mono (mg/L)								
<0.015	33	0.010	0.0011	11.0	30	0.010	0.0022	21.4
0.015 to 0.100	24	0.026	0.0017	6.4	29	0.031	0.0038	12.4
0.100 to 0.500	7	0.273	0.0033	1.2	6	0.283	0.0049	1.7
0.500 to 1.000	2	0.534	0.0045	0.8	0	--	--	--
>1.000	1	1.908	0.0030	0.2	0	--	--	--
All data	67	0.087	0.0019	2.1	65	0.045	0.0033	7.4
Al-nex (mg/L)								
<0.023	52	0.014	0.0022	15.8	47	0.016	0.0031	19.2
0.023 to 0.100	12	0.043	0.0033	7.6	18	0.042	0.0068	16.2
0.100 to 0.500	3	0.225	0.0139	6.2	0	--	--	--
All data	67	0.029	0.0039	13.4	65	0.023	0.0045	19.2
ANC (μeq/L)								
<0	2	-6.8	0.70	--	7	-42.4	2.94	--
0 to 50	4	21.9	1.05	--	9	27.9	7.83	--
>50	62	401.2	4.42	--	49	425.0	11.22	--
All data	68	366.9	4.23	--	65	319.7	10.22	--
BNC (μeq/L)								
0 to 50	24	32.6	2.68	--	28	36.2	6.38	--
>50	44	93.4	6.72	--	37	111.1	12.25	--
All data	68	71.9	5.64	--	65	78.8	10.14	--

(Continued)

Table 28. (Continued)

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]
Ca (mg/L)								
0.02 to 1.00	6	0.59	0.009	1.5	4	0.56	0.015	2.7
1.00 to 5.00	35	2.37	0.041	1.7	35	2.55	0.057	2.2
5.00 to 10.00	12	7.06	0.034	0.5	19	6.82	0.090	1.3
>10.00	15	25.59	0.529	2.1	7	28.90	1.604	5.6
All data	68	8.16	0.251	3.1	65	6.51	0.530	8.1
Cl^- (mg/L)								
<0.03	2	0.00	0.000	--	0	--	--	--
0.03 to 1.00	13	0.57	0.006	1.1	8	0.69	0.018	2.6
1.00 to 2.00	13	1.49	0.022	1.5	18	1.53	0.033	2.2
2.00 to 5.00	23	3.18	0.045	1.4	24	3.05	0.043	1.4
5.00 to 10.00	11	6.73	0.077	1.1	6	7.46	0.073	1.0
>10.00	6	15.89	0.336	2.1	9	18.21	0.734	4.0
All data	68	3.96	0.108	2.7	65	4.85	0.276	5.7
Cond-PL (µS/cm)								
<25.0	5	20.1	1.11	5.5	5	20.4	0.26	1.3
25.0 to 50.0	22	37.8	1.94	5.1	24	37.7	1.26	3.3
50.0 to 100.0	21	75.0	5.27	7.0	24	70.1	1.32	1.9
>100.0	19	244.3	16.99	7.0	12	192.0	3.12	1.6
All data	67	106.7	9.59	9.0	65	76.8	1.75	2.3
Cond-lab (µS/cm)								
<25.0	14	19.9	0.12	0.6	6	20.6	0.87	4.2
25.0 to 50.0	20	35.6	0.23	0.6	24	36.9	0.27	0.7
50.0 to 100.0	17	69.2	0.29	0.4	24	69.9	0.42	0.6
>100.0	17	206.9	0.49	0.2	11	197.4	1.22	0.6
All data	68	83.6	0.31	0.4	65	74.7	0.64	0.9
DIC-closed (mg/L)								
<1.00	9	0.60	0.025	4.2	7	0.46	0.020	4.3
1.00 to 2.00	8	1.56	0.040	2.5	9	1.43	0.063	4.4
2.00 to 5.00	26	3.42	0.068	2.0	31	3.37	0.092	2.7
5.00 to 10.00	18	6.50	0.168	2.6	13	6.65	0.184	2.8
>10.00	7	31.79	0.236	0.7	5	24.47	0.365	1.5
All data	68	6.56	0.123	1.9	65	5.07	0.147	2.9
DIC-eq (mg/L)								
<0.23	1	0.07	0.002	2.9	10	0.12	0.054	45.3
0.23 to 1.00	4	0.74	0.029	4.0	12	0.55	0.099	18.1
1.00 to 2.00	8	1.69	0.031	1.8	12	1.50	0.162	10.8
2.00 to 5.00	24	3.60	0.067	1.9	22	3.49	0.261	7.5
5.00 to 10.00	18	7.27	0.086	1.2	5	6.90	0.184	2.7
>10.00	13	19.21	0.639	3.3	4	26.22	1.050	4.0
All data	68	7.11	0.286	4.0	65	3.72	0.317	8.5

(Continued)

Table 28. (Continued)

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]
DIC-init (mg/L)								
<0.23	0	--	--	--	3	0.17	0.014	8.2
0.23 to 1.00	1	0.24	0.007	2.9	10	0.56	0.091	16.3
1.00 to 2.00	17	1.60	0.036	2.3	15	1.50	0.211	14.0
2.00 to 5.00	18	3.42	0.111	3.2	26	3.39	0.250	7.4
5.00 to 10.00	18	7.22	0.139	1.9	6	6.17	0.050	0.8
>10.00	14	18.82	0.780	4.1	5	24.46	1.009	4.1
All data	68	7.09	0.366	5.2	65	4.24	0.339	8.0
DOC (mg/L)								
<0.5	4	0.3	0.02	6.5	2	0.30	0.02	6.1
0.5 to 2.0	27	1.2	0.05	4.0	41	1.20	0.24	20.8
2.0 to 5.0	16	3.3	0.11	3.3	15	3.10	0.44	14.3
5.0 to 10.0	14	7.2	0.15	2.1	7	7.20	0.58	8.1
>10.0	7	17.2	0.15	0.9	0	--	--	--
All data	68	4.5	0.10	2.3	65	2.23	0.34	15.4
F^- (mg/L)								
0.010 to 0.050	29	0.037	0.0000	0.0	51	0.033	0.0011	3.4
>0.050	39	0.121	0.0021	1.7	14	0.082	0.0029	3.5
All data	68	0.085	0.0016	1.9	65	0.044	0.0017	3.8
Fe (mg/L)								
<0.02	2	0.01	0.001	7.4	14	0.01	0.003	66.0
0.02 to 0.05	3	0.02	0.001	2.7	18	0.03	0.009	34.0
0.05 to 0.10	11	0.08	0.002	3.0	10	0.08	0.029	34.3
0.10 to 0.50	36	0.19	0.004	2.2	18	0.22	0.117	53.3
0.50 to 1.00	3	0.72	0.008	1.2	3	0.64	0.163	25.6
>1.00	12	6.85	0.114	1.7	2	1.55	0.377	24.4
All data	67	1.38	0.048	3.5	65	0.16	0.098	61.4
K (mg/L)								
<0.15	1	0.11	0.001	1.3	1	0.03	0.003	8.3
0.15 to 0.35	10	0.26	0.006	2.5	5	0.26	0.003	1.0
0.35 to 0.45	2	0.39	0.002	0.6	6	0.40	0.005	1.3
>0.45	55	1.47	0.015	1.0	53	1.45	0.063	4.3
All data	68	1.24	0.014	1.1	65	1.24	0.056	4.5
Mg (mg/L)								
<1.00	26	0.54	0.004	0.8	19	0.68	0.013	2.0
1.00 to 2.00	21	1.47	0.013	0.9	23	1.48	0.014	0.9
2.00 to 5.00	15	3.42	0.033	1.0	18	3.04	0.058	1.9
>5.00	6	18.15	0.158	0.9	5	9.39	0.242	2.6
All data	68	3.02	0.050	1.7	65	2.29	0.075	3.3

(Continued)

Table 28. (Continued)

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]	Number of pairs	Grand mean	Pooled s	%RSD$_p$[a]
Mn (mg/L)								
<0.01	2	<0.01	0.001	11.8	19	<0.01	--	56.6
0.01 to 0.05	9	0.03	0.001	2.1	19	0.02	0.015	65.0
0.05 to 0.10	12	0.08	0.003	3.5	7	0.07	0.002	2.7
>0.10	44	1.00	0.017	1.7	20	0.24	0.023	9.5
All data	67	0.68	0.014	2.0	65	0.09	0.015	16.8
Na (mg/L)								
<0.50	4	0.23	0.001	0.6	4	0.28	0.001	0.5
0.50 to 1.00	8	0.73	0.027	3.7	4	0.84	0.030	3.6
1.00 to 2.00	11	1.47	0.017	1.1	24	1.42	0.028	1.9
2.00 to 5.00	33	3.40	0.028	0.8	25	3.52	0.037	1.1
>5.00	12	10.27	0.174	1.7	8	9.66	0.097	1.0
All data	68	3.80	0.076	2.0	65	3.14	0.045	1.4
NH_4^+ (mg/L)								
<0.02	12	0.02	0.000	0.0	33	0.01	0.006	42.0
0.02 to 0.05	20	0.03	0.001	3.2	21	0.03	0.007	23.4
0.05 to 0.10	14	0.07	0.002	3.1	7	0.07	0.014	19.7
>0.10	22	0.35	0.004	1.2	4	0.17	0.011	6.6
All data	68	0.14	0.003	1.9	65	0.03	0.008	22.8
NO_3^- (mg/L)								
<3.000	55	0.722	0.0162	2.2	53	0.686	0.0422	6.2
>3.000	13	8.182	0.1212	1.5	12	14.885	0.3239	2.2
All data	68	2.148	0.0549	2.6	65	3.307	0.1443	4.4
P (mg/L)								
<0.001	2	-0.001	0.0003	30.5	7	-0.001	0.0009	62.9
0.001 to 0.005	24	0.003	0.0003	8.6	31	0.003	0.0016	53.7
0.005 to 0.015	20	0.008	0.0004	4.6	20	0.008	0.0034	43.8
>0.015	22	0.105	0.0018	1.7	7	0.025	0.0040	16.1
All data	68	0.037	0.0010	2.8	65	0.006	0.0026	40.5
pH-closed (pH units)								
<4.00	4	3.75	0.000	--	0	--	--	--
4.00 to 5.00	4	4.51	0.012	--	6	4.55	0.032	--
5.00 to 6.00	7	5.61	0.041	--	13	5.69	0.036	--
6.00 to 7.00	25	6.63	0.012	--	24	6.66	0.036	--
7.00 to 8.00	26	7.43	0.028	--	19	7.26	0.036	--
>8.00	2	8.26	0.010	--	3	8.45	0.026	--
All data	68	6.59	0.023	--	65	6.53	0.035	--

(Continued)

Table 28. (Continued)

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	%RSD_p [a]	Number of pairs	Grand mean	Pooled s	%RSD_p [a]
pH-ANC (pH units)								
4.00 to 5.00	1	4.98	0.014	--	6	4.49	0.022	--
5.00 to 6.00	5	5.73	0.021	--	8	5.70	0.104	--
6.00 to 7.00	34	6.69	0.053	--	27	6.63	0.073	--
7.00 to 8.00	28	7.25	0.039	--	22	7.31	0.091	--
>8.00	0	--	--	--	2	8.10	0.034	--
All data	68	6.82	0.046	--	65	6.59	0.080	--
pH-BNC (pH units)								
4.00 to 5.00	0	--	--	--	6	4.50	0.035	--
5.00 to 6.00	6	5.61	0.034	--	8	5.69	0.092	--
6.00 to 7.00	34	6.70	0.042	--	26	6.63	0.065	--
7.00 to 8.00	28	7.25	0.040	--	23	7.31	0.086	--
>8.00	0	--	--	--	2	8.12	0.062	--
All data	68	6.83	0.041	--	65	6.60	0.075	--
pH-eq (pH units)								
<4.00	2	3.73	0.005	--	0	--	--	--
4.00 to 5.00	7	4.55	0.009	--	6	4.53	0.017	--
5.00 to 6.00	4	5.55	0.000	--	4	5.73	0.041	--
6.00 to 7.00	3	6.86	0.004	--	11	6.73	0.113	--
7.00 to 8.00	34	7.55	0.019	--	40	7.48	0.171	--
>8.00	18	8.45	0.028	--	4	8.37	0.083	--
All data	68	7.22	0.020	--	65	7.03	0.144	--
SiO_2 (mg/L)								
<0.50	1	0.44	0.007	1.6	0	--	--	--
0.50 to 1.50	1	1.03	0.035	3.4	1	0.57	0.064	11.3
1.50 to 5.50	29	3.71	0.052	1.4	25	3.31	0.071	2.2
5.50 to 10.50	19	7.73	0.141	1.8	21	7.70	0.122	1.6
>10.50	18	16.38	0.195	1.2	18	14.51	0.730	5.0
All data	68	8.10	0.129	1.6	65	7.79	0.393	5.0
SO_4^{2-} (mg/L)								
<0.05	2	<0.01	0.002	66.7	0	--	--	--
0.05 to 2.50	13	1.43	0.019	1.3	18	1.49	0.054	3.6
2.50 to 5.00	18	3.60	0.048	1.3	10	3.86	0.081	2.1
5.00 to 12.50	26	7.85	0.117	1.5	24	8.38	0.128	1.5
>12.50	9	22.56	0.239	1.1	13	21.28	0.243	1.1
All data	68	7.21	0.116	1.6	65	8.36	0.140	1.7
True color (PCU)								
<30	50	13	0.9	6.5	44	13	2.6	19.6
>30	17	129	8.7	6.8	21	50	1.9	3.9
All data	67	43	4.4	10.5	65	25	2.4	9.6

(Continued)

Table 28. (Continued)

Variable (units) and measurement range	Method-level precision (processing and analytical laboratory duplicate pairs)				System-level precision (field routine-duplicate pairs)			
	Number of pairs	Grand mean	Pooled s	$\%RSD_p$[a]	Number of pairs	Grand mean	Pooled s	$\%RSD_p$[a]
Turbidity (NTU)								
<2.0	28	1.1	0.07	6.5	27	1.0	0.08	8.6
2.0 to 20.0	38	8.1	0.26	3.2	36	7.7	0.39	5.0
20.0 to 100.0	1	33.5	0.71	2.1	2	24.0	1.41	5.9
All data	67	5.6	0.22	3.9	65	5.4	0.38	7.1

[a] $\%RSD_p$ = relative pooled standard deviation, calculated as (pooled s grand mean) x 100.

(largest $\%RSD_p$ = 4.0), specific conductance at the processing laboratory (largest $\%RSD_p$ = 3.3), specific conductance at the analytical laboratory (largest $\%RSD_p$ = 4.2), closed system DIC (largest $\%RSD_p$ = 4.4), fluoride (largest $\%RSD_p$ = 3.5), potassium (largest $\%RSD_p$ = 8.3), magnesium (largest $\%RSD_p$ = 2.6), sodium (largest $\%RSD_p$ = 3.6), nitrate (largest $\%RSD_p$ = 6.2), closed-system pH (largest s_p = 0.036 pH units), silica (largest $\%RSD_p$ = 11.3), sulfate (largest $\%RSD_p$ = 3.6), true color (largest s_p = 2.6 PCU), and turbidity (largest $\%RSD_p$ = 8.6). For these variables, random variations due to collection and processing appear to be minimal at all concentrations.

For six other variables, estimates of s_p or $\%RSD_p$ were not near the within-laboratory precision goal for one subrange. These variables were extractable aluminum ($\%RSD_p$ = 37.2 for the subrange 0.007 to 0.050 mg/L), nonexchangeable monomeric aluminum ($\%RSD_p$ = 16.2 for the subrange 0.023 to 0.10 mg/L), equilibrated DIC ($\%RSD_p$ = 18.1 for the subrange 0.23 to 1.00 mg/L), initial DIC ($\%RSD_p$ = 16.3 for the subrange 0.23 to 1.00 mg/L), DOC ($\%RSD_p$ = 20.8 for the subrange 0.5 to 2.0 mg/L), and manganese ($\%RSD_p$ = 65.0 for the subrange 0.01 to 0.05 mg/L).

For the extractable aluminum values outside the goal, the mean value (0.016 mg/L) indicates that routine-duplicate pairs in this subrange tended to be of low concentration. The s_p estimate (0.0059 mg/L) is approximately 12 percent of the upper limit of the subrange, and thus no data quality problem should exist for measurements approaching 0.050 mg/L. At lower concentrations, precision is probably affected by sample-to-sample differences in extraction efficiency at the processing laboratory.

For nonexchangeable monomeric aluminum, the majority of the duplicate pairs appears to have mean concentrations in the lower portion of the subrange between 0.023 and 0.100 mg/L (grand mean value = 0.042 mg/L). The observed precision is similar to that obtained for other aluminum measurements in the same general subrange of concentrations, and no data quality problem is indicated.

For the subranges of the two DIC determinations that were outside the precision goals, precision estimates at both laboratories for low concentrations (less than 1 mg/L) were outside the precision objective ($\%RSD_p$ = 16 to 22 for Laboratory 1; $\%RSD_p$ = 16 for Laboratory 2 for both variables). However, the s_p value was small (less than 0.1 mg/L), and a data quality problem is not indicated for this subrange.

For DOC, precision estimates at both laboratories ($\%RSD_p$ for Laboratory 1 = 27;

$\%RSD_p$ for Laboratory 2 = 17) were outside the precision objectives for concentrations less than 2.0 mg/L, suggesting an effect of collection, processing, or sample preparation. Variability in digestion efficiency, sporadic background contamination, or carry-over from a sample having a high DOC content into a sample having a low DOC content would all increase the variation of measurements within field routine-duplicate pair measurements. Interpretations of DOC measurements of approximately 2.0 mg/L or less should be conducted with this potential imprecision in mind.

For the manganese measurements that did not meet the precision objective, the grand mean value (0.02 mg/L) is near the limit of detection. Within this subrange (0.01 to 0.05, Table 27) one pair of measurements had a large difference (0.019 mg/L) between the measurements. Although removal of this pair improved precision estimates by approximately 25 percent, the precision within this subrange was still well outside the within-laboratory precision goals, indicating that measurements in this subrange are not reliable. Interpretation of streamwater concentrations in this subrange should be conducted with this variability in mind.

The three pH measurements determined at the analytical laboratories had s_p estimates (Table 28) that were not near the within-laboratory precision goals for several subranges. For pH-ANC and pH-BNC, s_p values were generally less than $\pm$ 0.1 pH unit. Other studies of pH measurements (Tyree, 1981; Davison and Gardner, 1986) report errors of this magnitude or greater within a laboratory in samples of low ionic strength unless stringent protocols are followed to minimize pH electrode errors. Thus, there is no reason to suspect a data quality problem in the pH-ANC and pH-BNC measurements. Equilibrated pH measurements in the circumneutral range (pH 6.00 to pH 8.00) had relatively high s_p estimates (0.113 to 0.171 pH units). Two sample pairs in the 7.00 to 8.00 subrange had large differences (0.8 to approximately 1.0 pH units) between the routine and duplicate measurements, resulting in large estimates of variance (5 to 8 times the next largest value). When the outlying pairs are removed, the value for S_p for the subrange improved from 0.171 to 0.094 pH units. When all measurements except the two outlying pairs were considered (n = 63), the s_p value based on all measurements improved from 0.144 to 0.090. This variation is probably the result of sample-to-sample differences in the carbon dioxide equilibration procedure. Interpretation of equilibrated pH measurements in the circumneutral range should consider this variability.

For total aluminum (largest $\%RSD_p$ = 37.8), iron (largest $\%RSD_p$ = 53.3), ammonium (largest $\%RSD_p$ = 23.4), and phosphorus (largest $\%RSD_p$ = 53.7), $\%RSD_p$ estimates for most or all subranges were not near within-laboratory precision objectives. For total aluminum, $\%RSD_p$ estimates for both laboratories for concentrations less than 0.5 mg/L were not near the precision objectives. Sample-to-sample differences in digestion appear to be the most likely source of the variance for total aluminum. For iron, $\%RSD_p$ values from both laboratories were also not near the precision objectives. For these two variables there appears to be a substantial effect from sample preparation, processing, or collection on precision.

Both laboratories also had $\%RSD_p$ estimates for ammonium measurements for concentrations less than 0.05 mg/L that were not near precision objectives. At concentrations between 0.05 and 0.10 mg/L, both laboratories had a $\%RSD_p$ estimate of approximately 23 to 24 percent. Precision estimates of $\%RSD_p$ for phosphorus measurements at both laboratories were not near precision objectives for any subrange. Measured values between 0.001 and 0.005 are close to the limit of detection and below the system decision limit (0.007 mg/L). At concentrations between 0.005 and 0.015 mg/L, precision was probably affected by background contamination or sample carryover. Good measurement precision for ammonium and phosphorus may not be achievable at these concentrations unless special precautions are taken during sample collection, processing, and preparation for analysis. Users of phosphorus data for the NSS-I should consider this variability.

Precision Estimates Derived From Audit Sample Measurements

For each type of audit sample (synthetic, Bagley Lake, and Big Moose Lake), individual measurements from both laboratories were combined to provide estimates of among-batch precision. For most variables, among-batch precision was estimated as the standard deviation of the combined audit sample measurements. Precision estimates for turbidity were not determined because the audit samples were filtered. Summary statistics based on the combined audit sample measurements are presented in Table 29.

For the synthetic audit samples, measurements of field and laboratory audit samples from both preparation lots were combined, except for extractable aluminum, total aluminum, and iron. Field audit and laboratory audit measurements were not significantly different for other variables (see preceding Accuracy subsection) and thus processing effects were minimal and did not contribute to among-batch variability. For extractable and total aluminum, the difference between lots was large enough to possibly inflate the precision estimate artificially; thus, precision estimates were calculated for each lot. For extractable aluminum, total aluminum, and iron, only laboratory audits were used because of the loss of aluminum and iron from field audit samples due to adsorption or precipitation. For the natural audit samples, field and laboratory samples were combined for all variables.

For each type of audit sample, several variables were present in very low concentrations, and analytical imprecision near the limit of detection is probably more influential than among-batch precision. For the synthetic audit sample, mean values for total monomeric (0.010 mg/L) and nonexchangeable monomeric aluminum (0.015 mg/L) were near the limits of detection (Table 16). In addition, the mean value for iron (0.03 mg/L) was also low, near the system decision limit (0.017 mg/L, Table 17). For the Bagley Lake audit sample, mean values for extractable aluminum (0.007 mg/L), total monomeric aluminum (0.01 mg/L), nonexchangeable aluminum (0.014 mg/L), iron (0.01 mg/L), manganese (<0.01 mg/L), ammonium (0.01 mg/L), nitrate (0.012 mg/L), and phosphorus (0.001 mg/L) were near or below limits of detection (Table 16). Mean values for total aluminum (0.027 mg/L), BNC (35.7 μeq/L), and DOC (0.4 mg/L) were near to the system decision limits (Table 17). In the Big Moose Lake sample, equilibrated DIC (mean = 0.09 mg/L) and phosphorus (mean = 0.002 mg/L) concentrations were near to or below limits of detection (Table 16).

As shown in Table 29, the major components of among-batch variability, assuming minimal preparation errors of the audit samples, will be interlaboratory differences and batch-to-batch differences within laboratories. Among-batch precision estimates for total monomeric aluminum, ANC, all DIC measurements, fluoride, potassium, magnesium, sodium, ammonium, nitrate, closed-system pH, and sulfate were within or near the within laboratory precision objectives for all audit sample types having mean concentrations of these variables greater than the system decision limit. Among-batch variability for these variables is not substantial for the range of concentrations represented by the audit samples.

Among-batch variability appears to be large for extractable aluminum at low concentrations (less than 0.030 mg/L). Examination of data from each laboratory (Tables 21 and 23) indicated that variability at both laboratories was large, and the most likely source of the variation is among-batch differences in extraction efficiency at low concentrations at the processing laboratory.

For total aluminum, among-batch variability is large at concentrations less than approximately 0.050 mg/L, and was observed for both synthetic and Bagley Lake audit samples (Table 29). Examination of data from both laboratories (Tables 21 through 24) indicates that variation in these measurements is large at both laboratories; among-batch differences due to the digestion procedure is the probable source of the variation.

Table 29. Summary Statistics for Among-Batch Precision Based on Pooled Audit Sample Data, National Stream Survey - Phase I

Variable	Units	Audit samples: Synthetic n[a]	Synthetic Mean	Synthetic s[b]	Bagley Lake n[a]	Bagley Lake Mean	Bagley Lake s[b]	Big Moose Lake n[a]	Big Moose Lake Mean	Big Moose Lake s[b]
Al-ext, Lot 14	mg/L	9[c]	0.024	0.0106	--	--	--	--	--	--
Al-ext, Lot 15	mg/L	26[c]	0.015	0.0030	38	0.007	0.0034	37	0.215	0.0329
Al-total, Lot 14	mg/L	19[d]	0.034	0.0222	--	--	--	--	--	--
Al-total, Lot 15	mg/L	32[d]	0.030	0.0075	38	0.027	0.0399	37	0.274	0.0255
Al-mono	mg/L	13[e]	0.010	0.0074	25[e]	0.013	0.0048	24[e]	0.195	0.0117
Al-nex	mg/L	13[e]	0.015	0.0080	25[e]	0.014	0.0045	24[e]	0.058	0.0127
ANC	μeq/l	56	113.9	9.24	38	127.8	11.73	39	-2.1	4.88
BNC	μeq/l	56	45.6	20.15	38	35.7	23.55	39	73.9	11.34
Ca	mg/L	56	0.19	0.031	39	1.57	0.065	39	1.90	0.077
Cl^-	mg/L	56	0.33	0.029	39	0.18	0.045	39	0.43	0.051
Cond-PL	μS/cm	12[e]	23.3	8.35	26	15.5	2.45	26	25.5	1.76
Cond-lab	μS/cm	56	18.5	1.58	38	13.1	1.91	39	24.8	0.89
DIC-closed	mg/L	14[e]	1.40	0.046	27	1.70	0.061	27	0.53	0.062
DIC-eq	mg/L	56	1.32	0.192	38	1.50	0.203	39	0.09	0.078
DIC-initial	mg/L	56	1.49	0.211	38	1.54	0.214	39	0.24	0.070
DOC	mg/L	56	1.1	0.27	39	0.4	0.42	39	3.9	0.28
F^-	mg/L	56	0.042	0.0026	39	0.025	0.0033	39	0.074	0.0024
Fe	mg/L	42[c]	0.03	0.018	39	0.01	0.017	39	0.05	0.017
K	mg/L	56	0.20	0.008	39	0.30	0.016	39	0.43	0.022
Mg	mg/L	56	0.43	0.032	39	0.17	0.016	39	0.32	0.021
Mn	mg/L	56	0.10	0.009	39	0.01	0.004	39	0.08	0.020
Na	mg/L	56	2.74	0.099	39	0.85	0.042	39	0.62	0.038
NH_4^+	mg/L	56	0.17	0.015	39	0.01	0.010	39	0.06	0.009
NO_3^-	mg/L	55	0.469	0.040	38	0.012	0.0189	39	1.228	0.0553
P	mg/L	53[f]	0.022	0.0039	39	0.001	0.0014	39	0.002	0.0022
pH-closed	pH units	14[d]	6.94	0.085	27	6.99	0.049	27	5.15	0.059
pH-ANC	pH units	56	6.80	0.187	38	7.01	0.100	39	5.19	0.159
pH-BNC	pH units	56	6.82	0.200	38	7.04	0.122	39	5.21	0.161
pH-eq	pH units	56	7.23	0.179	38	7.30	0.117	39	5.22	0.071
SiO_2	mg/L	56	1.01	0.143	39	9.32	0.537	39	4.25	0.419
SO_4^{2-}	mg/L	56	2.25	0.131	39	0.63	0.046	39	6.33	0.167
True color	PCU	12[d]	5	3.3	26	6	3.2	25	18	4.6

[a] n = number of measurements.

[b] s = standard deviation.

[c] Only laboratory audit samples were used for precision estimation.

[d] One significant outlier (Grubbs' test, $p \leq 0.05$; Grubbs, 1969) was excluded from the precision estimation.

[e] Only field audit samples were used for precision estimation.

[f] Three significant outliers (Grubbs test, $p \leq 0.05$; Grubbs, 1969) were excluded from the precision estimation.

For nonexchangeable monomeric aluminum, the Big Moose Lake audit samples provide the only estimate of among-batch precision (s = 0.0127 mg/L; Table 29) at a concentration greater than the system decision limit. The relative precision, even after removing three statistical outliers, was 22 percent at a mean concentration of 0.058 mg/L. Precision estimates derived from this methodology appear to be affected by day-to-day differences in performance. Further evaluations are needed to provide the level of precison required to interpret data in this range of concentrations on a routine basis.

Measurements of BNC have large estimates of among-batch variability (s = 11.34 to 23.55 μeq/L) for all three audit sample types (Table 29). The source of this variability is primarily a result of interlaboratory bias, as Laboratory 2 had measured values for BNC approximately twice those of Laboratory 1 for all types of audit samples (Tables 21 through 24).

Calcium measurements at low concen- trations (approximately 0.2 mg/L) may be affected by interlaboratory bias. Examination of data from each laboratory indicated that there was a consistent difference of approximately 0.02 mg/L between the two analytical laboratories for both lots of synthetic audit samples (Tables 21 and 23). Users of data from meas-urements of calcium at low concentrations should consider among-batch precision.

Chloride measurements at low concentrations (less than 0.2 mg/L, represented by the Bagley Lake sample) appear to be affected by among-batch variations (s = 0.045 mg/L). The source of this variation appears to be differences among batches within the laboratories and not interlaboratory bias (Tables 22 and 24).

The among-batch precision of specific conductance measurements at the processing laboratory was undoubtedly affected by the malfunctioning probe used for the first 28 batches. The majority of synthetic audit sample measurements were conducted by using the faulty probe. Measurements of natural audit samples had improved precision estimates (s = 1.76 to 2.45 μS/cm), which were close to the among-batch precision of analytical laboratory measurements (Table 26).

For specific conductance measurements at concentrations less than 20 μS/cm, among-batch precision estimates at both analytical laboratories were outside the within-laboratory precision objectives. The among-batch variation was also affected by interlaboratory bias (Tables 21 through 24).

Among-batch precision for laboratory DOC measurements at low concentrations (near 1 mg/L) appears to be affected by among-batch differences within the laboratories (especially Laboratory 1 for Lot 15, Table 21). A possible source of the variation is differences resulting from the sample digestion process at low concentrations.

For iron, among-batch precision of all concentrations represented by the audit samples (mean value = 0.05 mg/L or less) was poor. For the Big Moose Lake sample, which had the highest measured iron concentrations (mean = 0.05 mg/L), there were consistent differences between the mean values from the two laboratories (approximately 0.02 mg/L, Tables 22 and 24). Among-batch precision for synthetic field audit samples within Laboratory 2 (Table 23) was better than for Laboratory 1 (Table 21) as indicated by the larger confidence intervals for Laboratory 1.

Among-batch precision of manganese measurements is acceptable for all concentrations represented by the audit samples. In the Big Moose Lake audit sample, the precision estimate (s = 0.020 mg/L) was outside the within-laboratory precision objective. Measurements of manganese in the synthetic audit sample had a similar mean concentration (0.10 mg/L), but the precision estimate (s = 0.009) was within the within-laboratory precision objective. Two outlying values were present for the Big Moose Lake sample measurements (0.016 mg/L from a laboratory audit sample at Laboratory 2 and 0.172 mg/L from a field audit sample at Laboratory 1). The

among-batch precision estimated without these two values (n = 37, mean = 0.08 mg/L, s = 0.009 mg/L) provided a relative precision estimate (11%) that was near the within-laboratory precision objective.

Only the synthetic audit sample had concentrations of phosphorus above background levels. Initially, among-batch precision was large (n = 56, mean = 0.023 mg/L, s = 0.0069 mg/L, %RSD = 30.6). Examination of the measurements revealed that one measurement value was -0.0005 mg/L, while two other measurements were greater than 0.040 mg/L. Because these samples were all laboratory audit samples, the sample concentrations could have been affected during sample preparation at either the support laboratory or the analytical laboratory. When these three significant outliers (Grubbs test, $p \leq 0.05$; Grubbs, 1969) were excluded, the among-batch precision estimate improved by almost 50 percent (mean = 0.022 mg/L, s = 0.0039, %RSD = 17.7).

Estimates of among-batch precision for all of the analytical laboratory pH measurements were between 0.1 and 0.2 pH units. Differences between mean values from the two laboratories ranged from 0.1 to 0.3 pH units (Tables 21 through 24) and standard deviations of measurement within each laboratory ranged from approximately 0.1 to 0.15 pH units. This level of variation has been observed during other multilaboratory studies (Davison and Gardner, 1986) unless measurement protocols are stringently defined and followed. The primary purpose of the analytical laboratory pH measurements was to serve as a check on the ANC and BNC measurements on a sample-by-sample basis.

For silica, among-batch precision at low concentrations ($\simeq$1 mg/L, represented by the synthetic audit sample) was slightly greater (%RSD = 14.1) than the within-laboratory precision objectives. Among-batch precision within each laboratory was good for both types of natural audit samples (Tables 22 and 24). The negative bias observed for Laboratory 2 (Tables 23 and 24) appears to be the primary source of among-batch variation.

Comparison of Precision Estimates

Estimates of method-level (within-batch analytical), system-level (overall within-batch), and among-batch (including among-laboratory) precision were compiled from Tables 28 and 29 and are presented in Table 30. For subranges where more than one audit sample type was represented, the sample type having the largest standard deviation is presented in the table.

For all variables, the within-batch precision estimates presented here indicate that random errors occurring during sample preparation and analysis contributed only a small proportion to the overall measurement error during the NSS-I. For all variables except total aluminum, DOC, iron, and ammonium, among-batch precision estimates for most or all subranges were greater than corresponding system-level within-batch estimates. This difference is not totally unexpected, as the among-batch precision estimates are analogous to long-term standard deviations (or reproducibility), which are expected to be greater than a short-term standard deviation (or repeatability) (Taylor, 1987). This difference suggests that variation among batches within a laboratory or among laboratories (due to differences in calibrations or processing) contributes more to overall measurement error than does sample collection (including natural spatial variability in the stream over the 10-to-20 minute period during which samples were collected).

The most conservative estimates of measurement precision should be used for data interpretation activities. For total aluminum, DOC, iron, and ammonium measurements, system-level precision estimates were nearly the same as or larger than corresponding among-batch precision estimates. Sample-to-sample variability or collection effects are more important for these variables than day-to-day or among-laboratory variability in determining overall measurement error. For these four variables, system-level precision provides the most conservative estimate of overall measurement error and should be used for data interpretation. Among-batch

TABLE 30. Comparison of Method-Level, System-Level, and Among-Batch Precision Estimates, National Stream Survey - Phase I

Variable (units) and measurement range	Within-batch				Among-batch[c]		
	Method-level[a]		System-level[b]				
	Number of pairs	Pooled s[d]	Number of pairs	Pooled s[d]	Number of samples	s[e]	Sample type[f]
Al-ext (mg/L)							
<0.007	3	0.0002	29	0.0016	38	0.0034	BL
0.007 to 0.050	35	0.0009	25	0.0059	9	0.0106	S
0.050 to 0.100	15	0.0026	4	0.0021	--	--	
>0.100	15	0.0055	7	0.0166	37	0.0329	BM
Al-total (mg/L)							
<0.027	9	0.0018	4	0.0035	38	0.0399	BL
0.027 to 0.100	6	0.0034	16	0.0079	19	0.0222	S
0.100 to 0.500	51	0.0042	38	0.0807	37	0.0255	BM
0.500 to 1.000	1	0.0304	7	0.1953	--	--	
>1.000	1	0.0573	0	--	--	--	
Al-mono (mg/L)							
<0.015	33	0.0011	30	0.0022	13	0.0074	S
0.015 to 0.10	24	0.0017	29	0.0038	--	--	
0.10 to 0.50	7	0.0033	6	0.0049	24	0.0117	BM
0.50 to 1.00	2	0.0045	0	--	--	--	
>1.00	1	0.0030	0	--	--	--	
Al-nex (mg/L)							
<0.023	52	0.0022	47	0.0031	13	0.0080	S
0.023 to 0.10	12	0.0033	18	0.0068	24	0.0127	BM
0.10 to 0.50	3	0.0139	0	--	--	--	
ANC (μeq/L)							
<0	2	0.70	7	2.94	39	4.88	BM
0 to 50	4	1.05	9	7.83	--	--	
>50	62	4.42	49	11.22	38	11.73	BL
BNC (μeq/L)							
0 to 50	24	2.68	28	6.38	38	23.55	BL
>50	44	6.72	37	12.25	39	11.34	BM
Ca (mg/L)							
0.02 to 1.00	6	0.009	4	0.015	56	0.031	S
1.00 to 5.00	35	0.041	35	0.057	39	0.077	BM
5.00 to 10.00	12	0.034	19	0.090	--	--	
>10.00	15	0.529	7	1.604	--	--	

(Continued)

Table 30. (Continued)

Variable (units) and measurement range	Within-batch: Method-level[a] Number of pairs	Method-level[a] Pooled s[d]	System-level[b] Number of pairs	System-level[b] Pooled s[d]	Among-batch[c] Number of samples	Among-batch[c] s[e]	Sample type[f]
Cl^- (mg/L)							
<0.03	2	0.000	0	--	--	--	
0.03 to 1.00	13	0.006	8	0.018	39	0.051	BM
1.00 to 2.00	13	0.022	18	0.033	--	--	
2.00 to 5.00	23	0.045	24	0.043	--	--	
5.00 to 10.00	11	0.077	6	0.073	--	--	
>10.00	6	0.336	9	0.734	--	--	
Cond-PL (µS/cm)							
<25.0	5	1.11	5	0.26	12	8.35	S
25.0 to 50.0	22	1.94	24	1.26	26	1.76	BM
50.0 to 100.0	21	5.27	24	1.32	--	--	
>100.0	19	16.99	12	3.12	--	--	
Cond-lab (µS/cm)							
<25.0	14	0.12	6	0.87	38	1.91	BL
25.0 to 50.0	20	0.23	24	0.27	39	0.89	BM
50.0 to 100.0	17	0.29	24	0.42	--	--	
>100.0	17	0.49	11	1.22	--	--	
DIC-closed (mg/L)							
<1.00	9	0.025	7	0.020	27	0.062	BM
1.00 to 2.00	8	0.040	9	0.063	27	0.061	BL
2.00 to 5.00	26	0.068	31	0.092	--	--	
5.00 to 10.00	18	0.168	13	0.184	--	--	
>10.00	7	0.236	5	0.365	--	--	
DIC-eq (mg/L)							
<0.23	1	0.002	10	0.054	39	0.078	BM
0.23 to 1.00	4	0.029	12	0.099	--	--	
1.00 to 2.00	8	0.031	12	0.162	38	0.203	BL
2.00 to 5.00	24	0.067	22	0.261	--	--	
5.00 to 10.00	18	0.086	5	0.184	--	--	
>10.00	13	0.639	4	1.050	--	--	
DIC-init (mg/L)							
<0.23	0	--	3	0.014	--	--	
0.23 to 1.00	1	0.007	10	0.091	39	0.070	BM
1.00 to 2.00	17	0.036	15	0.211	38	0.214	BL
2.00 to 5.00	18	0.111	26	0.250	--	--	
5.00 to 10.00	18	0.139	6	0.050	--	--	
>10.00	14	0.780	5	1.009	--	--	

(Continued)

Table 30. (Continued)

Variable (units) and measurement range	Within-batch				Among-batch[c]		
	Method-level[a]		System-level[b]				
	Number of pairs	Pooled s[d]	Number of pairs	Pooled s[d]	Number of samples	s[e]	Sample type[f]
DOC (mg/L)							
<0.5	4	0.02	2	0.02	39	0.42	BL
0.5 to 2.0	27	0.05	41	0.24	56	0.27	S
2.0 to 5.0	16	0.11	15	0.44	39	0.28	BM
5.0 to 10.0	14	0.15	7	0.58	--	--	
>10.0	7	0.15	0	--	--	--	
F^- (mg/L)							
0.010 to 0.050	29	0.0000	51	0.0011	39	0.0033	BL
>0.050	39	0.0021	14	0.0029	39	0.0024	BM
Fe (mg/L)							
<0.02	2	0.001	14	0.003	39	0.017	BL
0.02 to 0.05	3	0.001	18	0.009	42	0.018	S
0.05 to 0.10	11	0.002	10	0.029	39	0.017	BM
0.10 to 0.50	36	0.004	18	0.117	--	--	
0.50 to 1.00	3	0.008	3	0.163	--	--	
>1.00	12	0.114	2	0.377	--	--	
K (mg/L)							
<0.15	1	0.001	1	0.003	--	--	
0.15 to 0.35	10	0.006	5	0.003	39	0.016	BL
0.35 to 0.45	2	0.002	6	0.005	39	0.022	BM
>0.45	55	0.015	53	0.063	--	--	
Mg (mg/L)							
<1.00	26	0.004	19	0.013	56	0.032	S
1.00 to 2.00	21	0.013	23	0.014	--	--	
2.00 to 5.00	15	0.033	18	0.058	--	--	
>5.00	6	0.158	5	0.242	--	--	
Mn (mg/L)							
<0.01	2	0.001	19	--	39	0.004	BL
0.01 to 0.05	9	0.001	19	0.015	--	--	
0.05 to 0.10	12	0.003	7	0.002	39	0.020	BM
>0.10	44	0.017	20	0.023	56	0.009	S
Na (mg/L)							
<0.50	4	0.001	4	0.001	--	--	
0.50 to 1.00	8	0.027	4	0.030	39	0.042	BL
1.00 to 2.00	11	0.017	24	0.028	--	--	
2.00 to 5.00	33	0.028	25	0.037	56	0.099	S
>5.00	12	0.174	8	0.097	--	--	

(Continued)

Table 30. (Continued)

Variable (units) and measurement range	Within-batch: Method-level[a] Number of pairs	Method-level[a] Pooled s[d]	System-level[b] Number of pairs	System-level[b] Pooled s[d]	Among-batch[c] Number of samples	Among-batch[c] s[e]	Sample type[f]
NH_4^+ (mg/L)							
<0.02	12	0.000	33	0.006	39	0.010	BL
0.02 to 0.05	20	0.001	21	0.007	--	--	
0.05 to 0.10	14	0.002	7	0.014	39	0.009	BM
>0.10	22	0.004	4	0.011	56	0.015	S
NO_3^- (mg/L)							
<3.000	55	0.0162	53	0.0422	39	0.0553	BM
>3.000	13	0.1212	12	0.3239	--	--	
P (mg/L)							
<0.001	2	0.0003	7	0.0009	39	0.0014	BL
0.001 to 0.005	24	0.0003	31	0.0016	39	0.0022	BM
0.005 to 0.015	20	0.0004	20	0.0034	--	--	
>0.015	22	0.0018	7	0.0040	53	0.0039	S
pH-closed (pH units)							
<4.00	4	0.000	0	--	--	--	
4.00 to 5.00	4	0.012	6	0.032	--	--	
5.00 to 6.00	7	0.041	13	0.036	27	0.059	BM
6.00 to 7.00	25	0.012	24	0.036	14	0.085	S
7.00 to 8.00	26	0.028	19	0.036	27	0.049	BL
>8.00	2	0.010	3	0.026	--	--	
pH-ANC (pH units)							
4.00 to 5.00	1	0.014	6	0.022	--	--	
5.00 to 6.00	5	0.021	8	0.104	39	0.159	BM
6.00 to 7.00	34	0.053	27	0.073	56	0.187	S
7.00 to 8.00	28	0.039	22	0.091	38	0.100	BL
>8.00	0	--	2	0.034	--	--	
pH-BNC (pH units)							
4.00 to 5.00	0	--	6	0.035	--	--	
5.00 to 6.00	6	0.034	8	0.092	39	0.161	BM
6.00 to 7.00	34	0.042	26	0.065	56	0.200	S
7.00 to 8.00	28	0.040	23	0.086	38	0.122	BL
>8.00	0	--	2	0.062	--	--	
pH-eq (pH units)							
<4.00	2	0.005	0	--	--	--	
4.00 to 5.00	7	0.009	6	0.017	--	--	
5.00 to 6.00	4	0.000	4	0.041	39	0.071	BM
6.00 to 7.00	3	0.004	11	0.113	--	--	
7.00 to 8.00	34	0.019	40	0.171	56	0.179	S
>8.00	18	0.028	4	0.083	--	--	

(Continued)

Table 30. (Continued)

Variable (units) and measurement range	Within-batch				Among-batch[c]		
	Method-level[a]		System-level[b]				
	Number of pairs	Pooled s[d]	Number of pairs	Pooled s[d]	Number of samples	s[e]	Sample type[f]
SiO_2 (mg/L)							
<0.50	1	0.007	0	--	--	--	
0.50 to 1.50	1	0.035	1	0.064	56	0.143	S
1.50 to 5.50	29	0.052	25	0.071	39	0.419	BM
5.50 to 10.50	19	0.141	21	0.122	39	0.537	BL
>10.50	18	0.195	18	0.730	--	--	
SO_4^{2-} (mg/L)							
<0.05	2	0.002	0	--	--	--	
0.05 to 2.50	13	0.019	18	0.054	56	0.131	S
2.50 to 5.00	18	0.048	10	0.081	--	--	
5.00 to 12.50	26	0.117	24	0.128	39	0.167	BM
>12.50	9	0.239	13	0.243	--	--	
True color (PCU)							
<30	50	0.9	44	2.6	25	4.6	BM
>30	17	8.7	21	1.9	--	--	
Turbidity (NTU)							
<2.0	28	0.07	27	0.08	--	--	
2.0 to 20.0	38	0.26	36	0.39	--	--	
20.0 to 100.0	1	0.71	2	1.41	--	--	

[a] Method-level = within-batch precision calculated from laboratory routine-duplicate sample pairs.
[b] System-level = overall within-batch precision calculated from field routine-duplicate sample pairs.
[c] Among-batch = among-batch precision (across laboratories) calculated from audit sample measurements.
[d] Pooled s = pooled standard deviation.
[e] s = standard deviation.
[f] Audit sample types from which precision estimate was derived: S = synthetic audit sample. BL = Bagley Lake sample, and BM = Big Moose Lake sample.

precision estimates for all other variables provide the most conservative estimate of overall measurement error available for the NSS-I.

Discussion and Summary: Precision

The grouping of measurements of routine-duplicate pairs into concentration subranges generally provides more useful information related to data quality than a single estimate of relative precision. However, there appear to be two limitations to this approach that must be considered. Unless there is existing information to allow field duplicate samples to be selected so that each subrange is represented by an adequate number of pairs, data for some subranges may be scarce. However, if samples are duplicated more or less at random, then the distribution of pairs among subranges should reflect the distribution of routine sample concentrations.

A more serious concern is that precision estimates based on pooled variances of sample pairs can be extremely sensitive to outlying values, as was observed for the equilibrated pH measurements, where removal of two values improved the precision estimate by almost 50 percent.

The utility of duplicate measurements is limited in that they do not provide an estimate of among-batch variability, as do the audit samples. Likewise, audit samples do not provide an estimate of random errors associated with sample collection and handling and are concentration-specific. A series of split samples prepared from a single bulk sample may provide a more useful design to monitor and assess total measurement uncertainty. The split samples could then be used to assess various components of within-batch precision, among-batch precision (by withholding a split until the next batch of samples is analyzed), and even among-laboratory precision (by sending splits to other laboratories for analysis within holding times). Natural audit samples of appropriate concentrations that have been well characterized chemically could also be taken to the field and processed through the sampling device to provide estimates of total measurement uncertainty.

For many variables, among-batch precision estimates are not available for subranges at high concentrations. The audit sample program used during the NSS-I was designed to control measurement processes at concentrations representative of surface waters sensitive to or impacted by acidic deposition. It cannot be said with certainty that measurement error will be reduced at concentrations greater than those represented by the NSS-I audit samples. Among-batch precision within a laboratory would probably improve at higher concentrations, but interlaboratory differences may be present at all concentrations. For those variables for which interlaboratory bias was indicated as a primary influence on among-batch precision (e.g., specific conductance), the among-batch precision estimate available for a lower concentration may pro-vide a more conservative estimate of overall measurement precision at higher concentrations than a system-level precision estimate. For future studies, the audit samples chosen should bracket the range of sample concentrations to the greatest extent feasible.

Summary of Data Quality Assessment

The overall data quality of the NSS-I analytical results, in terms of the data quality objectives established for the project, is adequate to achieve the project objectives. The samples collected are representative of the types of stream resources of interest to the project. The data base is sufficiently complete, in terms of the number of samples providing valid data, to allow the estimation of population sizes. Based on a comparison of audit sample measurements (see Appendix A), there is high confidence that the data collected and analyzed during the NSS-I are comparable and appear to be compatible with other NSWS data bases. A general summary of the detectability, accuracy, and precision of the analytical measurements is presented in Table 31. There are only a few cases where use of the data to meet the project objectives may be limited by data quality considerations. In most of these cases, the limitation affects only interpretation of measurements at low concentrations. More general limitations are summarized below:

1. Specific conductance measurements made at the processing laboratory during the first half of the NSS-I are suspect, because of systematic errors related to a faulty probe. Analytical laboratory measurements of specific conductance should be used for all interpretative activities.

2. Estimates of exchangeable monomeric aluminum, calculated as the difference between total monomeric and nonexchangeable monomeric aluminum, may be negative for some samples because the background levels of nonexchangeable monomeric aluminum are higher than those observed for total monomeric

Table 31. Summary of Data Quality Assessment for Chemical Variables with Respect to Detectability, Accuracy, and Precision, National Stream Survey - Phase I[a]

Variable	Detectability	Accuracy	Precision
Al-ext	+	+	-[b]
Al-total	+	-[b]	-[c]
Al-mono	+	+	-[b]
Al-nex	-[b]	+	-[b,e]
ANC	+	+	+
BNC	-[b]	-[b]	-[b]
Ca	+	+	+
Cl^-	+	+	-[b]
Cond-PL	-[d]	-[d,e]	-
Cond-lab	+	-[b]	+
DIC-closed	+	+	+
DIC-eq	+	+	+
DIC-init	+	+	+
DOC	+	-[b]	-[b,c]
F^-	+	-[b]	-[b,c]
Fe	+	+	+
K	+	+	+
Mg	+	+	+
Mn	+	+	-[b]
Na	+	+	+
NH_4^+	+	+	-[b,c]
NO_3^-	+	+	+
P	+	-[e]	-[b,c]
pH-closed	+	+	+
pH-ANC	+	+	+
pH-BNC	+	+	+
pH-eq	+	+	+
SiO_2	+	-[b]	+
SO_4^{2-}	+	+	+
True color	NE	+	+
Turbidity	NE	NE	+

[a] + = acceptable in terms of data quality objective or primary project objectives.
- = estimate not near data quality objective.
NE = not evaluated.

[b] Possible limitations at low concentrations.

[c] Overall within-batch precision estimate is larger than among-batch precision estimate.

[d] Estimates not of acceptable quality.

[e] Possible limitations at high concentrations.

aluminum. These negative values will result in an increased uncertainty in the estimate of exchangeable monomeric aluminum, especially at lower concentrations. In addition, the among-batch precision of nonexchangeable monomeric aluminum measurements may not be of adequate quality to effectively interpret patterns in aluminum chemistry.

3. At low concentrations of BNC, data interpretations may be limited because of random and systematic errors caused by changes in dissolved carbon dioxide concentration in the sample during collection, transport, or the titration procedure.

4. Although all DIC and pH measurements provided acceptable results, the closed-system measurements are less subject to changes in dissolved carbon dioxide concentration between the time of collection and analysis, and thus provide the best estimates of in situ conditions at the time of sampling.

5. Data users interested in phosphorus measurements should consider the possibility of a negative bias from Laboratory 1 during the last half of the NSS-I.

6. For total aluminum, DOC, iron, and ammonium, sample-to-sample variability within a batch was apparently larger than sample-to-sample variability among batches. For these four variables, the system-level precision estimates provide the most conservative estimate of meas-urement uncertainty available for the NSS-I. For all other variables, the estimates of among-batch precision are the best available for the evaluation of measurement uncertainty.

In addition to the variable-by-variable evaluations, three checks of overall data quality were related to interpretation of acidification effects: comparison of the total ionic charge of cations and anions; comparison of calculated and measured specific conductance values; and comparison of calculated carbonate alkalinity and measured values of ANC.

Charge Balances

For each sample the sum of concentrations (expressed in μeq/L) was calculated for both major cations and major anions. For cations, the summary value was the total of calcium, magnesium, sodium, potassium, ammonium, and hydrogen ion concentrations. For anions, the summary value was the total of sulfate, nitrate, chloride, fluoride, ANC, and hydrogen ion concentrations. Use of ANC and hydrogen ion concentrations in the equation accounted for bicarbonate and carbonate ions, hydroxide ions, and noncarbonate anions such as organic anions and metal oxides.

Because of the electroneutrality constraint, the sum of cations must equal the sum of anions. Realistically, the charge balance will never equal zero, because analytical errors in each of the measured cations or anions are additive in the summation process. In addition, the analytical errors due to different methodologies will be different for the various cations and anions.

The sum of cations is plotted against the sum of anions for each sample in Figure 13. Values from the enhanced data set were used for this comparison. The agreement between the sum of cations and the sum of anions is excellent, with 94.7 percent of the samples (n = 1,342) falling within 10 percent of the line of 1:1 correspondence (representing the values at which the sum of cations equals the sum of anions). A total of 88.4 percent of the samples fell within 5 percent of the line of 1:1 correspondence. The observed pattern of deviation from the identity line suggests either an underestimate of the sum of anions (due to systematic errors in measurement or to unmeasured anions) or an overestimate of the sum of cations. As no serious systematic errors were observed for measured anions or cations, the most likely reason for the observed pattern is unmeasured anions such as organics (indirectly measured as DOC), aluminum, or metals present in samples

collected from polluted streams (e.g., those receiving mine effluents).

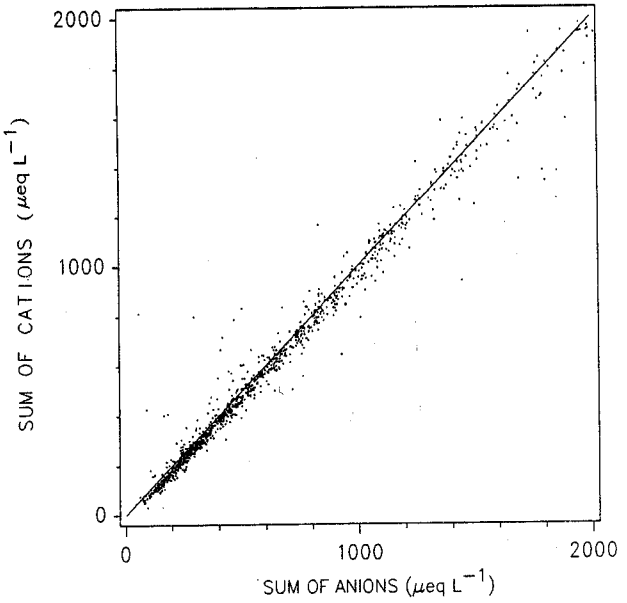


Figure 13. Sum of cations versus sum of anions, for routine samples, National Stream Survey - Phase I. (From Kaufmann et al., In press; line represents 1:1 correspondence, where sum of cations equals sum of anions).

Specific Conductance Check

Comparison of the calculated and measured values for specific conductance provides an additional check on analytical errors in measurements or the presence of unmeasured ionic species. The calculated conductance for each major ion was estimated using the Debye-Huckel-Onsager equation (Atkins, 1978), which corrects for concentration effects:

$$M_c = M^0_c - (A + BM^0_c) \times C^{\frac{1}{2}}$$

where M_c = corrected molar conductance at 25 °C
M^0_c = molar conductance at infinite dilution
A = 60.2
B = 0.229
C = molar concentration of ion

Figure 14 presents a plot of the measured versus calculated specific conductance for each routine sample. In general, the agreement of the two estimates is excellent for most samples. The linear regression equation that best described the observed relationship was:

Measured conductance = 0.981 (calculated conductance) + 4.87

This equation explained 98.7 percent of the total variance. Both the slope (0.981, standard error is $\pm$ 0.0031) and intercept (4.87, standard error is $\pm$ 0.522) indicated a deviation from the line of 1:1 correspondence (where measured conductance equals calculated conductance). Some of the deviation is probably due to systematic errors in measurement, at least at low values of specific conductance (see Accuracy section). Unmeasured ions, as indicated by the charge balance comparisons (Figure 13) also apparently affected the estimate of calculated conductance in some cases, because the measured conductance was larger than the calculated conductance.

Comparison of ANC Values

Comparison of an ANC value measured in a sample to that predicted assuming a carbonate system provides an indication of the reliability of pH, DIC, and ANC measurements, as well as the presence of unmeasured noncarbonate protolytes. Carbonate alkalinity represents the contributions to ANC from bicarbonate, carbonate, hydroxide, and hydrogen ions. Carbonate alkalinity, or calculated ANC, was estimated from measured values of pH-BNC and initial DIC (Hillman et al., 1987):

$$ANC_c = \left[\frac{[DIC]}{12{,}011}\left(\frac{[H^+]\,K_1 + 2\,K_1\,K_2}{[H^+]^2 + [H^+]\,K_1 + K_1\,K_2}\right) + \frac{K_w}{[H^+]} - [H^+]\right] \times 10^6$$

where ANC_c = calculated acid-neutralizing capacity

DIC = DIC-initial value in mg/L

$[H^+]$ = $10^{-(pH)}$ where pH = value from pH-BNC measurement

K_1 = 4.4463×10^{-7} at 25 °C

K_2 = 4.6881×10^{-11} at 25 °C

K_w = 1.0023×10^{-14} at 25 °C

For this relationship, the measured ANC should be greater than or equal to the carbonate alkalinity; otherwise, analytical errors may have influenced measurements of pH, DIC, or ANC.

Carbonate alkalinity values are plotted against measured ANC values in Figure 15. The majority of the data points fell below the line of 1:1 correspondence which shows that measured ANC is greater than carbonate alkalinity. Thus, measurements of pH-BNC, DIC-initial, and ANC appear to be reliable for most samples. The observed differences between carbonate alkalinity values and measured ANC values are apparently due to the presence of noncarbonate protolytes, such as DOC, aluminum, dissolved metal complexes, or particulate metal oxides.

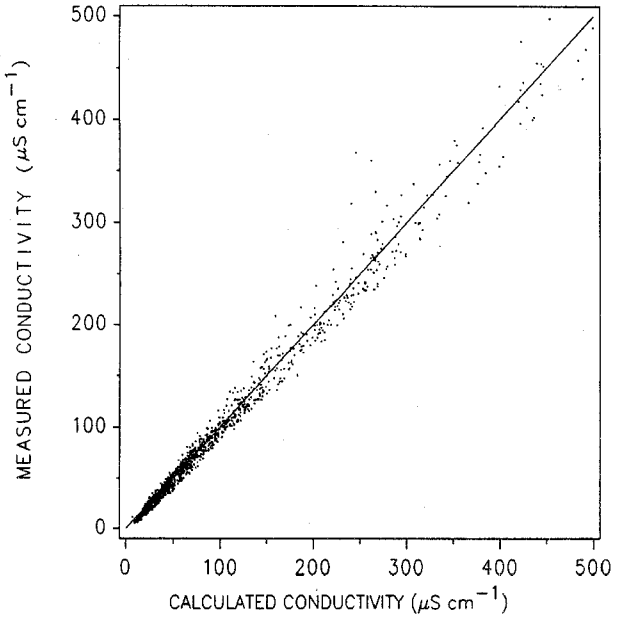


Figure 14. **Measured versus calculated specific conductance at 25°C for routine samples, National Stream Survey- Phase I (from Kaufmann et al., in press; line represents 1:1 correspondence, where measured conductance equals calculated conductance).**

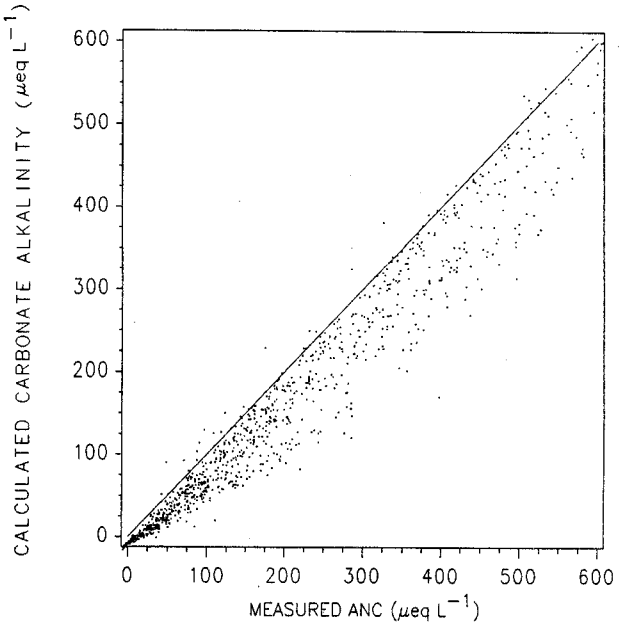


Figure 15. **Calculated carbonate akalinity versus measured ANC for routine samples, National Stream Survey - Phase I (from Kaufmann et al., in press; line represents 1:1 correspondence line, where carbonate alkalinity equals measured ANC).**

REFERENCES

American Chemical Society, Committee on Environmental Improvement. 1980. Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry. Analytical Chemistry, 52:2242-2249.

APHA (American Public Health Association), American Water Works Association, and Water Pollution Control Federation. 1985. Standard Methods for the Examination of Water and Wastewater, 16th Ed. APHA, Washington, D. C.

Arent, L. J., M. O. Morison, and C. S. Soong. In preparation. Eastern Lake Survey - Phase II and National Stream Survey - Phase I: Laboratory Operations Report. EPA/600/4-88/025. U. S. Environmental Protection Agency, Las Vegas, Nevada.

ASTM (American Society for Testing and Materials). 1984. Annual Book of ASTM Standards. Standard Specification for Reagent Water. D1193-77 (reapproved 1983). Vol. 11.01. ASTM, Philadelphia, Pennsylvania.

Atkins, P. W., 1978. Physical Chemistry. W. H. Freeman and Co., San Francisco, California.

Blick, D. J., J. J. Messer, D. H. Landers, and W. S. Overton. 1987. Statistical Basis for the Design and Interpretation of the National Surface Water Survey, Phase I: Lakes and Streams. Lake and Reservation Management. Vol. 3:470-475.

Burke, E. M., and D. C. Hillman. 1987. Syringe Sample Holding Time Study. Appendix A in C. M. Knapp, C. L. Mayer, D. V. Peck, J. R. Baker, and G. J. Filbin. National Surface Water Survey, National Stream Survey, Pilot Survey: Field Operations Report. EPA/600/8-87/019. U. S. Environmental Protection Agency, Las Vegas, Nevada.

Campbell, S., and H. Scott. 1985. Quality Assurance in Acid Precipitation Measurements. Pp. 272-283. In: J. K. Taylor and T. W. Stanley (eds.). Quality Assurance for Environmental Measurements. ASTM STP 867. American Society for Testing and Materials, Philadelphia, Pennsylvania.

Clayton, C. A., J. W. Hines, and P. D. Elkins. 1987. Detection Limits with Specified Assurance Properties. Analytical Chemistry, 59:2506-2514.

Currie, L. A. 1968. Limits of Qualitative Detection and Quantitative Determination: Application to Radiochemistry. Analytical Chemistry, 40:586-593.

Davison, W., and M. J. Gardner. 1986. Interlaboratory Comparisons of the Determinations of pH in Poorly-Buffered Fresh Waters. Analytica Chimica Acta, 182:17-31.

Drousé, S. K., D. C. Hillman, L. W. Creelman, J. F. Potter, and S. J. Simon. 1986a. National Surface Water Survey, Eastern Lake Survey - Phase I: Quality Assurance Plan. EPA/600/4-86/008, U. S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Drousé, S. K., D. C. Hillman, J. L. Engels, L. W. Creelman, and S. J. Simon, 1986b. National Surface Water Survey, National Stream Survey (Phase I - Pilot, Mid-Atlantic Phase I, Southeast Screening, and Episodes Pilot): Quality Assurance Plan. EPA/600/4-86/044. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Drousé, S. K. 1987. The National Surface Water Survey, National Stream Survey, Phase I - Pilot Survey: Summary of Quality Assurance Data Results. EPA/600/8-87/057. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Eshleman, K. N. In press. Predicting Regional Episodic Acidification of Surface Water Using Empirical Techniques. Water Resources Research.

Glaser, J. A., D. L. Forest, G. D. McKee, S. A. Quae, and W. L. Budde. 1981. Trace Analyses of Wastewaters. Environmental Science and Technology. 15:1426-1435.

Grosser, S. C., A. K. Pollack. 1987. National Stream Survey Data Base Audit. Final Report. Systems Applications, Inc., San Rafael, California

Grubbs, F. E. 1969. Procedures for Detecting Outlying Observations in Samples. Technometrics, 11:1-21.

Hagley, C. A., C. L. Mayer, and R. Hoenicke. In press. National Stream Survey, Phase I, Field Operations Report. EPA/600/4-88/023. U.S. Environmental Protection Agency, Las Vegas, Nevada.

Hillman, D. C., S. H. Pia, and S. J. Simon. 1987. National Surface Water Survey, Stream Survey (Pilot, Middle-Atlantic Phase I, Southeast Screening, and Middle-Atlantic Episode Pilot): Analytical Methods Manual. EPA/600/8-87/005. U.S. Environmental Protection Agency, Las Vegas, Nevada.

Hubaux, A., and G. Vos. 1970. Decision and Detection Limits for Linear Calibration Curves. Analytical Chemistry. 42:849-855.

Hunt, D. T. E., and A. L. Wilson. 1986. The Chemical Analysis of Water: General Principles and Techniques. The Royal Society of Chemistry, Burlington House, London.

Kaufmann, P., A. Herlihy, J. Elwood, M. Mitch, S. Overton, M. Sale, J. Messer, K. Reckhow, K. Cougan, D. Peck, J. Coe, A. Kinney, S. Christie, D. Brown, C. Hagley, and Y. Jager. 1988. Chemical Characteristics of Streams in the Mid-Atlantic and Southeastern United States. Volume I: Population Descriptions and Physico-Chemical Relationships. EPA/600/3-88/021a U. S. Environmental Protection Agency, Washington, D. C.

Keith, L. H., W. Crummett, J. Deegan, Jr., R. A. Libby, J. K. Taylor, and G. Wentler. 1983. Principles of Environmental Analysis. Analytical Chemistry. 55:2210-2218.

Knapp, C. M., C. L. Mayer, D. V. Peck, J. R. Baker, and G. J. Filbin. 1987. National Surface Water Survey: National Stream Survey, Pilot Survey: Field Operations Report. EPA/600/8-87/019. U.S. Environmental Protection Agency, Las Vegas, Nevada. 94 pp.

Landers, D. H., J. M. Eilers, D. F. Brakke, W. S. Overton, P. E. Kellar, M. E. Silverstein, R. D. Schonbrod, R. E. Crowe, R. A. Linthurst, J. M. Omernik, S. A. Teague, and E. P. Meier. 1987. Characteristics of Lakes in the Western United States. Volume I: Population Descriptions and Physico-Chemical Relationships. EPA/600/3-86/054a. U.S. Environmental Protection Agency, Washington, D.C. 176 pp.

Linthurst, R. A., D. H. Landers, J. M. Eilers, D. F. Brakke, W. S. Overton, E. P. Meier, and R. E. Crowe. 1986. Characteristics of Lakes in the Eastern United States. Volume I: Population Descriptions and Physico-Chemical Relationships. EPA/600/4-86/007a. U.S. Environmental Protection Agency, Washington, D.C. 136 pp.

Long, G. L., and J. D. Winefordner. 1983. Limit of Detection: A Closer Look at the IUPAC Definition. Analytical Chemistry. 55:712A-724A.

McQuaker, N. R., P. D. Kluckner, and D. K. Sandberg, 1983. Chemical Analysis of Acid Precipitation: pH and Acidity Determinations. Environmental Science and Technology. 17(7):431-435.

Mericas, C. E., and R.D. Schonbrod. 1987. Measurement Uncertainty in the National Surface Water Survey. Lake and Reservoir Management. 3:488-497.

Messer, J. J., C. W. Ariss, J. R. Baker, S. K. Drousé, K. N. Eshleman, P. R. Kaufmann, R. A. Linthurst, J. M. Omernik, W. S. Overton, M. J. Sale, R. D. Schonbrod, S. M. Stambaugh, and J. R. Tuschall, Jr. 1986. National Surface Water Survey, National Stream Survey, Phase I - Pilot Survey. EPA/600/4-86/026. U.S. Environmental Protection Agency, Washington, D.C.

Miller, J. C., and J. N. Miller. 1984. Statistics for Analytical Chemistry. Ellis Horwood Limited, Chichester, England. 201 pp.

Oak Ridge National Laboratory. 1984. National Surface Water Survey Project - Data Management Proposal. Environmental Sciences Division and Computer Sciences, UCC-ND, ORNL, Oak Ridge, Tennessee.

Oliver, B. G., E. M. Thurman, and R. K. Malcolm. 1983. The Contribution of Humic Substances to the Acidity of Colored Natural Waters. Geochemica et Cosmochemica Acta. 47:2031-2035.

Oppenheimer, L., T. P. Capizzi, R. M. Weppelman, and Hina Mehta. 1983. Determining the Lowest Limit of Reliable Assay Measurement. Analytical Chemistry. 55:638-643.

Overton, W. S. 1986. A Sampling Plan for Streams in the National Surface Water Survey. Technical Report 114. Department of Statistics, Oregon State University, Corvallis, Oregon.

Overton, W. S. 1987. A Sampling and Analysis Plan for Streams in the National Surface Water Survey Conducted by EPA. Technical Report 117. Department of Statistics, Oregon State University, Corvallis, Oregon.

Peden, M. E. 1981. Sampling Analytical and Quality Assurance Protocols for the National Atmospheric Deposition Program. ASTM D-22 Symposium and Workshop on Sampling and Analysis of Rain. ASTM, Philadelphia, Pennsylvania.

Permutt, T. J. and A. K. Pollack. 1986. Analysis of Quality Assurance Data for the Eastern Lake Survey. Appendix A in M. D. Best, S. K. Drousé, L. W. Creelman and D. J. Chaloud. National Surface Water Survey, Eastern Lake Survey (Phase I - Synoptic Chemistry): Quality Assurance Report. EPA 600/4-86/011. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

Royset, O. 1986. Flow-injection Spectrophotometric Determination of Aluminum in Water with Pyrocatechol Violet. Analytica Chimica Acta. 185:75-81.

Rogeberg, E. J. S., and A. Henriksen. 1985. An Automatic Method for Fractionation and Determination of Aluminum Species in Freshwaters. Vatten. 41:48-53.

Sale, M. J. (ed.). In press. Data Management and Analysis Procedures for the National Stream Survey. ORNL/TM-XXXX. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

SAS Institute, Inc. 1985. SAS User's Guide: Statistics, Version 5 Edition. SAS Institute, Inc., Cary, North Carolina. 956 pp.

Skougstad, M. W., J. Fishman, L. C. Friedman, D. E. Erdmary and S. S. Duncan (eds.). 1979. Methods for Determination of Inorganic Substances in Water and Fluvial Sediments: Techniques of Water-Resources Investigations of the United States Geological Survey, Book 5, Chapter A1. U.S. Government Printing Office, Washington, D.C.

Silverstein, M. E., M. L. Faber, S. K. Drousé, and T. E. Mitchell-Hall. 1987. National Surface Water Survey, Western Lake Survey (Phase I--Synoptic Chemistry): Quality Assurance Report. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. 292 pp.

Stanley, T. W., and S. S. Verner. 1985. The U.S. Environmental Protection Agency's Quality Assurance Program. In: Quality Assurance for Environmental Measurements, ASTM STP 867, J. K. Taylor and T. W. Stanley (eds.). American Society for Testing and Materials, Philadelphia, Pennsylvania. pp. 12-19.

Stapanian, M. A., A. K. Pollack, and B. C. Hess. 1987. Container Holding Time Study. Appendix B in C. M. Knapp, C. L. Mayer, D. V. Peck, J. R. Baker, and G. L. Filbin. National Surface Water Survey, National Stream Survey: Pilot Survey, Field Operations Report. EPA/600/8-87/019. U.S. Environmental Protection Agency, Las Vegas, Nevada.

Taylor, J. K. 1984. Guidelines for Evaluating the Blank Correction. Journal of Testing and Evaluation. 12:54-5.

Taylor, J. K. 1985. Principles of Quality Assurance of Chemical Measurements. NSBIR 85-3105. U.S. Department of Commerce, National Bureau of Standards, Gaithersburg, Maryland. 81 pp.

Taylor, J. K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Inc., Chelsea, Michigan. 328 pp.

Tyree, S. Y., Jr. 1981. Rainwater Acidity Measurement Problems. Atmospheric Environment, 5:57-60.

U.S. EPA (Environmental Protection Agency). 1979. Handbook for Analytical Quality Control in Water and Wastewater Laboratories. EPA/600/4-79/019. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio.

U.S. EPA (Environmental Protection Agency). 1983. Methods for Chemical Analysis of Water and Wastes. EPA/600/4-79/020. U.S. Environmental Protection Agency, Cincinnati, Ohio.

U.S. EPA (Environmental Protection Agency). 1987. Handbook of Methods for Acid Deposition Studies: Laboratory Analyses for Surface Water Chemistry. EPA/600/4-87/026. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. 342 pp.

Weast, R. C. (ed.). 1972. CRC Handbook of Chemistry and Physics, 53rd Ed., CRC Press, Cleveland, Ohio.

Young, T. C., J. V. DePinto, S. C. Martin, J. S. Bonner. 1985. Algal Available Particulate Phosphorous in the Great Lakes Basin. Journal of Great Lakes Research. 11(4):434-436.

Appendix A

Preparation of Audit Samples

Preparation of Natural Audit Samples

To ensure that all field natural audit samples of a particular lot were uniform, EMSL-LV instructed the support laboratory to follow the protocol specified below:

1. Clearly label the field and laboratory natural stock barrels with the lot number.

2. Label the 2-L bottles to be filled.

3. Operating in a clean environment, flush the Tygon tubing lines with lake water. Discard the water.

4. Pump 20 to 25 mL lake water into the audit bottle, cap the bottle, rinse the bottle to get complete coverage, and discard the rinse.

   NOTE: The Tygon tubing must not touch the sidewalls of the bottle.

5. Perform step 4 two more times. Discard the rinse water each time.

6. Fill the bottle to the top (no head space) with lake water that is filtered through a 0.40-micron filter.

   NOTE: The bottle must be capped immediately after it is filled to minimize the possibility of contamination.

7. Secure the cap to the bottle with tape.

8. Log in the total number of samples prepared, the date prepared, and the name of the analyst or technician.

9. Place samples in storage at 4 °C by lot and ID number to await shipment.

10. Discard any water remaining in Tygon tubing. Do not drain residual lake water into the stock barrel.

For laboratory natural audit samples, the contents of the 2-L bottle were divided into analytical aliquots and preserved (Figure 9) by the support laboratory before shipment to the processing laboratory.

Preparation of Synthetic Audit Samples

To prepare the field synthetic audit samples of the desired concentrations, support laboratory technicians diluted the lot stock concentrates with ASTM Type I reagent-grade water.

Each diluted 2-L synthetic audit sample was prepared for shipment on the same day to the processing laboratory as follows:

1. Fill a 2-L volumetric flask with 1.5-L deionized water.

2. Add a predetermined volume of each of the four stock concentrates (see Table A-1) to the flask.

3. Fill the flask to volume and mix the solution thoroughly.

4. When the dilution is complete, transfer the 2-L sample to a carboy. (If 10 sample were prepared in one day, the carboy would eventually contain 20-L of diluted stock, prepared 2-L at a time.)

5. When these dilution and transfer steps are completed, sparge the audit sample solution in the carboy with 300 ppm CO_2 and equilibrate. (The equilibration raises the acidity of the sample, thereby counteracting the effect of adding the strong base Na_2SiO_3. It also restores any DIC lost during sample preparation steps and, by stabilizing the sample, it minimizes day-to-day sample variation caused by shipping and handling).

Table A-1. Composition of the Field and Laboratory Synthetic Audit Sample Concentrates, National Stream Survey - Phase I

Stock concentrate	Chemical formula	Analytes to be measured
1	$Al_2(SO_4)_3 \cdot (NH_4)_2SO_4 \cdot 24H_2O$	Al-ext, Al-total NH_4^+, SO_4^{2-}
2	$FeNH_4(SO_4)_2 \cdot 12H_2O$	Fe, NH_4^+, SO_4^{2-}
3	Na_2SiO_3	Na, SiO_2
4	$CaCl_2$	Ca, Cl^-
	$NaHCO_3$	DIC, Na
	$C_6H_4(COOH)_2$	DOC
	$MgSO_4$	Mg, SO_4^{2-}
	NaF	F^-, Na
	$MnSO_4 \cdot H_2O$	Mn, SO_4^{2-}
	NH_4NO_3	NH_4^+, NO_3^-
	Na_2HPO_4	Na, P
	$KHC_8H_4O_4$	DOC, K

6. After the sample is sparged, transfer the sample to 2-L bottles and ship the same day to the processing laboratory.

For laboratory synthetic samples, aliquots were prepared from a 2-L bottle, preserved, and shipped on the same day to the processing laboratory.

Summary of Audit Sample Measurements

The data tables presented in this section provide information concerning the verification of the composition of the synthetic audit samples used for the NSS-I and also measured values for chemical variables in both synthetic and natural audit samples reported by other laboratories during other NSWS programs besides the NSS-I. Table A-2 summarizes the verification analyses that were conducted at the support laboratory for the synthetic audit samples. Table A-3 presents results of analyses of EPA reference samples that were analyzed at the support laboratory during the verification analyses of synthetic and natural audit samples (including other lots of synthetic audit samples used in other NSWS programs). These data provide an indication of the reliability of the verification analyses, as well as estimates of among-batch standard deviations. Tables A-4 through A-7 provide summary statistics from all laboratories (including the NSS-I participants) that analyzed the particular audit samples used during the NSS-I. These data provide an indication of the compatibility of the NSS-I data base with the data bases of other NSWS programs. Tables A-8 and A-9 present summary statistics associated with the calculation of index values for each audit sample type. These index values were developed based on the data reported from the various laboratories that analyzed each type of audit sample. These index values can be used to evaluate the performance of a particular analytical laboratory with respect to other laboratories that analyzed the same sample. Index values were calculated using weighted mean values from each laboratory because of the differences in sample size and precision.

Table A-2. Levels of Analytes in Synthetic Audit Samples Measured at the Support Laboratory, National Stream Survey - Phase I

Variable	Units	Theoretical value[b]	Lot 14 (n = 6)[a] Mean	Lot 14 (n = 6)[a] s	Lot 15 (n = 6)[a] Mean	Lot 15 (n = 6)[a] s
Al-ext	mg/L	0.020	0.024*	0.0038	0.009*	0.0005
Al-total	mg/L	0.0199	0.026*	0.0006	0.016*	0.0032
ANC	μeq/L	NC	102	2.90	109	2.93
BNC	μeq/L	NC	32	4.37	22.8	3.85
Ca	mg/L	0.194	0.179	0.0179	0.208*	0.0098
Cl^-	mg/L	0.343	0.315*	0.003	0.300*	0.0068
Cond-lab	μS/cm	17.5	15.2	0.03	15.23	0.034
DIC	mg/L	0.959[c]	1.141*	0.0720	1.263*	0.0197
DOC	mg/L	1.00	1.10*	0.064	0.98	0.062
F^-	mg/L	0.042	0.045*	0.0012	0.044*	0.0008
Fe	mg/L	0.059	0.039*	0.0018	0.041*	0.0024
K	mg/L	0.203	0.199	0.0077	0.217*	0.012
Mg	mg/L	0.447	0.41*	0.024	0.43*	0.0103
Mn	mg/L	0.098	0.097	0.0068	0.093*	0.0046
Na	mg/L	2.75	3.1*	0.098	2.9	0.20
NH_4^+	mg/L	0.168	0.14*	0.013	0.14*	0.009
NO_3^-	mg/L	0.467	0.460	0.0074	0.456	0.0154
P	mg/L	0.0273	0.025*	0.0015	0.023*	0.0016
pH	pH units	NC	7.22*	0.035	7.29*	0.024
SiO_2	mg/L	1.070	1.065	0.0139	1.057*	0.0082
SO_4^{2-}	mg/L	2.280	2.207*	0.0534	2.172*	0.0287

[a] n = number of measurements.
* = mean value significantly different from theoretical at $P \leq 0.05$.
s = within-batch standard deviation.

[b] The theoretical value is the expected value of the synthetic audit sample assuming no preparation error or external effect. NC = theoretical value was not calculated.

[c] Theoretical value does not include carbon dioxide added to the solution during the air-equilibration process.

Table A-3. Summary Statistics for EPA Reference Standards Measured With Audit Samples at the Support Laboratory, National Stream Survey - Phase I[a]

Variable (units)	Low concentration					High concentration				
	True value	Mean	s[b]	n[c]	Bias[d]	True value	Mean	s	n	Bias[d]
Al-ext (mg/L)	0.021	0.023	0.0052	5	+0.002	0.146	0.151	0.0235	5	+0.005
Al-total (mg/L)	0.021	0.024	0.0024	4	+0.003	0.146	0.143	0.0170	5	-0.003
ANC (μeq/L)	35.0	34.2	4.18	6	-0.8	68.8	67.1	0.77	5	-1.7
BNC (μeq/L)	--	--	--	0	--	--	--	--	0	--
Ca (mg/L)	0.53	0.53	0.026	6	0.00	4.06	3.85	0.179	4	-0.21
Cl^- (mg/L)	1.15	1.18	0.169	5	+0.03	8.08	8.22	0.186	3	+0.14
Cond-lab (μS/cm)	9.27	8.7	0.49	6	-0.57	55.2	45.9	1.69	5	-9.3
DIC (mg/L)	1.13	1.20	0.048	6	+0.07	5.68	5.19	0.426	3	-0.49
DOC (mg/L)	1.02	1.02	0.024	5	0.00	4.10	4.08	0.050	3	-0.02
Fe (mg/L)	0.04	0.04	0.003	4	<+0.01	1.56	1.39	0.244	3	-0.17
F^- (mg/L)	0.043	0.047	0.058	6	+0.004	0.130	0.152	0.0285	4	+0.022
K (mg/L)	0.21	0.21	0.013	5	0.00	0.98	1.04	0.092	4	+0.06
Mg (mg/L)	0.18	0.18	0.009	6	+0.00	0.84	0.80	0.025	4	-0.04
Mn (mg/L)	0.026	0.025	0.004	4	-0.001	0.696	0.681	0.051	3	-0.015
Na (mg/L)	0.82	1.05	0.122	6	+0.23	4.65	5.72	0.608	4	+1.07
NH_4^+ (mg/L)	0.19	0.18	0.013	5	-0.01	1.94	1.77	0.136	4	-0.17
NO_3^- (mg/L)	0.354	0.353	0.0066	3	-0.001	--	--	--	0	--
pH	5.70	5.62	0.034	6	-0.08	7.80	7.76	0.042	6	-0.04
P (mg/L)	0.130	0.120	0.0069	5	-0.01	1.030	0.897	0.093	4	+0.133
SiO_2 (mg/L)	4.28	4.45	0.348	6	+0.17	21.40	22.24	1.187	4	+0.84
SO_4^{2-} (mg/L)	0.72	0.73	0.021	6	+0.01	9.53	9.29	0.250	4	-0.24

[a] Source: Radian Corporation, 1987. Quality Assurance Audit Sample Support for the National Surface Water Survey Field Programs. Final Report submitted to U.S. Environmental Protection Agency, Las Vegas, Nevada. DCN 87-203-023-87-02. Contract 68-02-3994, Work Assignment Numbers 47 and 87.

[b] s = within-batch standard deviation.

[c] n = number of measurements.

[d] Bias = true value - mean value.

Table A-4. Summary Statistics for Measurements of Synthetic Audit Sample Lot 14, From Four Analytical Laboratories and Processing Laboratory (Field and Laboratory Audit Samples Combined)[a]

	NSS-I						ELS-II					
	LAB 1			LAB 2			LAB 1			LAB 2		
Variable (units)	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Al-ext (mg/L)	4	0.027	0.0160	5	0.021	0.0019	1	0.017	--	1	0.028	--
Al-total (mg/L)	6	0.026	0.0212	5	0.039	0.0102	1	0.022	--	1	0.023	--
ANC (μeq/L)	14	104.707	2.7008	9	114.378	7.4547	6	113.500	7.8656	2	113.200	2.8284
BNC (μeq/L)	14	20.850	5.6349	9	50.656	4.8706	6	36.817	12.1733	2	18.850	1.0607
Ca (mg/L)	14	0.218	0.0455	9	0.193	0.0216	6	0.189	0.0081	2	0.198	0.0156
Cl^- (mg/L)	14	0.335	0.0171	9	0.326	0.0349	6	0.248	0.0497	2	0.452	0.1570
Cond-lab (μS/cm)	14	16.329	0.3197	9	19.544	0.2698	6	18.950	0.6473	2	19.500	0.1414
DIC-eq (mg/L)	14	1.327	0.0846	9	1.374	0.2625	6	1.746	0.3101	2	1.237	0.0339
DIC-init (mg/L)	14	1.406	0.0524	9	1.624	0.0965	6	1.716	0.2083	2	1.345	0.0240
DOC (mg/L)	14	1.063	0.2768	9	1.281	0.1253	6	0.820	0.1942	2	1.335	0.1202
Fe (mg/L)	9	0.030	0.0138	5	0.040	0.0028	1	0.045	--	1	0.044	--
F^- (mg/L)	14	0.041	0.0033	9	0.043	0.0015	6	0.050	0.0220	2	0.043	0.0023
K (mg/L)	14	0.195	0.0048	9	0.203	0.0074	6	0.189	0.0081	2	0.246	0.0099
Mg (mg/L)	14	0.426	0.0036	9	0.436	0.0211	6	0.427	0.0078	2	0.445	0.0057
Mn (mg/L)	14	0.089	0.0068	9	0.109	0.0022	6	0.089	0.0012	2	0.095	0.0000
Na (mg/L)	14	2.819	0.0390	9	2.729	0.1162	6	2.717	0.1177	2	2.850	0.0396
NH_4^+ (mg/L)	14	0.167	0.0073	9	0.188	0.0130	6	0.137	0.0132	2	0.162	0.0099
NO_3^- (mg/L)	14	0.467	0.0544	9	0.454	0.0378	6	0.432	0.0807	2	0.493	0.0156
pH-ANC (pH units)	14	7.021	0.0804	9	6.734	0.0635	6	6.853	0.0647	2	7.100	0.0424
pH-BNC (pH units)	14	7.049	0.0866	9	6.740	0.1007	6	6.910	0.0603	2	7.145	0.0495
pH-eq (pH units)	14	7.171	0.2757	9	7.253	0.0892	6	7.087	0.0848	2	7.240	0.0141
P (mg/L)	14	0.023	0.0066	9	0.022	0.0034	6	0.020	0.0044	2	0.022	0.0004
SiO_2 (mg/L)	14	1.203	0.0267	9	0.957	0.0568	6	0.963	0.0327	2	1.089	0.0148
SO_4^{2-} (mg/L)	14	2.214	0.0495	9	2.151	0.2048	6	2.313	0.0495	2	1.722	0.8436

	Processing Laboratory					
	NSS-I			ELS-II		
	N	Mean	SD	N	Mean	SD
Al-mono (mg/L)	8	0.011	0.0089	5	0.009	0.0037
Al-nex (mg/L)	8	0.017	0.0091	5	0.016	0.0038
Cond-PL (μS/cm)	7	26.400	10.0411	0	--	--
DIC-closed (mg/L)	9	1.379	0.0394	6	1.370	0.1345
pH-closed (pH units)	9	6.958	0.0964	6	6.863	0.0787
True color (PCU)	8	5.000	2.6726	5	11.000	13.4164
Turbidity (NTU)	8	0.115	0.0288	6	0.147	0.0413

[a] For Al-ext, Al-total, and Fe, only laboratory audit samples are included. For processing laboratory measurements, only field audit samples are included.

Table A-5. Summary Statistics for Measurements of Synthetic Audit Sample Lot 15, From Three Analytical Laboratories and Processing Laboratory (Field and Laboratory Audit Samples Combined)[a]

Variable (units)	NSS-I LAB 1 N	NSS-I LAB 1 Mean	NSS-I LAB 1 SD	NSS-I LAB 2 N	NSS-I LAB 2 Mean	NSS-I LAB 2 SD	ELS-II LAB 1 N	ELS-II LAB 1 Mean	ELS-II LAB 1 SD
Al-ext (mg/L)	3	0.015	0.0035	23	0.015	0.0017	2	0.008	0.0045
Al-total (mg/L)	3	0.029	0.0092	24	0.036	0.0263	2	0.018	0.0081
ANC (μeq/L)	4	105.650	3.5977	29	118.993	7.1318	6	135.333	65.3334
BNC (μeq/L)	4	18.800	1.5188	29	58.890	12.4716	6	40.100	11.3434
Ca (mg/L)	4	0.199	0.0293	29	0.177	0.0105	6	0.179	0.0097
Cl^- (mg/L)	4	0.329	0.0075	29	0.336	0.0334	6	0.381	0.1519
Cond-lab (μS/cm)	4	15.950	0.1291	29	19.562	0.2194	6	18.967	0.1633
DIC-eq (mg/L)	4	1.233	0.0507	29	1.321	0.2169	6	1.795	0.3757
DIC-init (mg/L)	4	1.386	0.0621	29	1.511	0.2674	6	1.906	0.3993
DOC (mg/L)	4	0.995	0.2651	29	1.092	0.2930	6	0.843	0.2496
Fe (mg/L)	4	0.031	0.0259	24	0.024	0.0195	2	0.026	0.0269
F^- (mg/L)	4	0.039	0.0027	29	0.043	0.0017	6	0.044	0.0024
K (mg/L)	4	0.194	0.0025	29	0.200	0.0098	6	0.183	0.0145
Mg (mg/L)	4	0.430	0.0049	29	0.437	0.0426	6	0.439	0.0072
Mn (mg/L)	4	0.090	0.0046	29	0.103	0.0032	6	0.103	0.0034
Na (mg/L)	4	2.778	0.0228	29	2.703	0.0989	6	2.670	0.0874
NH_4^+ (mg/L)	4	0.155	0.0067	29	0.174	0.0156	6	0.099	0.0344
NO_3^- (mg/L)	4	0.477	0.0197	28	0.473	0.0341	6	0.365	0.1387
pH-ANC (pH units)	4	7.055	0.0520	29	6.687	0.1250	6	6.902	0.2022
pH-BNC (pH units)	4	7.075	0.0493	29	6.689	0.1299	6	6.907	0.1878
pH-eq (pH units)	4	7.307	0.1040	29	7.234	0.1464	6	7.157	0.1033
P (mg/L)	4	0.011	0.0083	29	0.024	0.0064	6	0.020	0.0062
SiO_2 (mg/L)	4	1.189	0.0228	29	0.974	0.0955	6	1.000	0.0245
SO_4^{2-} (mg/L)	4	2.250	0.0173	29	2.297	0.1198	6	2.081	0.0724

	Processing Laboratory NSS-I N	NSS-I Mean	NSS-I SD	ELS-II N	ELS-II Mean	ELS-II SD
Al-mono (mg/L)	5	0.009	0.0052	4	0.013	0.0035
Al-nex (mg/L)	5	0.013	0.0058	4	0.013	0.0021
Cond-PL (μS/cm)	5	19.060	1.1238	0	--	--
DIC-closed (mg/L)	5	1.439	0.0326	4	1.431	0.0802
pH-closed (pH units)	5	6.920	0.0628	4	6.927	0.0556
True color (PCU)	4	3.750	4.7871	4	3.750	2.5000
Turbidity (NTU)	4	0.102	0.0457	4	0.092	0.0287

[a] For Al-ext, Al-total, and Fe, only laboratory audit samples are included. For processing laboratory measurements, only field audit samples are included.

Table A-6. Summary Statistics for Measurements of Bagley Lake Natural Audit Sample, From Four Analytical Laboratories and Processing Laboratory (Field and Laboratory Audit Samples Combined)[a]

	NSS-I						ELS-II					
	LAB 1			LAB 2			LAB 1			LAB 2		
Variable (units)	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Al-ext (mg/L)	29	0.006	0.0024	8	0.008	0.0021	14	0.010	0.0038	24	0.006	0.0022
Al-total (mg/L)	29	0.016	0.0023	8	0.012	0.0026	14	0.019	0.0266	24	0.031	0.0452
ANC (μeq/L)	29	120.934	0.5440	8	122.450	4.0733	14	120.664	3.1750	24	131.958	12.9136
BNC (μeq/L)	29	31.303	1.9126	8	18.537	3.3517	14	23.000	4.1870	24	43.146	26.9478
Ca (mg/L)	29	1.574	0.0317	8	1.643	0.0202	15	1.500	0.0354	24	1.610	0.0374
Cl^- (mg/L)	29	0.162	0.0066	8	0.169	0.0147	15	0.175	0.0349	24	0.185	0.0507
Cond-lab (μS/cm)	29	14.290	0.4798	8	13.975	0.4950	14	10.743	0.5639	24	14.517	0.4603
DIC-eq (mg/L)	29	1.472	0.0923	8	1.524	0.0488	14	1.557	0.0549	24	1.470	0.2480
DIC-init (mg/L)	29	1.424	0.0508	8	1.626	0.0836	14	1.680	0.0500	24	1.464	0.2332
DOC (mg/L)	29	0.196	0.0864	8	0.407	0.0785	15	0.274	0.1418	24	0.540	0.5043
Fe (mg/L)	29	0.014	0.0080	8	0.002	0.0019	15	0.014	0.0252	24	0.004	0.0019
F^- (mg/L)	29	0.021	0.0011	8	0.022	0.0041	15	0.021	0.0015	24	0.027	0.0016
K (mg/L)	29	0.291	0.0076	8	0.294	0.0109	15	0.298	0.0077	24	0.310	0.0178
Mg (mg/L)	29	0.173	0.0028	8	0.176	0.0048	15	0.163	0.0231	24	0.178	0.0033
Mn (mg/L)	29	0.004	0.0097	8	0.000	0.0007	15	0.002	0.0063	24	0.001	0.0009
Na (mg/L)	29	0.815	0.0175	8	0.812	0.0179	15	0.835	0.0249	24	0.853	0.0487
NH_4^+ (mg/L)	29	0.000	0.0069	8	-0.007	0.0146	15	0.007	0.0111	24	0.018	0.0061
NO_3^- (mg/L)	28	0.007	0.0054	8	0.050	0.0747	14	0.007	0.0180	24	0.015	0.0186
pH-ANC (pH units)	29	7.095	0.0693	8	7.094	0.0746	14	7.077	0.0882	24	6.969	0.0847
pH-BNC (pH units)	29	7.046	0.0693	8	7.141	0.0522	14	7.134	0.1173	24	6.981	0.0851
pH-eq (pH units)	29	7.263	0.1569	8	7.216	0.1046	14	7.295	0.1782	24	7.307	0.0634
P (mg/L)	29	0.002	0.0033	8	0.001	0.0013	15	0.000	0.0024	24	0.002	0.0016
SiO_2 (mg/L)	29	9.650	0.3869	8	8.329	0.4407	15	9.906	0.2019	24	8.953	0.2939
SO_4^{2-} (mg/L)	29	0.639	0.0248	8	0.615	0.0339	15	0.631	0.0127	24	0.634	0.0588

	Processing Laboratory					
	NSS-I			ELS-II		
	N	Mean	SD	N	Mean	SD
Al-mono (mg/L)	25	0.013	0.0046	0	--	--
Al-nex (mg/L)	25	0.014	0.0046	0	--	--
Cond-PL (μS/cm)	26	15.505	2.4486	0	--	--
DIC-closed (mg/L)	27	1.703	0.0612	37	1.647	0.0593
pH-closed (pH units)	27	6.987	0.0493	36	7.126	0.0688
True color (PCU)	26	5.615	3.2010	32	1.250	2.1997
Turbidity (NTU)	27	0.175	0.3332	35	0.066	0.0591

[a] For processing laboratory measurements, only field audit samples were included.

Table A-7. Summary Statistics for Measurements of Big Moose Lake Natural Audit Sample, From Four Analytical Laboratories and Processing Laboratory (Field and Laboratory Audit Samples Combined)[a]

Variable (units)	NSS-I						ELS-II					
	LAB 1			LAB 2			LAB 1			LAB 2		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Al-ext (mg/L)	13	0.223	0.0490	24	0.210	0.0194	29	0.141	0.0562	24	0.162	0.0420
Al-total (mg/L)	13	0.259	0.0325	24	0.281	0.0168	35	0.259	0.0633	27	0.246	0.0388
ANC (μeq/l)	15	-2.907	2.2381	24	-1.625	5.9768	38	-1.987	4.2976	31	-3.677	2.1180
BNC (μeq/l)	15	64.413	6.6485	24	79.846	9.4930	38	72.768	14.5541	31	74.677	7.2473
Ca (mg/L)	15	1.829	0.0770	24	1.937	0.0382	38	1.983	0.6425	31	1.925	0.0551
Cl^- (mg/L)	15	0.430	0.0492	24	0.434	0.0536	38	0.406	0.1930	31	0.417	0.0233
Cond-lab (μS/cm)	15	23.753	0.3399	24	25.479	0.2187	38	25.380	1.0147	31	25.023	0.4938
DIC-eq (mg/L)	15	0.069	0.0469	24	0.103	0.0906	38	0.267	0.1231	31	0.096	0.0439
DIC-init (mg/L)	15	0.252	0.0392	24	0.230	0.0829	38	0.428	0.1253	31	0.454	0.0472
DOC (mg/L)	15	3.696	0.1552	24	4.065	0.2366	38	3.408	0.2587	31	3.419	0.1512
Fe (mg/L)	15	0.037	0.0188	24	0.057	0.0108	38	0.078	0.1159	31	0.044	0.0067
F^- (mg/L)	15	0.074	0.0036	24	0.074	0.0015	38	0.086	0.0754	31	0.073	0.0194
K (mg/L)	15	0.425	0.0106	24	0.427	0.0264	38	0.404	0.0177	31	0.393	0.0345
Mg (mg/L)	15	0.321	0.0049	24	0.326	0.0266	38	0.346	0.1334	31	0.322	0.0121
Mn (mg/L)	15	0.079	0.0263	24	0.086	0.0154	38	0.072	0.0041	31	0.071	0.0040
Na (mg/L)	15	0.632	0.0444	24	0.613	0.0316	38	0.609	0.0270	31	0.566	0.1660
NH_4^+ (mg/L)	15	0.062	0.0084	24	0.067	0.0086	38	0.056	0.0717	31	0.059	0.0322
NO_3^- (mg/L)	15	1.222	0.0341	24	1.232	0.0656	38	1.205	0.3054	31	1.199	0.0320
pH-BNC (pH units)	15	5.160	0.0939	24	5.235	0.1879	38	5.143	0.0773	31	5.137	0.0727
pH-ANC (pH units)	15	5.139	0.0748	24	5.224	0.1879	38	5.120	0.0892	31	5.081	0.0457
pH-Eq (pH units)	15	5.169	0.0144	24	5.245	0.0767	38	5.180	0.3357	31	5.135	0.3797
P (mg/L)	15	0.002	0.0028	24	0.002	0.0013	38	0.002	0.0018	31	0.002	0.0031
SiO_2 (mg/L)	15	4.668	0.1127	24	3.987	0.3106	38	4.418	0.2728	31	4.364	0.2541
SO_4^{2-} (mg/L)	15	6.391	0.0478	24	6.290	0.2018	38	6.567	1.3722	31	6.364	0.3626

	Processing Laboratory					
	NSS-I			ELS-II		
	N	Mean	SD	N	Mean	SD
Al-mono (mg/L)	24	0.195	0.0117	39	0.189	0.0205
Al-nex (mg/L)	24	0.058	0.0127	39	0.046	0.0190
Cond-PL (μS/cm)	26	25.473	1.7644	0	--	--
DIC-closed (mg/L)	27	0.526	0.0620	39	0.564	0.0513
pH-closed (pH units)	27	5.148	0.0594	39	5.129	0.0508
True color (PCU)	25	18.000	4.5644	38	15.789	4.8666
Turbidity (NTU)	25	0.607	2.1666	38	0.143	0.0934

[a] For processing laboratory measurements, only field audit samples were included.

Table A-8. Calculated Index Values for Measurements of Synthetic Audit Samples (Lots 14 and 15) (Field and Laboratory Audit Samples Combined)[a]

Variable (units)	Lot 14 Index Value N	Grand mean[b]	SD[c]	CI[d]	Lot 15 Index Value[a] N	Grand mean[b]	SD[c]	CI[d]
Al-ext (mg/L)	4	0.021	0.00086	0.00136	3	0.015	0.00035	0.00088
Al-total (mg/L)	4	0.036	0.00403	0.00642	3	0.028	0.00315	0.00781
ANC (μeq/L)	4	106.578	0.64174	1.02115	3	114.337	1.06563	2.64718
BNC (μeq/L)	4	24.033	0.61562	0.97960	3	23.105	0.71298	1.77114
Ca (mg/L)	4	0.192	0.00283	0.00451	3	0.178	0.00174	0.00431
Cl^- (mg/L)	4	0.330	0.00416	0.00662	3	0.331	0.00322	0.00800
Cond-lab (μS/cm)	4	18.335	0.05164	0.08217	3	18.624	0.03061	0.07604
DIC-eq (mg/L)	4	1.295	0.01605	0.02554	3	1.268	0.02124	0.05275
DIC-init (mg/L)	4	1.410	0.01018	0.01619	3	1.433	0.02599	0.06457
DOC (mg/L)	4	1.181	0.03082	0.04903	3	1.032	0.04512	0.11210
Fe (mg/L)	4	0.039	0.00120	0.00191	3	0.024	0.00373	0.00926
F^- (mg/L)	4	0.043	0.00042	0.00067	3	0.043	0.00030	0.00074
K (mg/L)	4	0.197	0.00106	0.00168	3	0.196	0.00102	0.00253
Mg (mg/L)	4	0.427	0.00088	0.00140	3	0.434	0.00182	0.00453
Mn (mg/L)	4	0.095	0.00040	0.00064	3	0.103	0.00053	0.00131
Na (mg/L)	4	2.813	0.00929	0.01479	3	2.751	0.00935	0.02324
NH_4^+ (mg/L)	4	0.167	0.00165	0.00262	3	0.164	0.00217	0.00540
NO_3^- (mg/L)	4	0.472	0.00705	0.01122	3	0.473	0.00537	0.01334
pH-BNC (pH units)	4	6.968	0.01384	0.02203	3	6.879	0.01682	0.04179
pH-ANC (pH units)	4	6.907	0.01200	0.01909	3	6.853	0.01694	0.04209
pH-eq (pH units)	4	7.230	0.00907	0.01444	3	7.227	0.02092	0.05198
P (mg/L)	4	0.022	0.00024	0.00038	3	0.023	0.00105	0.00260
SiO_2 (mg/L)	4	1.120	0.00519	0.00826	3	1.066	0.00692	0.01719
SO_4^{2-}(mg/L)	4	2.241	0.01092	0.01737	3	2.244	0.00779	0.01936

Variable (units)	Processing Laboratory Index Value N	Grand mean[b]	SD[c]	CI[d]	Index Value[a] N	Grand mean[b]	SD[c]	CI[d]
Al-mono (mg/L)	2	0.010	0.00145	0.01300	2	0.012	0.00139	0.01245
Al-nex (mg/L)	2	0.016	0.00152	0.01363	2	0.013	0.00098	0.00880
Cond-PL (μS/cm)	1	26.400	3.79517	--	1	19.060	0.50259	--
DIC-closed (mg/L)	2	1.379	0.01277	0.11473	2	1.438	0.01372	0.12326
pH-closed (pH units)	2	6.911	0.02271	0.20408	2	6.924	0.01977	0.17759
True color (PCU)	2	5.145	0.93341	8.38633	2	3.750	1.10801	9.95504
Turbidity (NTU)	2	0.123	0.00871	0.07829	2	0.095	0.01216	0.10927

[a] For Al-ext, Al-total, and Fe, only laboratory audit samples were used. For processing laboratory measurements, only field audit samples were used.

[b] Grand mean calculated from weighted mean from four analytical laboratories.

[c] SD = standard deviation, estimated as the square root of the reciprocal of the sum of individual laboratory weighting factors.

[d] CI = One-sided 95% confidence interval.

Table A-9. Index Values for Measurements of Bagley Lake and Big Moose Lake Natural Audit Samples (Field and Laboratory Audit Sample Combined)[a]

Variable (units)	Bagley Lake Index Value N	Grand mean[b]	SD[c]	CI[d]	Big Moose Lake Index Value[a] N	Grand mean[b]	SD[c]	CI[d]
Al-ext (mg/L)	4	0.006	0.00028	0.00044	4	0.197	0.00329	0.00524
Al-total (mg/L)	4	0.016	0.00039	0.00062	4	0.272	0.00284	0.00451
ANC (μeq/L)	4	120.954	0.09999	0.15911	4	-3.110	0.28133	0.44766
BNC (μeq/L)	4	29.685	0.32494	0.51705	4	72.897	0.85272	1.35687
Ca (mg/L)	4	1.588	0.00359	0.00571	4	1.924	0.00584	0.00929
Cl^- (mg/L)	4	0.163	0.00118	0.00188	4	0.420	0.00371	0.00590
Cond-lab (μS/cm)	4	13.844	0.05626	0.08953	4	25.120	0.03545	0.05640
DIC-eq (mg/L)	4	1.520	0.00920	0.01465	4	0.106	0.00594	0.00945
DIC-init (mg/L)	4	1.515	0.00736	0.01172	4	0.359	0.00581	0.00925
DOC (mg/L)	4	0.257	0.01288	0.02050	4	3.568	0.01833	0.02917
Fe (mg/L)	4	0.004	0.00032	0.00051	4	0.046	0.00103	0.00164
F^- (mg/L)	4	0.023	0.00016	0.00025	4	0.074	0.00029	0.00046
K (mg/L)	4	0.295	0.00106	0.00168	4	0.414	0.00178	0.00283
Mg (mg/L)	4	0.175	0.00040	0.00063	4	0.322	0.00107	0.00170
Mn (mg/L)	4	0.001	0.00014	0.00023	4	0.072	0.00048	0.00077
Na (mg/L)	4	0.820	0.00254	0.00405	4	0.612	0.00343	0.00546
NH_4^+ (mg/L)	4	0.008	0.00084	0.00134	4	0.065	0.00132	0.00210
NO_3^- (mg/L)	4	0.007	0.00096	0.00153	4	1.209	0.00451	0.00718
pH-ANC (pH units)	4	7.059	0.00890	0.01417	4	5.100	0.00659	0.01049
pH-BNC (pH units)	4	7.057	0.00867	0.01379	4	5.147	0.00827	0.01317
pH-eq (pH units)	4	7.293	0.01096	0.01744	4	5.173	0.00360	0.00573
P (mg/L)	4	0.002	0.00022	0.00036	4	0.002	0.00018	0.00029
SiO_2 (mg/L)	4	9.476	0.03370	0.05362	4	4.485	0.02032	0.03234
SO_4^{2-} (mg/L)	4	0.633	0.00255	0.00406	4	6.382	0.01162	0.01849

	Processing Laboratory							
	Bagley Lake Index Value				Big Moose Lake Index Value[a]			
Variable (units)	N	Grand mean[b]	SD[c]	CI[d]	N	Grand mean[b]	SD[c]	CI[d]
Al-mono (mg/L)	1	0.013	0.00093	--	2	0.193	0.00193	0.01735
Al-nex (mg/L)	1	0.014	0.00092	--	2	0.053	0.00198	0.01776
Cond-PL (μS/cm)	1	15.505	0.48021	--	1	25.473	0.34604	--
DIC-closed (mg/L)	2	1.670	0.00751	0.06746	2	0.552	0.00677	0.06080
pH-closed (pH units)	2	7.043	0.00731	0.06567	2	5.136	0.00663	0.05954
True color (PCU)	2	2.461	0.33057	2.97009	2	16.735	0.59714	5.36510
Turbidity (NTU)	2	0.068	0.00987	0.08872	2	0.144	0.01514	0.13607

[a] For processing laboratory measurements, only field audit samples were used.

[b] Grand mean calculated from weighted means from four analytical laboratories.

[c] SD = standard deviation, estimated as the square root of the reciprocal of the sum of individual laboratory weighting factors.

[d] CI = One-sided 95% confidence interval.

Appendix B

Data Qualifiers

Qualifiers were assigned to data points in the data sets to identify unusual conditions or results outside the expected criteria or limits. Tags assigned during collection and analyses activities are listed in Table B-1. Flags assigned as a result of verification activities are listed in Table B-2.

Table B-1. Field and Laboratory Data Qualifiers (Tags), National Stream Survey - Phase I

Qualifier	Indicates
A	Instrument unstable.
B	Redone, first reading not acceptable.
C	Instruments, sampling gear not vertical in water column.
D	Slow stabilization.
E	Result not available; sample destroyed during shipment.
F	Result outside QA criteria (with consent of QA manager).
G	Atypical result; already reanalyzed and confirmed by the laboratory manager.
H	Holding time exceeded criteria.
J	Result not available; insufficient sample volume shipped to analytical laboratory from the processing laboratory.
K	Result not available; entire aliquot not shipped.
L	Not analyzed because of interference.
M	Result not available; sample lost or destroyed by laboratory.
N	Not required.
P	Result outside QA criteria, but insufficient volume for reanalysis.
Q	Result outside QA criteria.
R	Result from reanalysis.
S	Contamination suspected.
T	Leaking container.
U	Result not required by procedure; unnecessary.
V	% ion balance difference value (Form 16) outside criteria because of high DOC.
W	% difference calculation for calculated ANC (Form 14) outside criteria because of high DOC.
X, Y, Z	Available for miscellaneous comments in the field and processing laboratory only.

Table B-2. Verification Data Qualifiers (Flags), National Stream Survey - Phase I

FLAGS USED WITH ANION/CATION BALANCE CHECK PROGRAM:

AO Anion/Cation % Ion Balance Difference is outside criteria due to an unknown cause.

A1 Anion/Cation % Ion Balance Difference is outside criteria due to unmeasured anions/cations (other anions/cations not considered in % ion balance difference calculation).

A2 Anion/Cation % Ion Balance Difference is outside criteria due to anion (flag suspect anion) contamination.

A3 Anion/Cation % Ion Balance Difference is outside criteria due to to cation contamination.

A4 Anion/Cation % Ion Balance Difference is outside criteria due to unmeasured organic protolytes (fits Oliver Model).

A5 Anion/Cation % Ion Balance Difference is outside criteria due to possible analytical error - anion concentration too high (flag suspect anion).

A6 Anion/Cation % Ion Balance Difference is outside criteria due to possible analytical error - cation concentration too low (flag suspect cation).

A7 Anion/Cation % Ion Balance Difference is outside criteria due to possible analytical error - anion concentration too low (flag suspect anion).

A8 Anion/Cation % Ion Balance Difference is outside criteria due to possible analytical error - cation concentration too high (flag suspect cation).

A9 Anion/Cation % Ion Balance Difference is outside criteria due to possible analytical error - alkalinity (ANC) measurement.

FLAGS GENERATED BY APPROPRIATE BLANK EXCEPTION PROGRAM:

BO External (field) blank is above expected criteria for pH, DIC, DOC, specific conductance, ANC, and BNC determinations.

B1 Internal (laboratory) blank is >2 x required detection limit for DIC, DOC, and specific conductance determinations.

B2 External (field) blank is above expected criteria and contributed >20% to sample concentrations. (This flag is not used for pH, DIC, DOC, specific conductance, ANC, and BNC determinations.)

B3 Internal (laboratory) blank is 2 x required detection limit and contributes >10% to the sample concentrations. (This flag is not used for DIC, DOC, and specific conductance determinations.)

(Continued)

Table B-2. (Continued.)

FLAGS GENERATED BY APPROPRIATE BLANK EXCEPTION PROGRAM (continued):

B4 Potential negative sample bias based on internal (laboratory) blank data.

B5 Potential negative sample bias based on external (field) blank data.

FLAGS USED WITH CONDUCTANCE BALANCE CHECK PROGRAM:

CO % Conductance Difference is outside criteria due to unknown cause.

C1 % Conductance Difference is outside criteria due to possible analytical error--anion concentration too high (flag suspect anion).

C2 % Conductance Difference is outside criteria due to anion contamination.

C3 % Conductance Difference is outside criteria due to cation contamination.

C4 % Conductance Difference is outside criteria due to unmeasured organic ions (fits Oliver Model).

C5 % Conductance Difference is outside criteria due to possible analytical error in specific conductance measurement.

C6 % Conductance Difference is outside criteria due to possible analytical error--anion concentration too low (flag suspect anion).

C7 % Conductance Difference is outside criteria due to unmeasured anions/cations (other anions/cations not measured in % conductance difference calculation).

C8 % Conductance Difference is outside criteria due to possible analytical error--cation concentration too low (flag suspect cation).

C9 % Conductance Difference is outside criteria due to possible analytical error--cation concentration too high (flag suspect cation).

FLAGS GENERATED BY DUPLICATE PRECISION EXCEPTION PROGRAM:

D2 External (field) duplicate precision exceeded the maximum expected % relative standard deviation, and both the routine and duplicate sample concentrations were $\geq$10 x required detection limit.

D3 Internal (laboratory) duplicate precision exceeded the maximum required % relative standard deviation, and both the routine and duplicate sample concentrations were $\geq$10 x required detection limit.

(Continued)

Table B-2. (Continued.)

FLAGS USED WHEN FIELD DATA ARE OUTSIDE CRITERIA:

FO % Conductance difference exceeded criteria when in situ field conductance value was substituted.

F1 Hillman/Kramer protolyte analysis program indicated field pH problem when stream site pH value was substituted.

F2 Hillman/Kramer protolyte analysis program indicated unexplained problem with stream site pH or processing laboratory DIC values when stream site pH value was substituted.

F3 Hillman/Kramer protolyte analysis program indicated field problem processing laboratory pH.

F4 Hillman/Kramer protolyte analysis program indicated field problem processing laboratory DIC.

F5 Hillman/Kramer protolyte analysis program indicated unexplained problem with processing laboratory pH or DIC values when processing laboratory pH value was substituted.

F6 % Conductance Difference exceeded criteria when processing laboratory specific conductance value was substituted.

FLAGS GENERATED BY HOLDING TIME EXCEPTION PROGRAM:

H0 The maximum holding time criteria were not met.

H1 No "Date Analyzed" data were submitted for reanalysis data.

FLAG GENERATED BY DETECTION LIMIT EXCEPTION PROGRAM:

L1 Instrumental Detection Limit exceeded required detection limit and sample concentration was <10 x instrumental detection limit.

FLAGS GENERATED BY AUDIT CHECK PROGRAM:

NO Audit sample value exceeded upper control limit.

N1 Audit sample value was below control limit.

FLAGS GENERATED BY HILLMAN/KRAMER PROTOLYTE ANALYSIS PROGRAM:

PO Laboratory problem--initial pH from alkalinity (ANC) titration.

P1 Laboratory problem--initial pH from acidity (BNC) titration.

P2 Laboratory problem--unexplained - initial pH from ANC or BNC titration.

(Continued)

Table B-2. (Continued.)

FLAGS GENERATED BY HILLMAN/KRAMER PROTOLYTE ANALYSIS PROGRAM (Continued):

P3 Laboratory problem--initial DIC determination.

P4 Laboratory problem--air-equilibrated pH or DIC determinations.

P5 Laboratory problem--unexplained - initial pH from ANC or BNC titrations or initial DIC determinations.

P6 Laboratory problem--alkalinity (ANC) determination.

P7 Laboratory problem--CO_2-acidity (BNC) determination.

FLAGS GENERATED BY QCCS EXCEPTION PROGRAM(S):

Q1 Quality Control Check Sample was above contractual criteria.

Q2 Quality Control Check Sample was below contractual criteria.

Q3 Insufficient number of Quality Control Check Samples were measured.

Q4 No Quality Control Check Sample was analyzed.

Q5 Detection Limit Quality Control Check Sample was not 2 to 3 x Required Detection Limit and measured value was not within 20% of the theoretical concentration.

MISCELLANEOUS FLAGS:

M0 Value obtained using a method which is unacceptable as specified in the Invitation for Bid contract.

M1 Value reported is questionable due to limitations of the laboratory methodology.

X0 Invalid but confirmed data based on QA review.

X1 Extractable aluminum concentration is greater than total aluminum concentration by 0.010 mg/L where extractable aluminum $\geq$ 0.015 mg/L.

X2 Invalid but confirmed data--potential aliquot switch.

X3 Invalid but confirmed data--potential gross contamination of aliquot or parameter.

X4 Invalid but confirmed data--potential sample (all aliquots) switch.

Values for flags X0 through X4 should not be included in any statistical analysis.

(Continued)

Table B-2. (Continued.)

MISCELLANEOUS FLAGS (continued):

X7 Site disturbance in watershed (e.g., strip mine).

MISSING CODE VALUE

"." Value never reported.

(Note: This code appears in numeric fields only.)

APPENDIX C

ACCEPTANCE CRITERIA

Appendix C consists of control limits for field blank, performance audit, and field duplicate pair samples used in the exception-generating programs found in AQUARIUS II.

CALCULATION OF FIELD BLANK SAMPLE CONTROL LIMITS

Criteria for determining contamination were needed in order to check for systematic contamination problems during sample collection and analysis and before preparing a verified data set. Some control limits were established on the basis of specifications provided by the instrument manufacturer; others reflected DQOs (i.e., the level of detectability needed to meet the goals of the survey). Control limits for some analytes could be defined only in terms of analytical experience and intuitive assumptions based on that experience, because there were not any acceptable precedents.

Upper control limits for the NSS-I blank samples were determined statistically on the basis of NSS-I Pilot experience and the analytical results obtained for NSS-I Pilot field blanks. The same type of sampling apparatus was used to collect field blanks for both surveys. The 95th percentile (P_{95}) nonparametric test was used to calculate the upper limit at which blank values would be flagged. The value of the required detection limit was used when the P_{95} statistic was below the required detection limit. The lower limit was designated as the negative value of the required detection limit, except for pH measurements for which the lower limit was the 5th percentile of the NSS-I Pilot field blanks. Anything less than this negative value was unacceptable and was attributed to excessive instrumental drift or to inaccurate calibration of the instrument.

Table C-I presents the field blank control limits for the NSS-I. Field blank concentrations that were outside these limits were considered suspect and were flagged. Establishing these limits prior to a full-scale statistical analysis was essential in order to identify contamination trends as they occurred. The detailed statistical analysis of the NSS-I field blank values for estimates of detectability was performed after data verification was completed.

TABLE C-1. Field Blank Control Limits, National Stream Survey - Phase I

Variable[a]	Low limit[b]	High limit[c]
Al-ext	-0.0050	0.0100
Al-total	-0.0050	0.0619
ANC (μeq/L)	-5.0000	5.6160
BNC (μeq/L)	-5.0000	26.3000
Ca	-0.0100	0.0400
Cl^-	-0.0100	0.0632
Cond-lab (μS/cm)	-0.9000	1.0000
DIC-eq	-0.0500	0.3620
DIC-init	-0.0500	0.2040
DOC	-0.1000	0.5400
F^-	-0.0050	0.0050
Fe	-0.0100	0.0100[d]
K	-0.0100	0.0100[d]
Mg	-0.0100	0.0100[d]
Mn	-0.0100	0.0100[d]
Na	-0.0100	0.0114
NH_4+	-0.0100	0.0210
NO_3-	-0.0050	0.0354
P	-0.0020	0.0084
pH-ANC (pH units)	5.4880[e]	5.9000
pH-BNC (pH units)	5.5600[e]	5.9680
pH-eq (pH units)	-5.6420[e]	6.7120
SiO_2	-0.0500	0.0622
$SO_4 2-$	-0.0500	0.0500[d]

[a] Units are in mg/L unless otherwise noted.
[b] The low limit is the negative value of the required detection limit.
[c] The high limit is the 95th percentile of the NSS-I Pilot field blanks.
[d] The required detection limit is substituted for the 95th percentile of the NSS-I Pilot field blanks.
[e] The 5th percentile is substituted for the negative value of the required detection limit.

PERFORMANCE AUDIT SAMPLE CONTROL LIMITS

Final audit sample control limits were generated after all analytical laboratory data (68 batches) had been entered into the raw data set. The QA plan (Drousé et al., 1986a) provides information about how to calculate the control limits. Values that were outside the control limits were considered suspect and were the basis for requesting confirmation of the values reported by the analytical laboratories. Tables C-2 through C-5 give the control limits for the two natural and two synthetic samples used during the NSS-I.

TABLE C-2 Control Limits for Performance Audit Samples From Bagley Lake, National Stream Survey - Phase I

	Control limits			
	Field audit		Laboratory audit	
Variable	Lower limit	Upper limit	Lower limit	Upper limit
Al-ext (mg/L)	0.0010	0.0105	0.0030	0.0136
Al-total (mg/L)	-0.0010	0.0330	0.0010	0.0290
ANC (μeq/L)	113.5960	133.6830	113.5220	139.2050
BNC (μeq/L)	13.5300	48.1790	10.4010	53.7490
Ca (mg/L)	1.4202	1.6990	1.4448	1.7258
Cl^- (mg/L)	0.0841	0.2976	0.1386	0.1790
Cond-lab (μS/cm)	9.3027	17.2895	8.2771	17.2394
DIC-eq (mg/L)	1.0590	1.8913	1.3650	1.7007
DIC-init (mg/L)	1.0370	2.0243	1.3480	1.8676
DOC (mg/L)	0.0590	0.6909	0.1640	0.3786
F^- (mg/L)	0.0186	0.0312	0.0155	0.0340
Fe (mg/L)	-0.0060	0.0115	-0.0080	0.0199
K (mg/L)	0.2846	0.3347	0.2752	0.3260
Mg (mg/L)	0.1666	0.1836	0.1658	0.1859
Mn (mg/L)	-0.0020	0.0036	-0.0020	0.0031
Na (mg/L)	0.7509	0.9428	0.7687	0.9173
NH_4+ (mg/L)	-0.0030	0.0340	-0.0150	0.0339
NO_3- (mg/L)	-0.0260	0.0518	-0.0070	0.0124
P (mg/L)	-0.0040	0.0066	0.0000	0.0037
pH-ANC (pH units)	6.7789	7.2233	6.9026	7.0813
pH-BNC (pH units)	6.7898	7.2565	6.8278	7.2387
pH-eq (pH units)	7.0722	7.5293	7.2024	7.3812
SiO_2 (mg/L)	8.0746	10.5190	8.3863	10.3577
$SO_4 2-$ (mg/L)	0.5660	0.6788	0.5637	0.6945

TABLE C-3 Control Limits for Performance Audit Samples From Big Moose Lake, National Stream Survey - Phase I

Variable	Control limits			
	Field audit		Laboratory audit	
	Lower limit	Upper limit	Lower limit	Upper limit
Al-ext (mg/L)	0.1400	0.2719	0.1840	0.3004
Al-total (mg/L)	0.2120	0.3241	0.2370	0.3260
ANC (μeq/L)	-6.7820	-0.8270	-9.3400	1.5040
BNC (μeq/L)	51.0900	96.2952	44.6450	104.1552
Ca (mg/L)	1.7282	2.0497	1.7949	2.0581
Cl^- (mg/L)	0.3243	0.5501	0.3677	0.4388
Cond-lab (μS/cm)	22.8934	26.7436	22.8268	26.7897
DIC-eq (mg/L)	-0.0540	0.2149	-0.1220	0.3457
DIC-init (mg/L)	0.0970	0.3571	0.1100	0.3760
DOC (mg/L)	3.3390	4.4967	3.2830	4.5851
F^- (mg/L)	0.0704	0.0774	0.0691	0.0793
Fe (mg/L)	0.0110	0.0889	0.0420	0.0620
K (mg/L)	0.3839	0.4737	0.3709	0.4680
Mg (mg/L)	0.3100	0.3441	0.3128	0.3448
Mn (mg/L)	0.0610	0.1023	0.0640	0.1054
Na (mg/L)	0.5482	0.6949	0.5807	0.6398
NH_4+ (mg/L)	0.0460	0.0840	0.0470	0.0841
NO_3- (mg/L)	1.1480	1.3045	1.0870	1.3434
P (mg/L)	-0.0010	0.0052	-0.0010	0.0033
pH-ANC (pH units)	4.9903	5.3424	4.9817	5.2787
pH-BNC (pH units)	5.0226	5.3069	4.8747	5.5468
pH-eq (pH units)	5.0818	5.3304	5.0834	5.3237
SiO_2 (mg/L)	3.3512	5.1708	3.3070	5.1361
$SO_4 2-$ (mg/L)	5.9858	6.6752	5.9118	6.7373

TABLE C-4 Control Limits for Performance Audit Samples From Synthetic Lot 14, National Stream Survey - Phase I

Variable	Control limits			
	Field audit		Laboratory audit	
	Lower limit	Upper limit	Lower limit	Upper limit
Al-ext (mg/L)	0.0000	0.0179	0.0020	0.0517
Al-total (mg/L)	-0.0340	0.1079	-0.0100	0.0664
ANC (μeq/L)	87.6412	130.6250	97.6354	116.3490
BNC (μeq/L)	-1.7160	71.6280	-6.0570	67.9430
Ca (mg/L)	0.1227	0.2715	0.1179	0.3127
Cl^- (mg/L)	0.2635	0.4017	0.3143	0.3386
Cond-lab (μS/cm)	13.5135	22.0642	13.8833	21.0309
DIC-eq (mg/L)	0.8790	1.7018	1.1280	1.5440
DIC-init (mg/L)	1.1890	1.8158	1.1820	1.7862
DOC (mg/L)	0.8550	1.4353	0.6240	1.5587
F^- (mg/L)	0.0341	0.0514	0.0355	0.0469
Fe (mg/L)	-0.0170	0.0317	0.0150	0.0561
K (mg/L)	0.1852	0.2172	0.1925	0.1996
Mg (mg/L)	0.4078	0.4426	0.4215	0.4332
Mn (mg/L)	0.0630	0.1284	0.0750	0.1207
Na (mg/L)	2.7904	2.8668	2.7060	2.8847
NH_4+ (mg/L)	0.1470	0.2036	0.1390	0.2106
NO_3- (mg/L)	0.3790	0.5638	0.3350	0.5757
P (mg/L)	0.0140	0.0234	0.0160	0.0310
pH-ANC (pH units)	6.4931	7.2957	6.5537	7.2834
pH-BNC (pH units)	6.4869	7.3419	6.5200	7.3542
pH-eq (pH units)	6.8688	7.4534	6.8165	7.7434
SiO_2 (mg/L)	0.7382	1.4360	0.8442	1.3932
$SO_4 2-$ (mg/L)	1.7440	2.6095	1.9712	2.4235

TABLE C-5 Control Limits for Performance Audit Samples From Synthetic Lot 15, National Stream Survey - Phase I

Variable	Control limits			
	Field audit		Laboratory audit	
	Lower limit	Upper limit	Lower limit	Upper limit
Al-ext (mg/L)	0.0000	0.0125	0.0110	0.0186
Al-total (mg/L)	0.0110	0.0421	0.0140	0.0465
ANC (μeq/L)	114.6760	125.6740	99.0350	134.1080
BNC (μeq/L)	29.4950	98.8250	14.3340	90.1090
Ca (mg/L)	0.1469	0.1914	0.1556	0.2039
Cl^- (mg/L)	0.2744	0.3803	0.2752	0.3910
Cond-lab (μS/cm)	19.3235	20.1164	16.3188	21.7168
DIC-eq (mg/L)	0.8080	1.8036	0.8630	1.7594
DIC-init (mg/L)	1.0060	2.3103	0.9380	1.9948
DOC (mg/L)	-0.8090	2.6054	0.7810	1.3563
F^- (mg/L)	0.0382	0.0472	0.0419	0.0452
Fe (mg/L)	-0.0100	0.0161	-0.0170	0.0667
K (mg/L)	0.1806	0.2233	0.1788	0.2197
Mg (mg/L)	0.4074	0.4801	0.4204	0.4654
Mn (mg/L)	0.0970	0.1097	0.0920	0.1126
Na (mg/L)	2.2153	3.1446	2.5396	2.8954
NH_4+ (mg/L)	0.1270	0.2120	0.1430	0.1966
NO_3- (mg/L)	0.4170	0.4886	0.4230	0.5173
P (mg/L)	0.0020	0.0454	0.0190	0.0273
pH-ANC (pH units)	6.1860	7.1059	6.3907	7.1035
pH-BNC (pH units)	6.1841	7.0878	6.3810	7.1254
pH-eq (pH units)	6.6952	7.8487	6.9524	7.5225
SiO_2 (mg/L)	0.6790	1.2451	0.7600	1.2530
$SO_4 2-$ (mg/L)	1.8906	2.5333	2.0754	2.5360

FIELD ROUTINE-DUPLICATE PAIR PRECISION LIMITS

Precision limits for field routine-duplicate pairs were designed using the DQO for the intralaboratory precision estimates as a reference. Some flexibility was allowed since these samples pass through the system from the field to the analytical laboratory, whereas laboratory duplicates do not.

A data qualifier flag (Appendix B) was given to all the samples in the batch when the precision estimates for the field routine-duplicate pair exceeded the allowable limit. The measurements of the field routine-duplicate pairs for this variable are considered suspect and the analytical laboratories are asked to confirm the value.

TABLE C-6 Precision Limits for Field Routine-Duplicate Pairs

Variable	Field Routine-Duplicate Pair Precision Limit Percent Relative Standard Deviation (%RSD)[a]
Al-ext	10 (if Al-ext concentration > 0.01 mg/L) 20 (if Al-ext concentration ≤ 0.01 mg/L)
Al-total	10 (if Al-total concentration > 0.01 mg/L) 20 (if Al-total concentration ≤ 0.01 mg/L)
ANC	10
BNC	10
Ca	5
Cl^-	5
Cond-lab	3
DIC-eq	10
DIC-init	10
DOC	10
F^-	10
Fe	10
K^+	10
Mg	5
Mn	10
Na	10
NH_4^+	10
NO_3^-	10
P	10 (if P concentration > 0.01 mg/L) 20 (if P concentration ≤ 0.01 mg/L)
pH-ANC	±0.1 (pH unit)
pH-BNC	±0.1 (pH unit)
pH-eq	±0.1 (pH unit)
SiO_2	5
SO_4^{2-}	5

[a] This limit was the %RSD at all concentrations, unless otherwise noted.

APPENDIX D

ESTIMATING RELATIVE INTERLABORATORY BIAS FOR THE NATIONAL STREAM SURVEY

The following report assesses relative interlaboratory bias during Phase I of the National Stream Survey. The report was submitted to Lockheed Engineering and Management Services Company by Systems Applications, Inc.

ESTIMATING RELATIVE INTERLABORATORY BIAS
FOR THE NATIONAL STREAMS SURVEY

SYSAPP-87/093

15 June 1987

Prepared for

Mohammed Miah

Lockheed EMSCO
1050 East Flamingo Road, Suite 209
Las Vegas, Nevada 89119-7432

(EPA Contract 68-03-3249, Task 2)

Prepared by

S. D. Edland
T. J. Permutt
T. S. Stocking

Systems Applications, Inc.
101 Lucas Valley Road
San Rafael, CA 94903

CONTENTS

Errata

Page 8, line 20: ℓ_3 should be ℓ_1

Page 10, line 2: ℓ_1 should be ℓ_3

Page 16, Table 4: Column head "vaiable" should be "variable"

App. B, page 3, line 7: μ should be μ_1

App. B, page 4, line 9: "m.l.e" should be m.l.e."

App. B, page 5, line 1: "γ_i = 1, 2, ...7" should read "γ_i, i = 1, 2, ...7,"

1 INTRODUCTION

During the course of the National Streams Survey over 1300 water samples were collected from 450 streams in the mid-Atlantic and Southeast regions. These samples were sent to one of two laboratories contracted for the survey, New York State Department of Health (NYSDOH) or Global, for analysis of 24 critical chemical constituents (Table 1). In addition, 134 "performance audit" water samples were sent to the laboratories to aid in assessing the quality of the data they produced. We analyzed this performance audit data set to assess and estimate the relative interlaboratory bias of the measurements.

Relative interlaboratory bias can be defined as the mean difference in measurement by two laboratories of identical water samples. Different laboratory facilities, personnel, and instrumentation provide many opportunities for inconsistent treatment of the samples, and hence for the introduction of some interlaboratory bias. This interlaboratory bias may then confound the analysis of the data. Differences between two streams ascribed to properties of the water may actually be the result of measurement bias. Apparent regional differences in stream water quality may actually be due to the fact that streams from different regions were analyzed by different laboratories. For these reasons it is important to assess and, if possible, correct for any interlaboratory bias within the survey.

TABLE 1. Variables measured in National Streams Survey.

Calcium	CA
Magnesium	MG
Potassium	K
Sodium	NA
Manganese	MN
Iron	FE
Aluminum (extractable)	ALEX
Chloride	CL
Sulfate	SO4
Nitrate	NO3
Silica	SIO2
Fluoride (total)	FTL
Dissolved Organic Carbon	DOC
Ammonium	NH4
pH (equilibrated)	PHEQ
pH (alkalinity)	PHAL
pH (acidity)	PHAC
Acidity	ACCO
Alkalinity (μeq/l)	ALKA
Conductivity (μS/cm)	COND
Dissolved Inorganic Carbon (equilibrated)	DICE
Dissolved Inorganic Carbon (initial)	DICI
Phosphorus (total)	PTL
Aluminum (total)	ALTL

2 STATISTICAL METHODS

Interlaboratory bias can be estimated if we have similar water samples measured by both of the labs. This is exactly the case for the performance audit data in the National Streams Survey (NSS). The performance audit data were grouped into eight types of samples; of these, seven were analyzed by both NYSDOH and Global.

Audit Group	Number of Samples	
	Global	NYSDOH
FL-14	5	4
LL-14	9	5
FL-15	0	5
LL-15	4	24
FN-6	10	17
LN-6	5	7
FN-8	11	16
LN-8	4	8
	48	86

These audit groups are distinguished by the source of the samples--L14 and L15 are low concentration synthetic audit samples (produced by a contract laboratory), and N6 and N8 are natural audit samples (Bagley Lake, Washington, and Big Moose Lake, New York, respectively). The groups are further distinguished by the sample processing protocol used--the F prefix indicates samples preprocessed at the NSS field lab, and the L prefix indicates samples preprocessed at the contract laboratory.

The samples within each of the natural audit groups were produced by subdividing a single large, homogeneous, and presumably chemically stable water sample into several two-liter aliquots. The synthetic audit samples

were produced from stock solutions using a recipe designed to give concentrations close to the limits of quantitation for most analytes. Measurements will be considered as repeated measures of the same solution.

Ideally, the mean measurements by Global and NYSDOH should approximately agree for each of the seven data groups. Consistent or large deviations between these pairs of means would be an indication of interlaboratory bias. These seven pairs of means, for each of the 24 parameters measured in the survey, are summarized in the scatterplots in Appendix A. Each of the scatterplots includes the line of identity (of slope 1 and intercept 0) about which we would expect the data points to be clustered if there were no interlaboratory bias. The scatterplots will be discussed in more detail in Section 5. Deviations from the line of identity may be the result of random error, or they may be the result of bias.

In order to quantify these deviations from the line of identity and thus be able to test hypotheses about the deviations, we need models that describe these deviations as a function of the concentrations of the analytes. Since there are only seven data points with which to estimate these functional relationships, we considered only the simplest linear functions. Fitting more complex functions would risk overfitting the data, and would be appropriate only if we had previous knowledge of the functional relationships to expect, as for example, if a more complex relationship had been suggested by data on the laboratory instrumentation in an earlier study. The linearity assumption is, however, a practical alternative. It is often appropriate in statistical applications, since functions are often approximately linear when considered over a limited range. Also, it may be that many of the factors that lead to measurement error, such as sample contamination or errors in instrument calibration, result in biases that are linear functions of concentration.

We considered four linear models: (1) no bias, (2) constant bias, (3) bias proportional to concentration, and (4) the linear model with a constant term and a proportionality term. These models are described in

detail in Appendix B. They translate into the following functional relationships between the expected values of measurements by Global and NYSDOH:

(1) NYSDOH measurements = Global measurements
(2) NYSDOH measurements = Global measurements + α
(3) NYSDOH measurements = $(1 + \beta)$ (Global measurements)
(4) NYSDOH measurements = $(1 + \beta)$ (Global measurements) + α

where α represents the constant term and β represents the proportionality term.

The positions of Global and NYSDOH in these relationships is completely arbitrary and is by no means intended to suggest that one lab has performed better than the other. The roles of NYSDOH and Global could be switched in the following analyses and none of the conclusions would be changed.

If the audit samples are representative of the water samples encountered in the survey, these relationships can be used to adjust the NSS data base and correct for interlaboratory bias. Our goal is then to estimate α and/or β (for model numbers 2, 3, and 4) and provide statistical procedures for comparing the four models.

Assuming that the audit samples are measured with independent normal errors, this can be treated as a maximum likelihood estimation problem.* The maximum likelihood estimation procedure identifies for each model the parameters that are most likely to have produced the observed audit

---

* Permutt, T., M. Moezzi, and S. C. Grosser. 1986. "Relative Interlaboratory Bias in the Western Lake Survey." Systems Applications, Inc., San Rafael, California (SYSAPP-86/173).

data. The maximum likelihood estimates (m.l.e.) can then be used to test various hypotheses about the four models. The derivation of the m.l.e. is described in Appendix B. The hypotheses tests are as follows.

First, we evaluate the probability density functions corresponding to each of the models at their maximum likelihood estimates. Call these values L_1, L_2, L_3, and L_4.

Second, we define three statistics ℓ_1, ℓ_2, and ℓ_3 as:

$$\ell_1 = -2 \ln \frac{L_4}{L_1}$$

$$\ell_2 = -2 \ln \frac{L_4}{L_2}$$

$$\ell_3 = -2 \ln \frac{L_4}{L_3}$$

Finally, we note that these statistics have approximately chi-square distributions (ℓ_2 and ℓ_3 are chi-square with one degree of freedom, ℓ_1 with two degrees of freedom) under the appropriate hypotheses, and so provide test statistics for these hypotheses.

For example, ℓ_3 can be used to compare model 4 ($\alpha \neq 0$ and $\beta \neq 0$) with model 3 ($\alpha = 0$). Under the assumption that α actually is equal to 0, ℓ_3 is distributed as chi-square. If in fact α is not equal to 0, then observed values of ℓ_3 would be tend to be larger than we would expect under the hypothesis. If ℓ_3 is large enough (e.g., greater than 3.85 at the 95 percent confidence level) we conclude that $\alpha \neq 0$. Otherwise we can assume that the simpler model with $\alpha = 0$ is adequate to explain the functional relationship of Global and NYSDOH measurements. In an analogous fashion, ℓ_1 and ℓ_2 can be used to test the models with $\alpha = \beta = 0$ and $\beta = 0$ against the general linear model.

We can test for our basic assumption of linearity in a similar fashion. If we define a general bias model with seven distinct bias terms to

describe the bias among our seven pairs of samples (see Appendix B for the derivation of m.l.e.), and we define L_5 as this bias model's density function evaluated at its m.l.e. values, then,

$$\ell_4 = -2 \log \frac{L_5}{L_4}$$

is distributed as a chi-square with five degrees of freedom under the hypothesis that bias is indeed linear. High values of ℓ_4 (greater than 10.07 at the 95 percent confidence level) cause rejection of the linear bias model.

The observed values of the statistics ℓ_1 through ℓ_4 and an interpretation of their meaning relative to our problem are given in the following section.

3 RESULTS

We tested the following four hypotheses about the functional relationship of interlaboratory bias to concentration. (Observed test statistics ℓ_1 through ℓ_4 corresponding to these hypotheses are given, for each of the 24 analytes in the study, in Table 2.)

To choose among the linear functions considered, we test:

(1) H_0: no bias (i.e., $\alpha = \beta = 0$) (versus H_A: $\alpha \neq 0$, $\beta \neq 0$); this hypothesis is accepted at the 95 percent confidence level if $\ell_1 < 5.99$.

(2) H_0: bias is constant (i.e., $\beta = 0$) (versus H_A: $\alpha \neq 0$, $\beta \neq 0$); this hypothesis is accepted at the 95 percent confidence level if $\ell_2 < 3.85$.

(3) H_0: bias is proportional to concentration (i.e., $\alpha = 0$) (versus H_A: $\alpha \neq 0$, $\beta \neq 0$); this hypothesis is accepted at the 95 percent confidence level if $\ell_3 < 3.85$.

(4) H_0: bias is a linear function of concentration (versus H_A: there are seven unique bias parameters); we reject H_0 (and accept H_A) at the 95 percent confidence level if $\ell_4 > 10.07$.

Using the chi square scores in Table 2 as a guide, we can choose from among the linear models considered.

Based on ℓ_3, the no bias model is acceptable for CL, SO4, NO3, DICE, DICI, and PTL. Of the remaining variables, the $\beta = 0$ model is acceptable for

TABLE 2. Chi square statistics.

Variable	ℓ_1	ℓ_2	ℓ_3	ℓ_4
CA	57.444	46.794	10.211	3.591
MG	18.781	0.415	6.188	3.503
K	18.223	0.001	1.441	3.853
NA	12.316	9.187	1.197	14.015
MN	46.279	60.129	0.629	7.040
FE	17.288	15.031	4.636	1.714
ALEX	25.273	7.094	11.161	6.605
CL	1.150	1.146	1.023	4.830
SO4	2.348	2.162	1.616	10.416
NO3	1.097	1.075	0.684	2.426
SIO4	193.560	67.310	45.345	11.227
FTL	76.548	21.332	48.723	5.496
DOC	40.050	5.901	7.315	7.719
NH4	8.265	3.401	0.029	6.368
PHEQ	17.904	0.285	1.586	7.629
PHAL	28.428	17.114	13.282	36.951
PHAC	36.253	21.804	17.953	31.901
ACCO	93.509	26.108	50.790	19.497
ALKA	40.402	24.106	1.555	1.461
COND	379.168	73.498	90.049	20.262
DICE	2.791	0.603	2.733	6.400
DICI	0.832	0.295	0.018	37.994
PTL	2.201	0.412	2.138	13.396
ALTL	12.778	0.490	5.481	5.271

MG, K, NH4, PHEQ and ALTL (see ℓ_2), while the $\alpha = 0$ model is acceptable for K, NA, MN, NH4, PHEQ, and ALKA (see ℓ_1). Both the $\beta = 0$ and the $\alpha = 0$ models are acceptable for K, NH4, and PHEQ. The final choice of models for these three variables can rest on considerations of which form of bias, constant or proportional to concentrations, is more likely to have been introduced during the respective measurement procedures. The maximum likelihood estimates of α and β for the chosen models are listed in Table 3.

We use the statistic ℓ_4 to test hypothesis (4). Rejection of hypothesis (4), indicated in Table 3 by an asterisk, means that our basic assumption of linearity is in doubt. In practical terms this means that we are less certain of the functional relationships suggested by hypotheses 1 through 3 and hence of the associated transformations in Table 3. The implications of this are discussed in detail in Section 5.

TABLE 3. Estimated transformation coefficients.

Variable	α	β	Nonlinear[b]
CA	-0.042	0.096	--
MG	0.0070	0.0	--
K[a]	0.0071	0.0	--
K[a]	0.0	0.030	--
NA	0.0	-0.020	*
MN	0.0	0.187	--
FE	-0.0205	1.053	--
ALEX	-0.0038	-0.122	--
CL	0.0	0.0	--
SO4	0.0	0.0	*
NO3	0.0	0.0	--
SIO2	-0.18	-0.083	*
FTL	0.0081	-0.112	--
DOC	0.17	0.064	--
NH4[a]	0.005	0.0	--
NH4[a]	0.0	0.064	--
PHEQ[a]	0.071	0.0	--
PHEQ[a]	0.0	0.012	--
PHAL	0.661	-0.116	*
PHAC	1.13	-0.208	*
ACCO	37.7	-0.394	*
ALKA	0.0	0.106	--
COND	5.91	-0.174	*
DICE	0.0	0.0	--
DICI	0.0	0.0	*
PTL	0.0	0.0	*
ALTL	0.014	0.0	--

[a] Two models are acceptable for these variables.

[b] An asterisk indicates rejection of the linearity hypothesis at the 95 percent confidence level.

4 APPLYING THE RESULTS

There are several ways in which the models in Table 3 can be applied. Where bias is identified, the measurements by the two labs need to be in effect calibrated so that they are expressed on the same scale. Hence if one has reason to believe that NYSDOH's measurements are more accurate than Global's, then the Global measurements would be transformed to the NYSDOH measurement scale:

$$\text{Global}_{NEW} = (1 + \beta)\ (\text{Global}_{OLD}) + \alpha \quad .$$

Conversely, if one believes that the Global measurements are more accurate, the NYSDOH measurements would be transformed to the Global measurement scale:

$$\text{NYSDOH}_{NEW} = (\text{NYSDOH}_{OLD} - \alpha)/(1 + \beta) \quad .$$

In the absence of knowledge of absolute bias, a reasonable alternative would be to split the difference between the measurement scales for the two labs and transform the measurements by each lab to this new scale:

$$\text{Global}_{NEW} = (1 + \tfrac{\beta}{2})\ (\text{Global}_{OLD}) + \tfrac{\alpha}{2} \quad ,$$

$$\text{NYSDOH}_{NEW} = (\text{NYSDOH}_{OLD} - \tfrac{\alpha}{2})/(1 + \tfrac{\beta}{2}) \quad .$$

5 CONCLUSIONS

DISCUSSION

The coefficients listed in Table 3 can be used to transform the measurement data from one laboratory to the measurement scale of the other when interlaboratory bias is identified. Even given non-zero coefficients we cannot immediately assume that the NSS data would be improved by applying the transformations. Instead, we must decide whether the expected improvements in relative accuracy justify the loss in precision that will result from applying these transformations. Loss of precision results from the fact that the values of α and β in Table 3 are estimates only; the uncertainty of the α and β estimates (i.e., the uncertainty in the estimated functional relationship) would produce uncertainty in the transformed NSS data.

The figures in Appendix A illustrate the potential gains and losses from calibration. The distance from the best fitting line to the line of identity at a given concentration is our best estimate of the bias at that concentration. Other things being equal, the larger this estimated bias, the more important the correction transformation (i.e., calibration) becomes. On the other hand, the curved band represents the uncertainty of this estimate. If, for example, this band straddles the line of identity at a relevant concentration, there is a significant doubt that the proposed correction is even in the right direction. This happens for potassium (K) for example. In less extreme cases, the direction of the bias is clear, e.g., extractable aluminum (ALEX), but there is still substantial uncertainty about its magnitude. In such cases calibration therefore introduces a substantial uncertainty in exchange for eliminating a likely bias.

The figures show in each case the line of best fit. In many cases, identified in Table 3, the fit is not significantly worse for a line constrained to have zero intercept or unit slope. In these cases, if calibration seems desirable, the simpler forms are probably to be preferred.

An additional factor to consider before applying the transformations in Table 2 is the representativeness of the range of the audit data as compared to the range of the NSS data. As an illustration of the importance of this consider the calcium audit data. The range from zero to the highest calcium audit group mean (1.95 mg/l) contains only 23.0 percent of the NSS calcium measurements. This is the range for which we can be most certain that the audit data has represented the NSS data, and hence most certain of the validity of the estimated functional relationship. Applying this transformation beyond the range of the audit data relies heavily on the assumption of linearity, especially when the NSS data extend well beyond the range of the audit data. For calcium, for example, the range of the audit data represents only 2.0 percent of the entire range of the NSS data. Even if we eliminate the top 5 percent, and thus any outliers, the audit data still only represents 4.6 percent of lowest 95 percent of the NSS data. This is dramatically illustrated in Figure 1, which shows the estimated transformation and 95 percent confidence bounds extrapolated over approximately 95 percent of the range of the NSS calcium data. (The audit data means are clustered in the bottom left corner of the graph.)

The measures of representativeness outlined above are provided, for each of the variables in the study, in Table 4.

Figure 1 emphasizes the importance of the linearity assumption when the audit data do not cover the range of the NSS data. Even given this assumption we see how the 95 percent confidence bounds widen substantially when we extrapolate the estimated transformation beyond the range of the audit data. If this assumption is not true, these bounds should be even wider than they are now represented. If the relationship is nearly

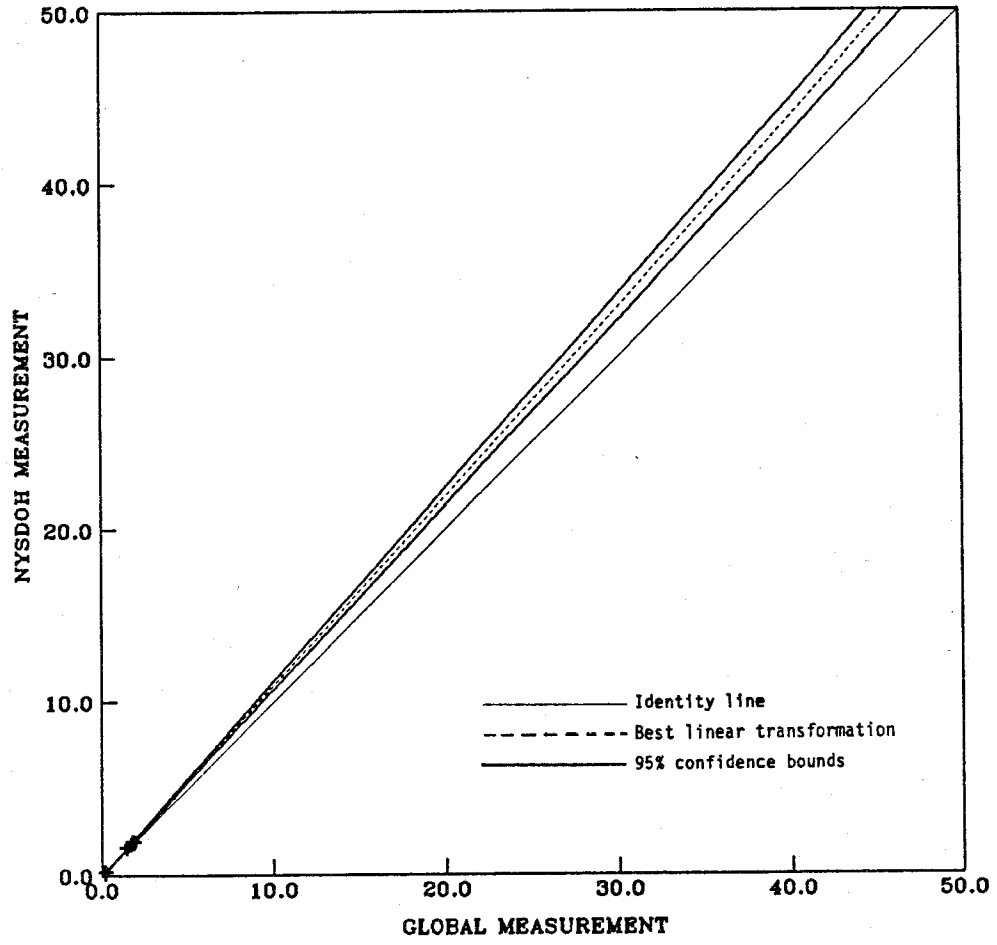


FIGURE 1 **Comparison of means of GLOBAL vs. NYSDOH measurements for CA11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).**

TABLE 4. Representativeness of audit data.

Vaiable	Coverage by Audit Data		
	% of Data[a]	% of Truncated Range[b]	% of Range[c]
CA	23.0	4.6	2.0
MG	5.2	3.2	1.2
K	14.5	14.8	4.9
NA	57.3	23.0	1.5
MN	81.6	27.8	0.9
FE	56.5	7.3	0.2
ALEX	94.0	82.3	2.7
CL	1.8	1.9	0.1
SO4	38.4	12.0	1.9
NO3	58.2	6.7	0.9
SIO2	83.1	67.1	29.1
FTL	86.8	64.5	14.3
DOC	81.4	35.8	2.4
NH4	96.1	146.0	5.9
PHEQ	39.7	54.9[d]	38.2[e]
PHAL	53.7	56.5[d]	32.9[e]
PHAC	58.5	58.6[d]	34.0[e]
ACCO	67.0	33.3	3.4
ALKA	31.8	6.1[d]	1.5[e]
COND	12.0	6.8	2.0
DICE	44.3	6.4	2.2
DICI	38.5	6.7	1.8
PTL	91.7	62.6	1.9
ALT1	76.2	26.1	0.8

a The % of the survey data ≤ the maximum audit mean.

b $$\frac{\text{maximum audit mean}}{\text{95th quantile of the survey data}} * 100\ \%$$

c $$\frac{\text{maximum audit mean}}{\text{maximum survey value}} * 100\ \%$$

d $$\frac{\text{maximum audit mean - minimum audit mean}}{\text{95th quantile - 5th quantile}} * 100\ \%$$

e $$\frac{\text{maximum audit mean - minimum audit mean}}{\text{maximum survey value - minimum survey value}} * 100\ \%$$

linear, but only wavers slightly about this straight line, then the confidence bounds are approximately correct. If, however, the relationship could be a quadratic or more complicated function, then these widths are significantly underestimated particularly when we must extrapolate over a relatively large range, as with the calcium data.

SUMMARY

The decision to transform the NSS data depends then on a careful weighing of the expected improvements in accuracy against the possible losses in precision that can result. This information is summarized in the graphs in Appendix A. The width of the confidence bounds in these graphs however depends heavily on the assumption of the linear relationship of bias to concentration across the range of NSS concentrations. Hence, this assumption should be carefully considered for those variables for which the linearity hypothesis was rejected (indicated by an asterisk in Table 2) and for those variables for which there is no audit data to test this assumption for the entire range of the NSS data (as indicated by Table 4).

These are decisions that should be made by someone very familiar with the laboratory analytical procedures involved and with the ways bias can be introduced. However, we can make the following specific recommendations:

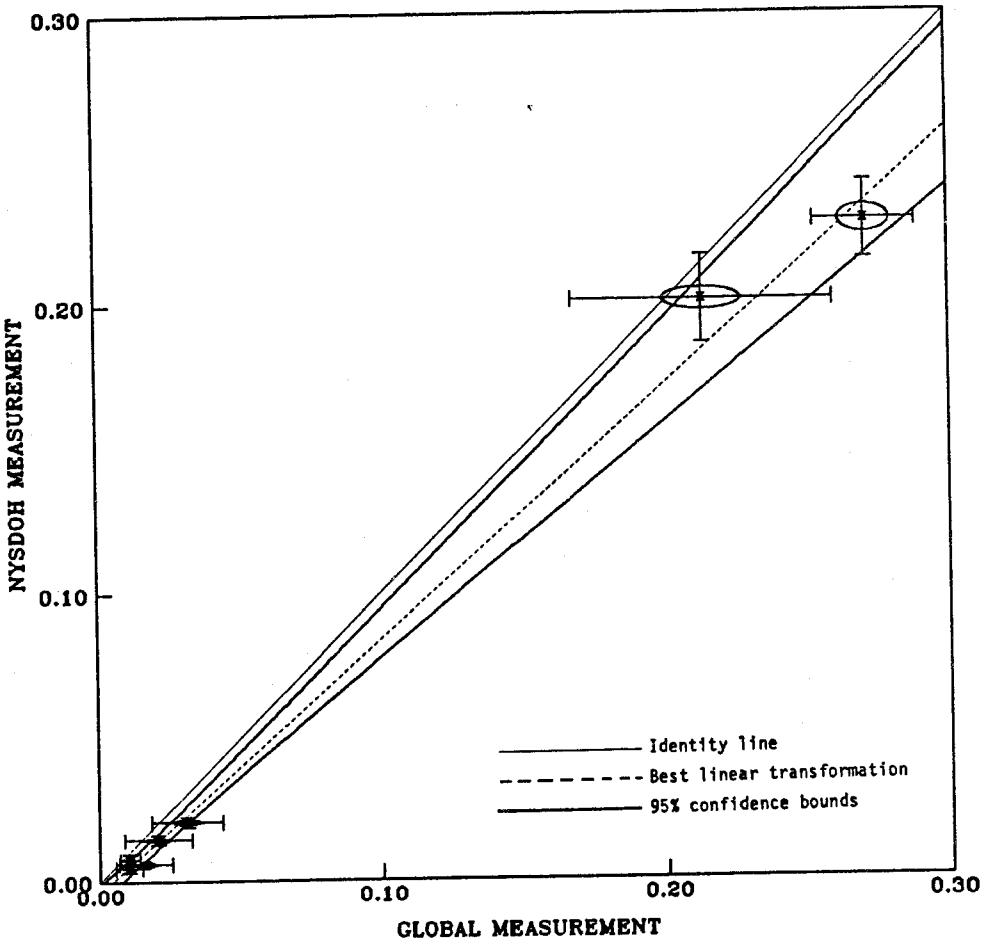
(1) Measurements of those analytes for which the no bias model was recommended, Cl, SO_4, NO_3, DICI, and PTL cannot be improved by transformation.

(2) If policy decisions are based on a measurement range of the data that is within the range of the performance audit data, for instance in the low concentrations for most of the variables in the study, then transformations may be in order.

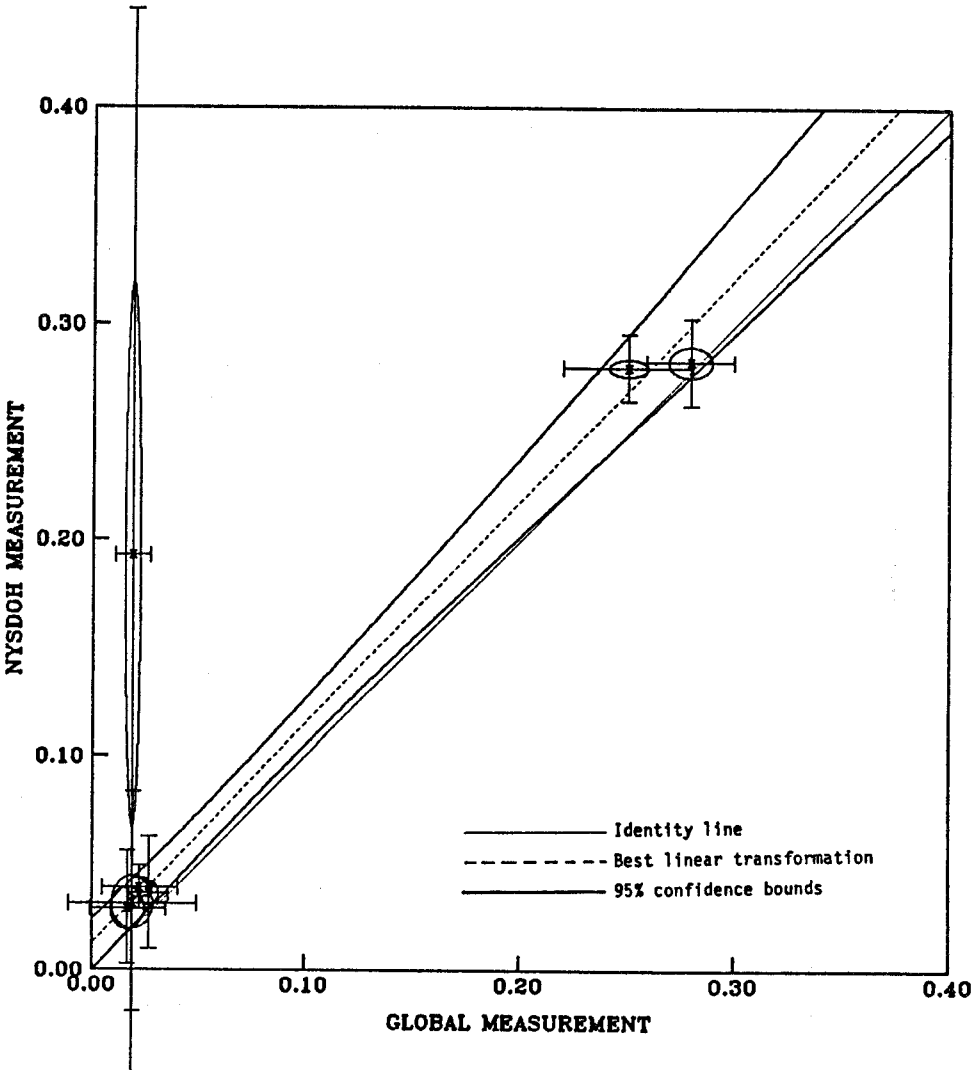
(3) Transformations that involve extrapolation over much of the range of the data should only be carried out after careful consideration of the linearity assumption.

Appendix A

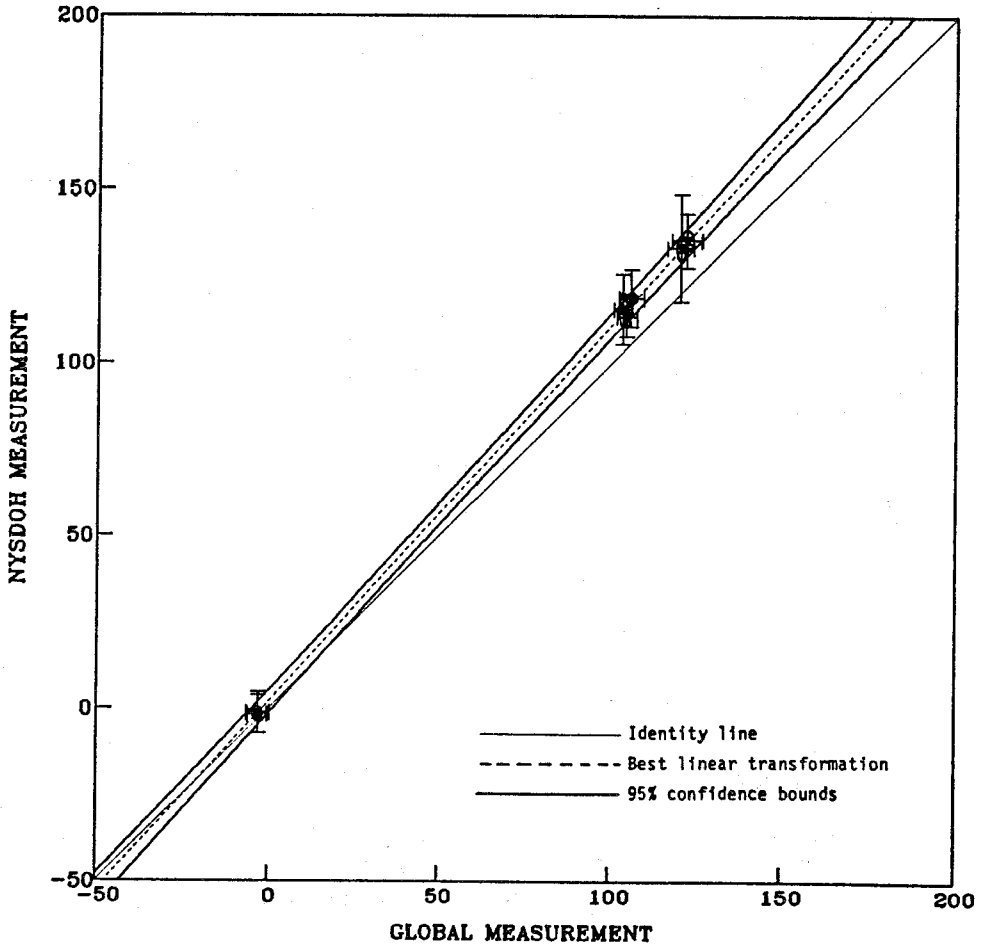
SCATTERPLOTS OF THE MEANS OF PERFORMANCE AUDIT DATA



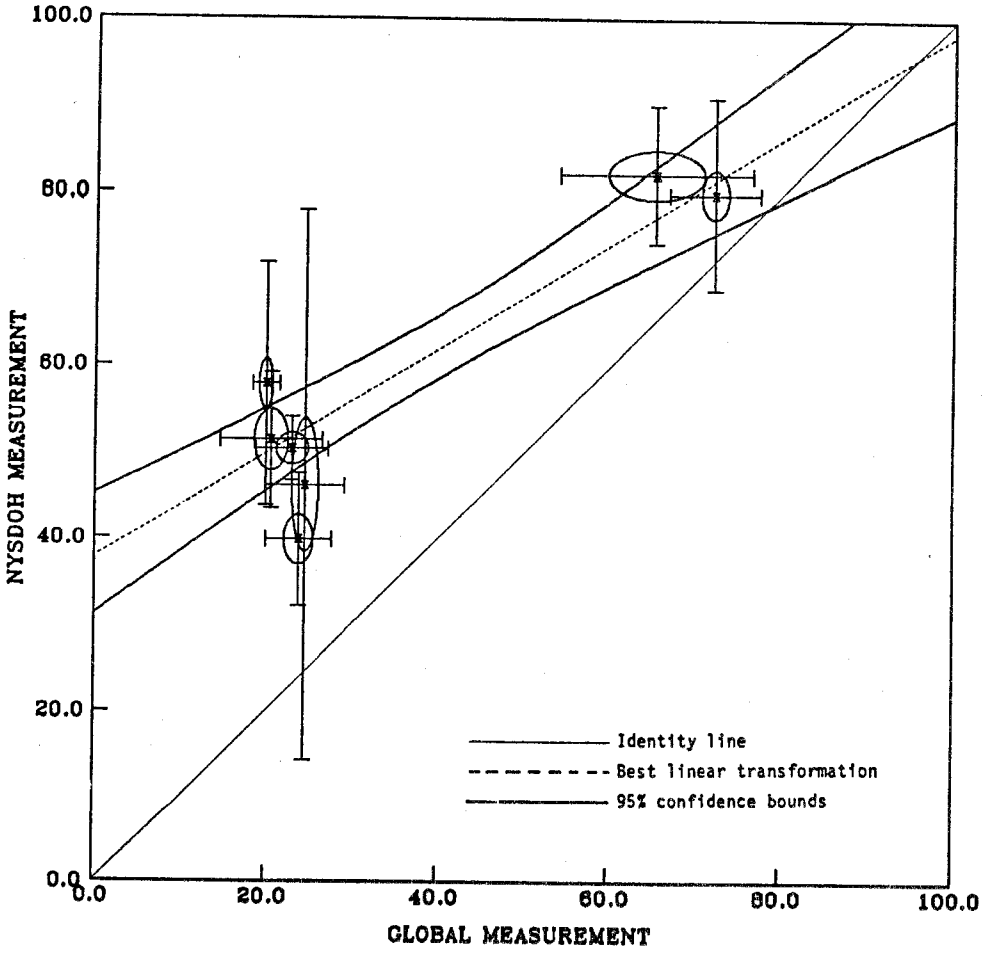
Comparison of means of GLOBAL vs. NYSDOH measurements for ALEX11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



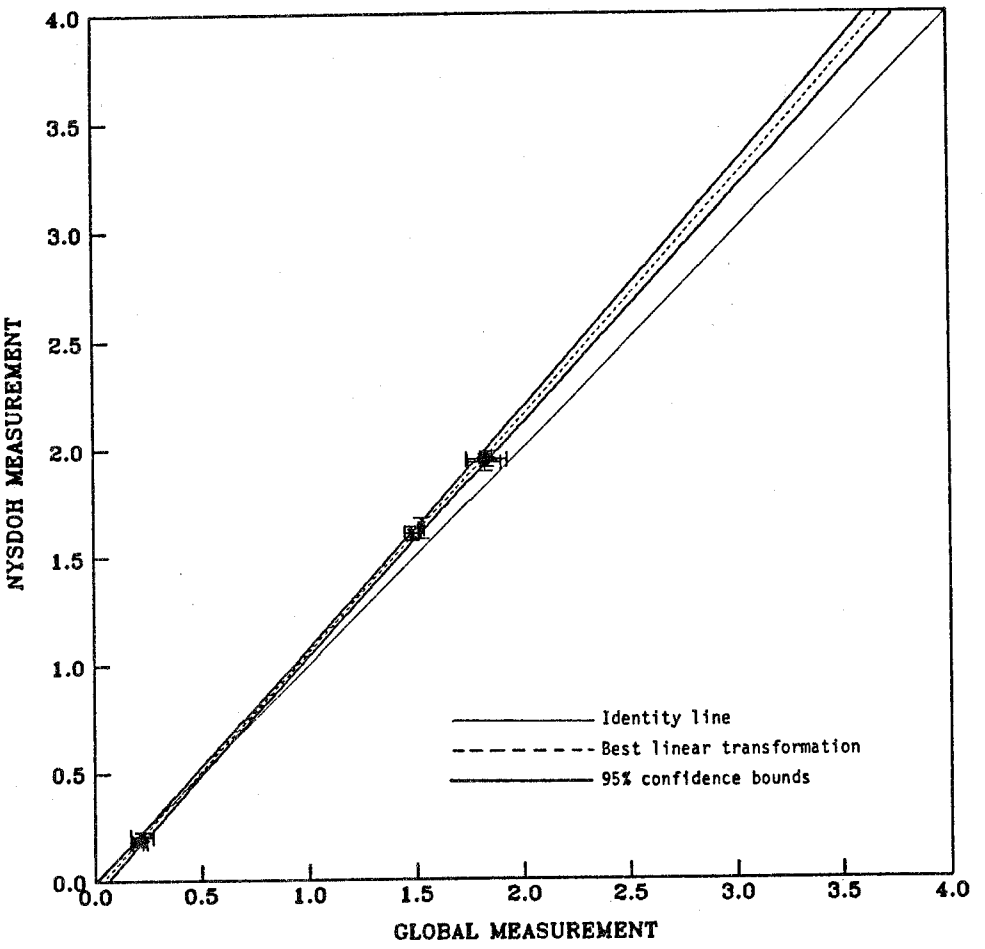
Comparison of means of GLOBAL vs. NYSDOH measurements for ALTL11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



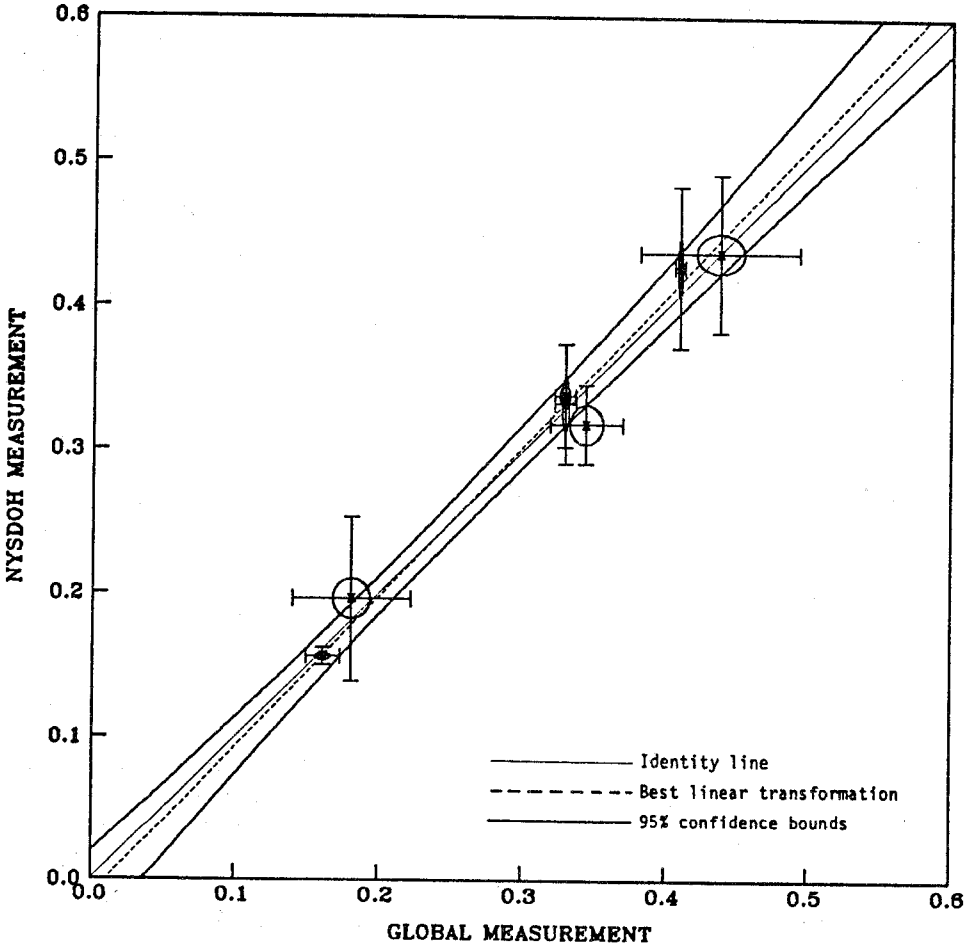
Comparison of means of GLOBAL vs. NYSDOH measurements for ALKA11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



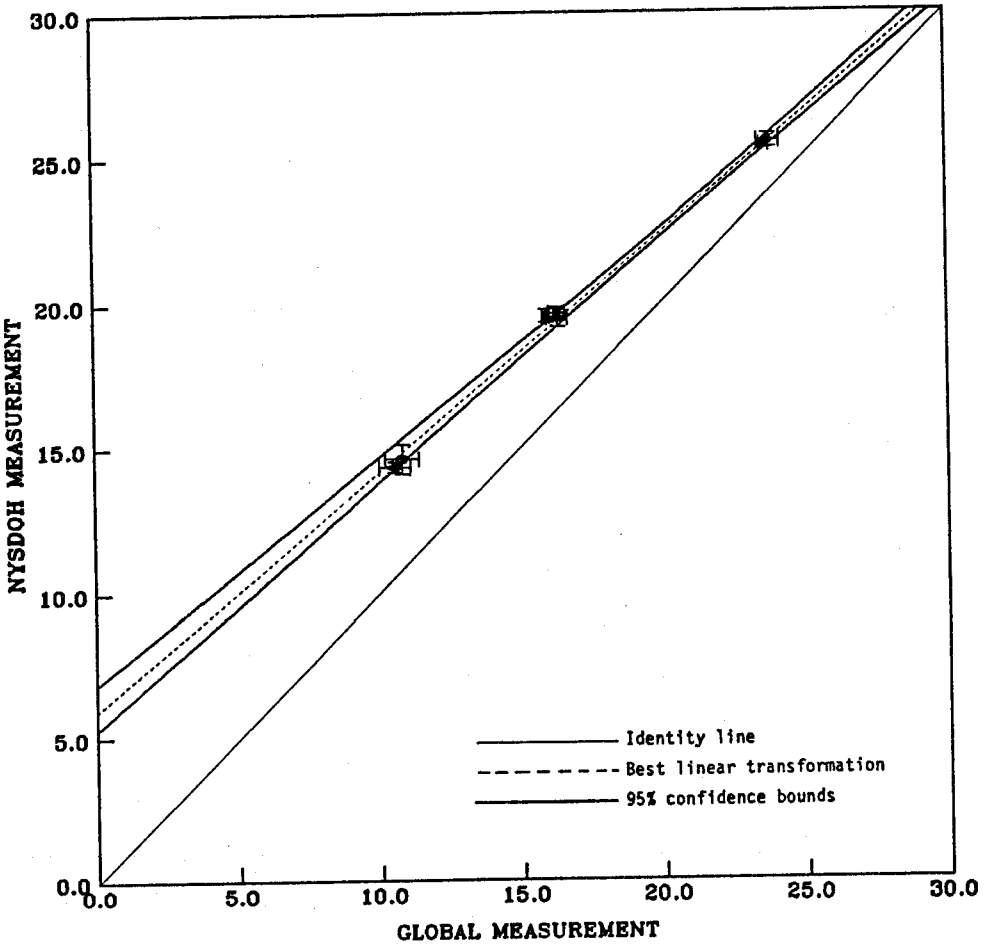
Comparison of means of GLOBAL vs. NYSDOH measurements for ACC011. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



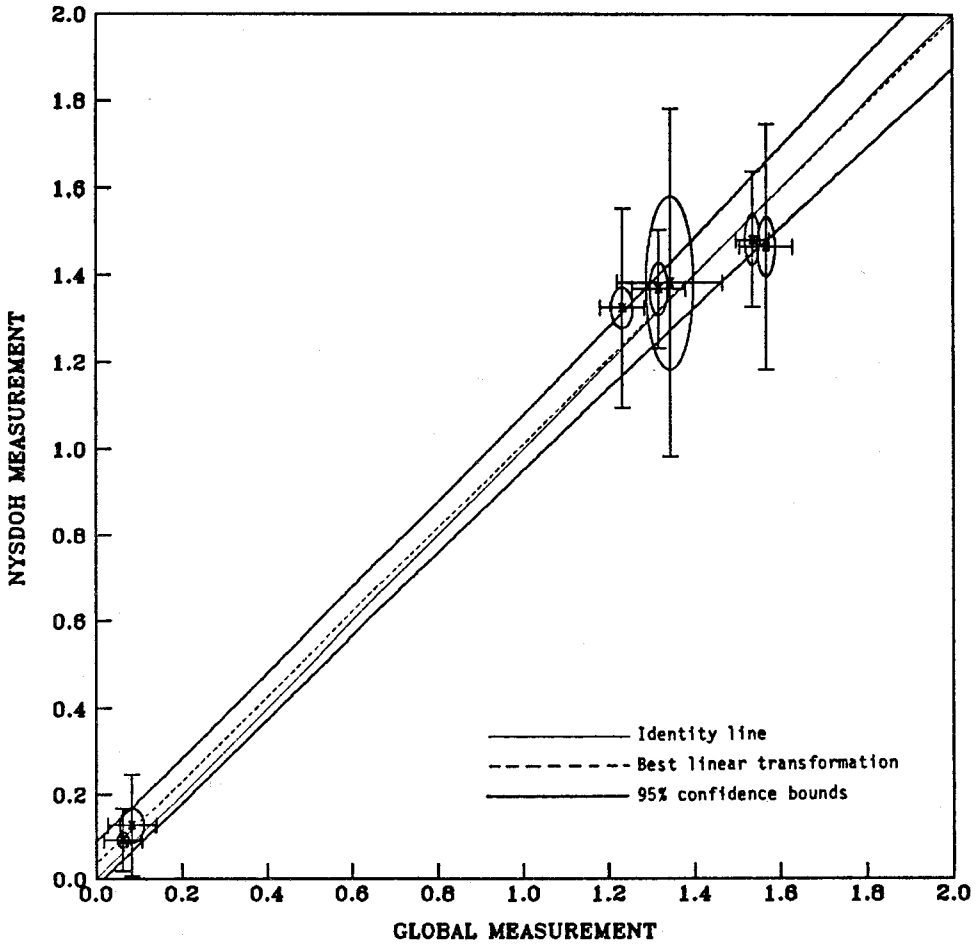
Comparison of means of GLOBAL vs. NYSDOH measurements for CA11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



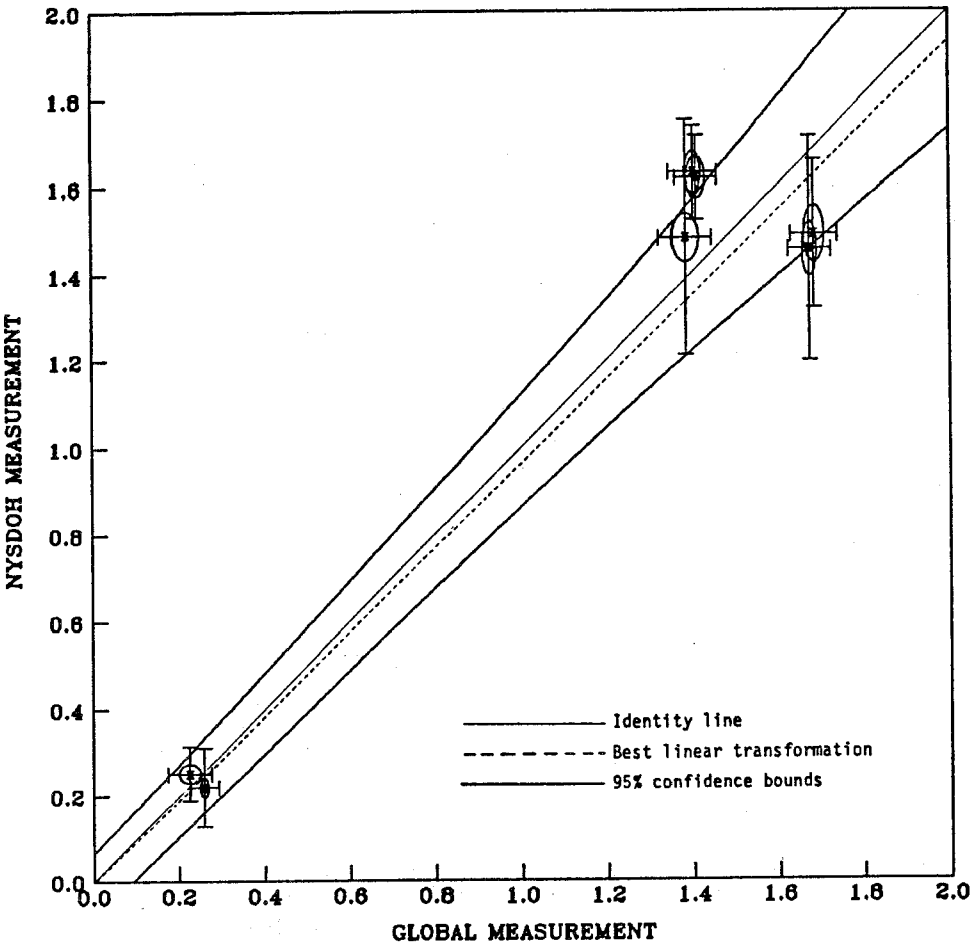
Comparison of means of GLOBAL vs. NYSDOH measurements for CL11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



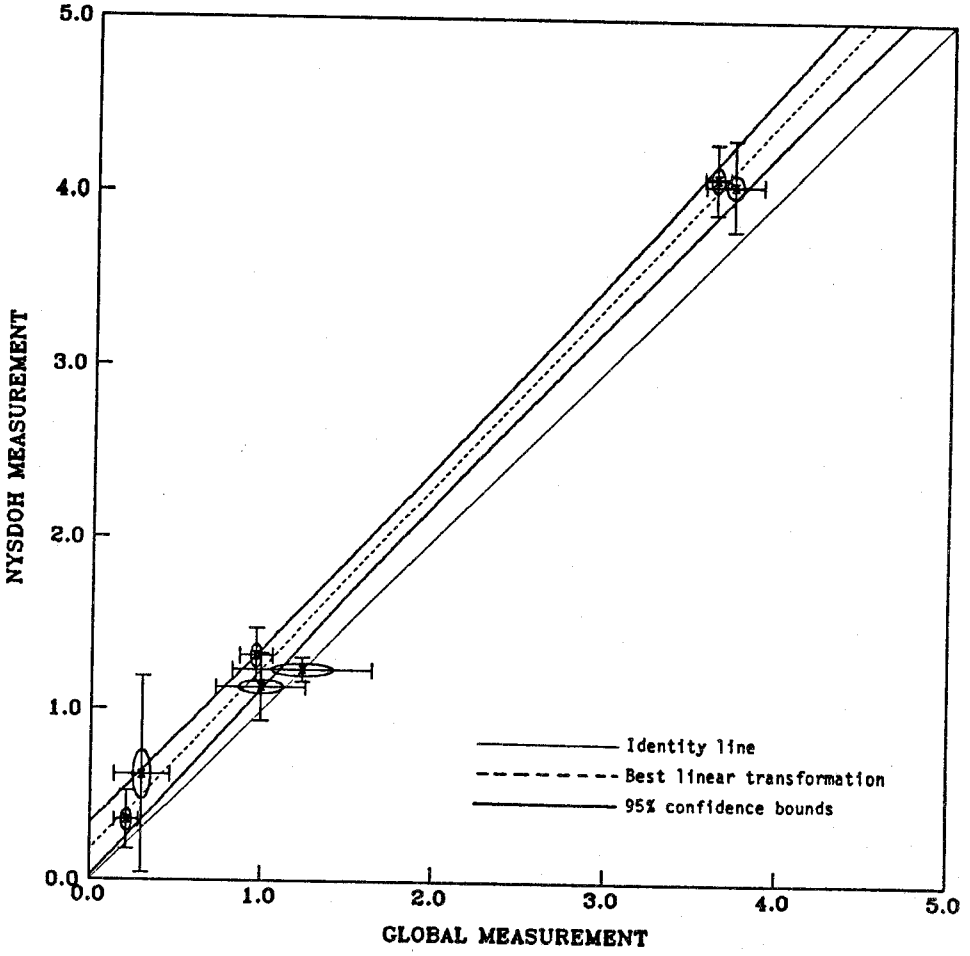
Comparison of means of GLOBAL vs. NYSDOH measurements for COND11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



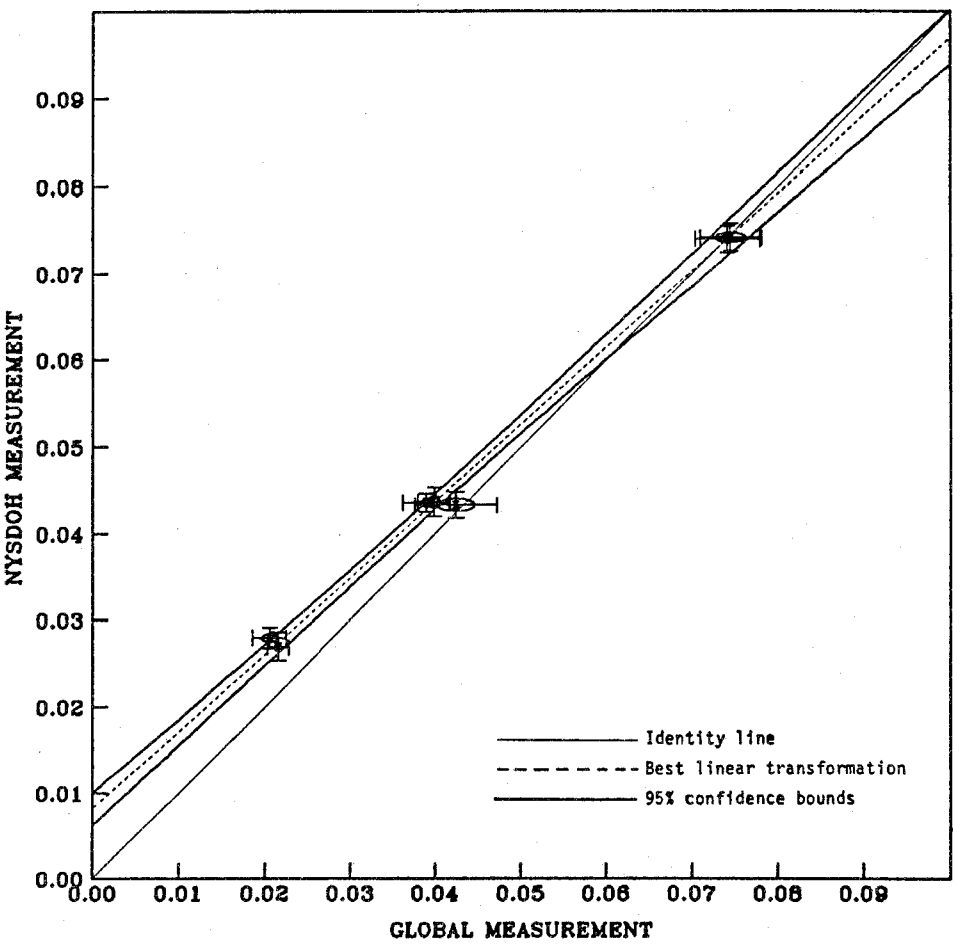
Comparison of means of GLOBAL vs. NYSDOH measurements for DICE11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



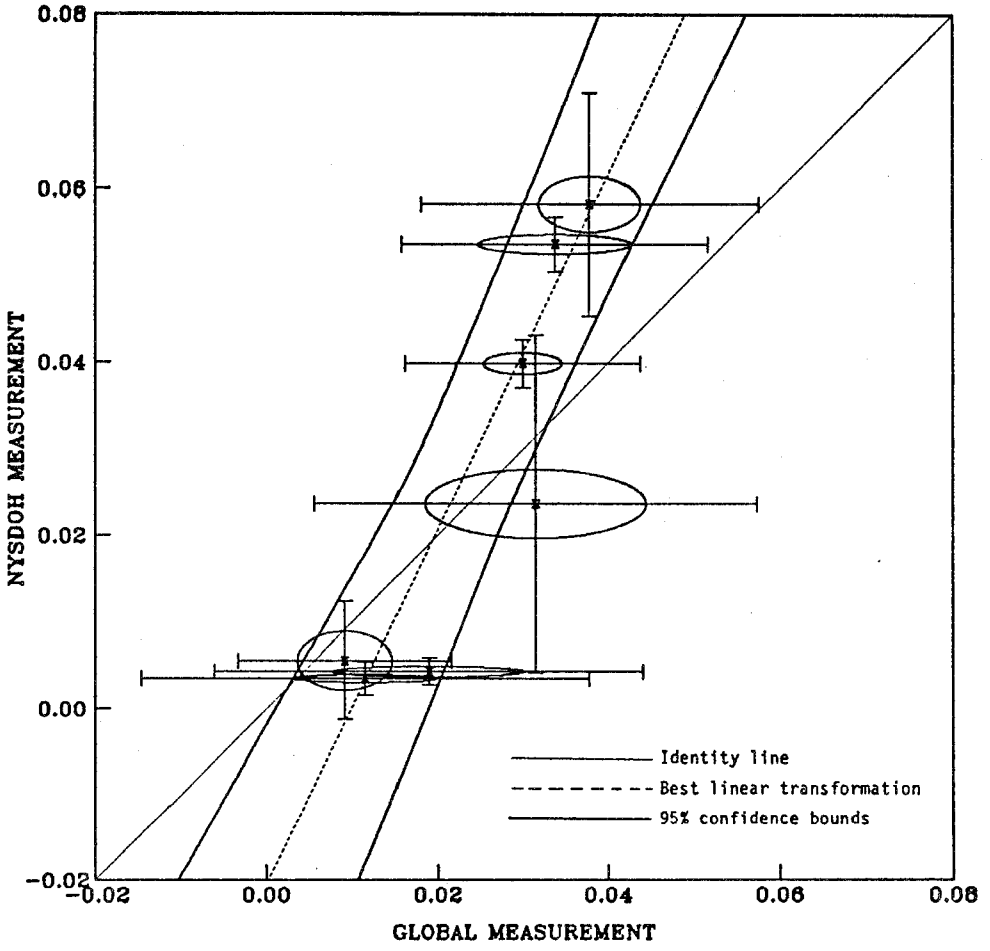
Comparison of means of GLOBAL vs. NYSDOH measurements for DICI11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



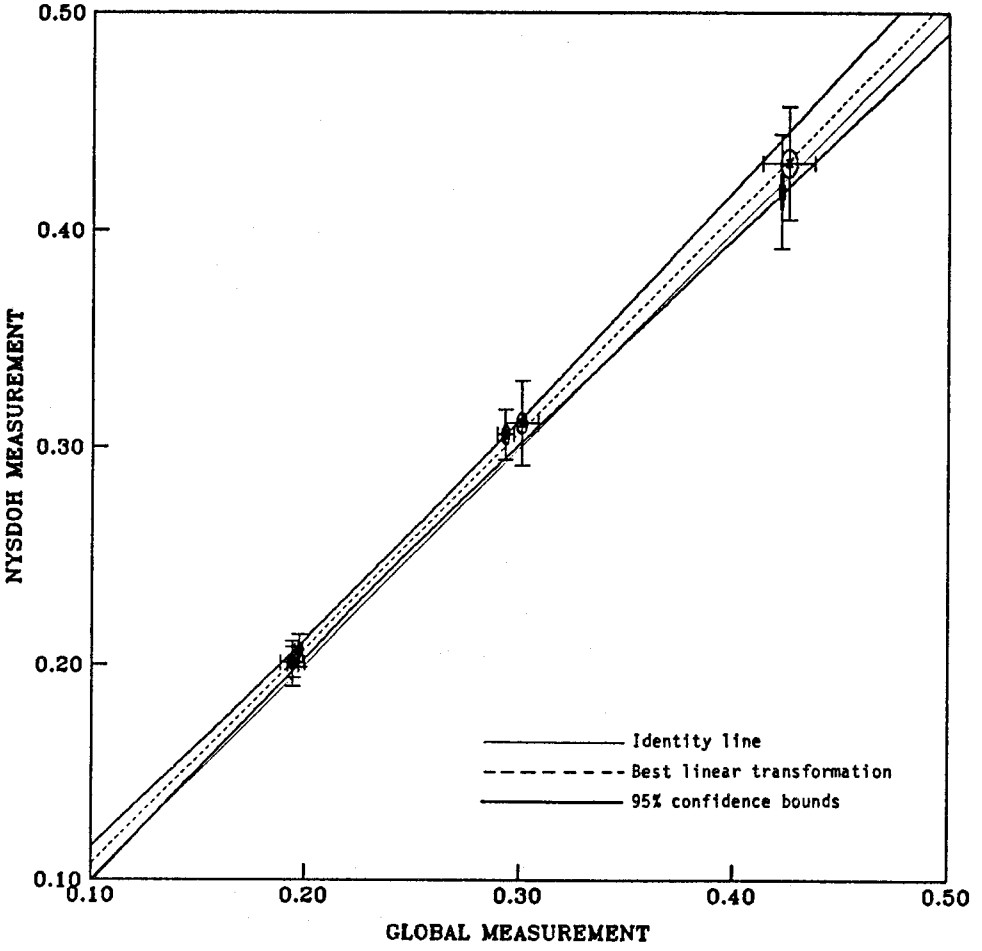
Comparison of means of GLOBAL vs. NYSDOH measurements for DOC11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



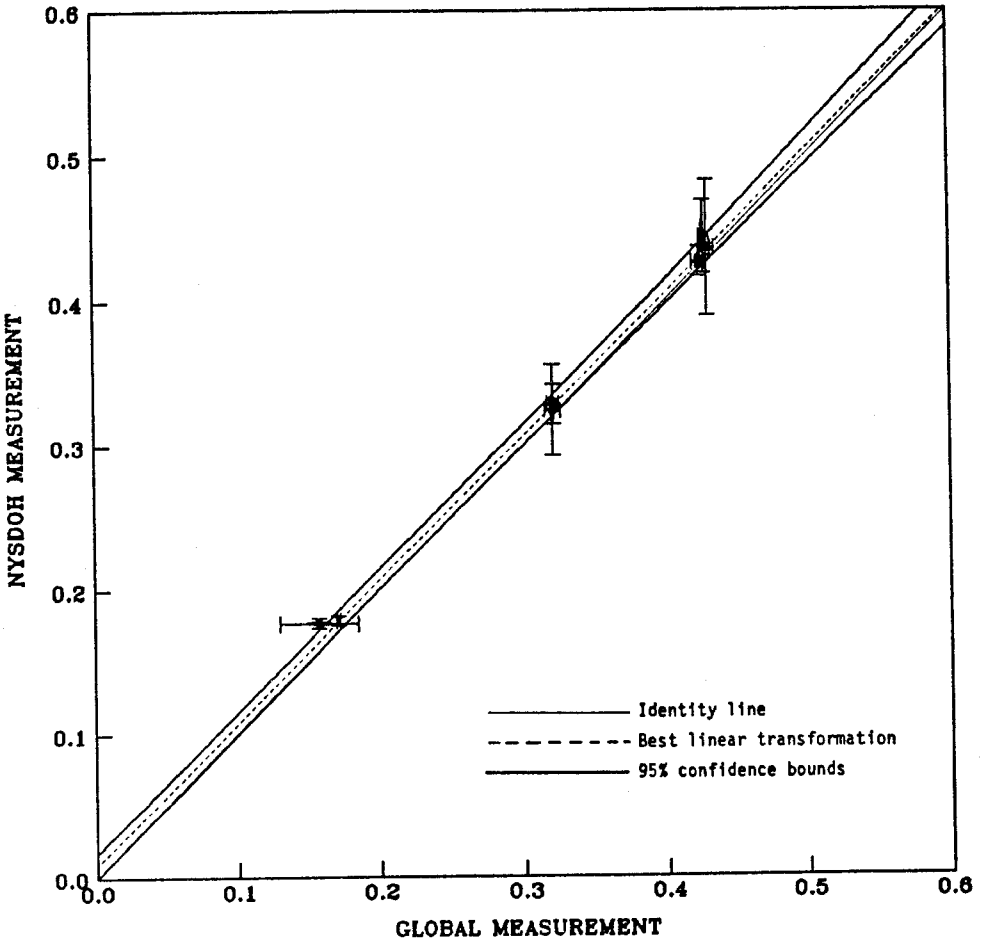
Comparison of means of GLOBAL vs. NYSDOH measurements for FTL11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



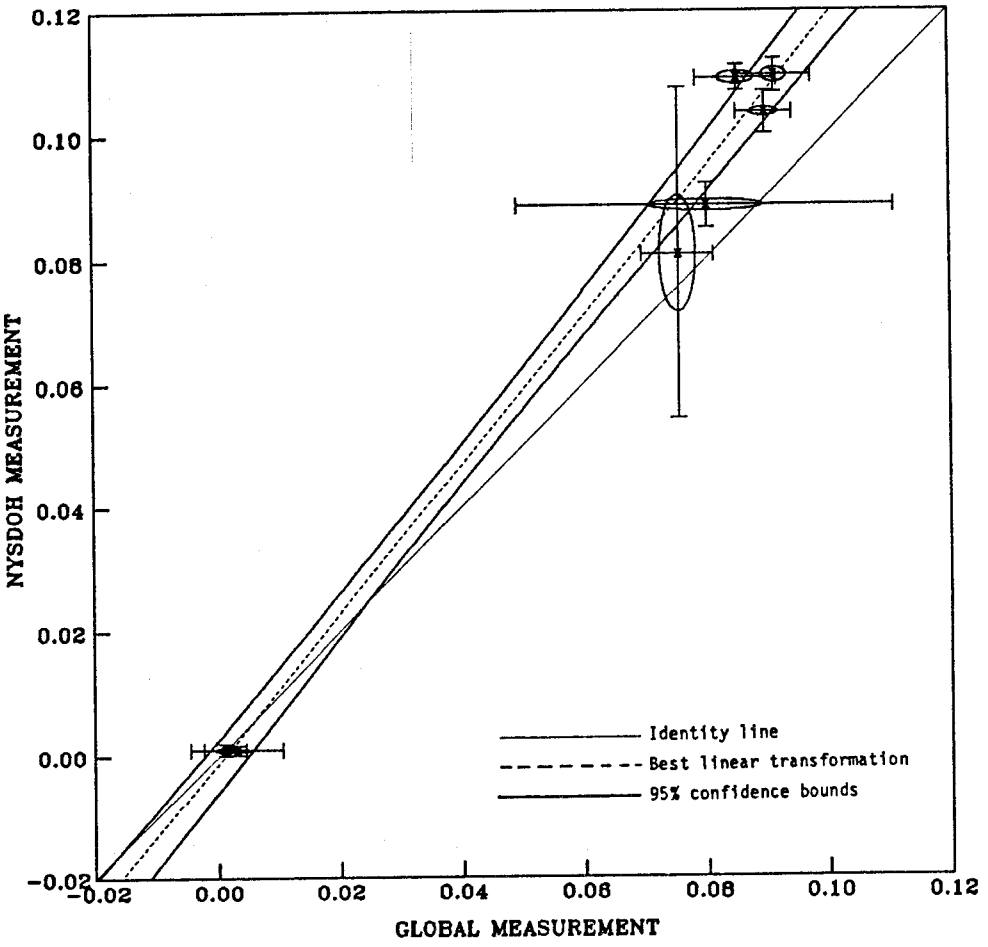
Comparison of means of GLOBAL vs. NYSDOH measurements for FE11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



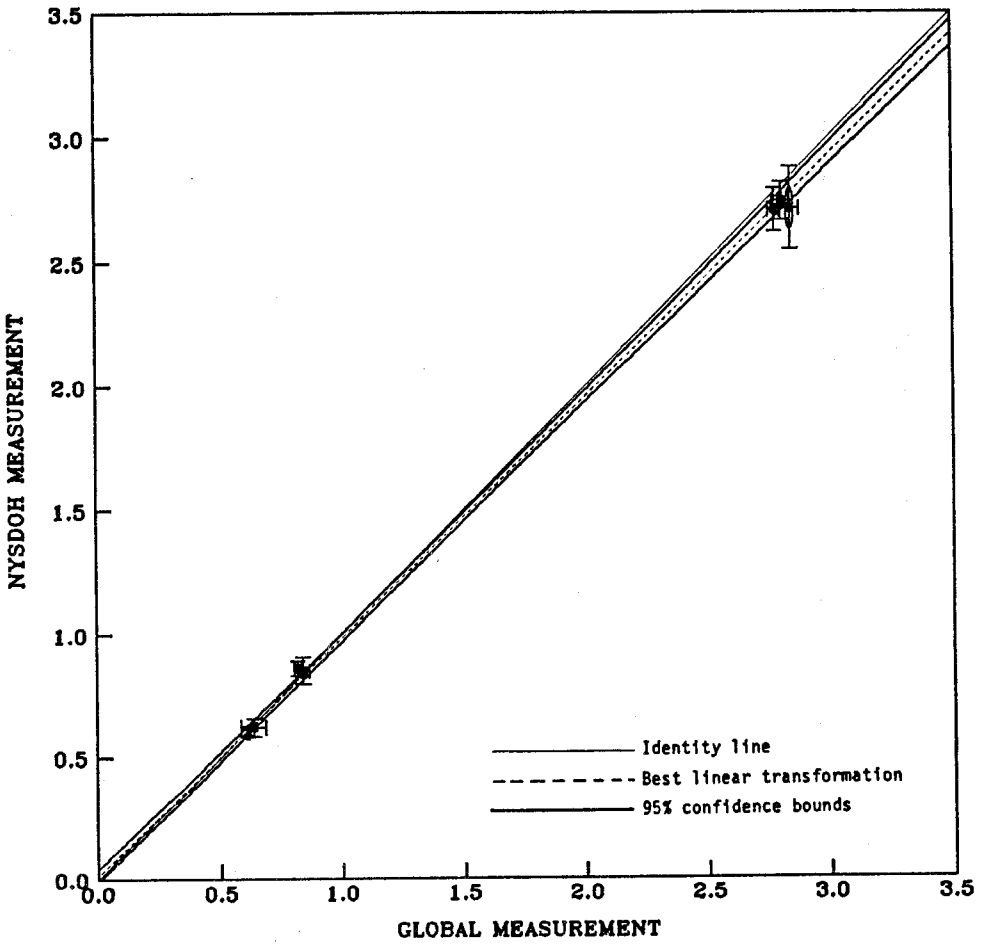
Comparison of means of GLOBAL vs. NYSDOH measurements for K11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



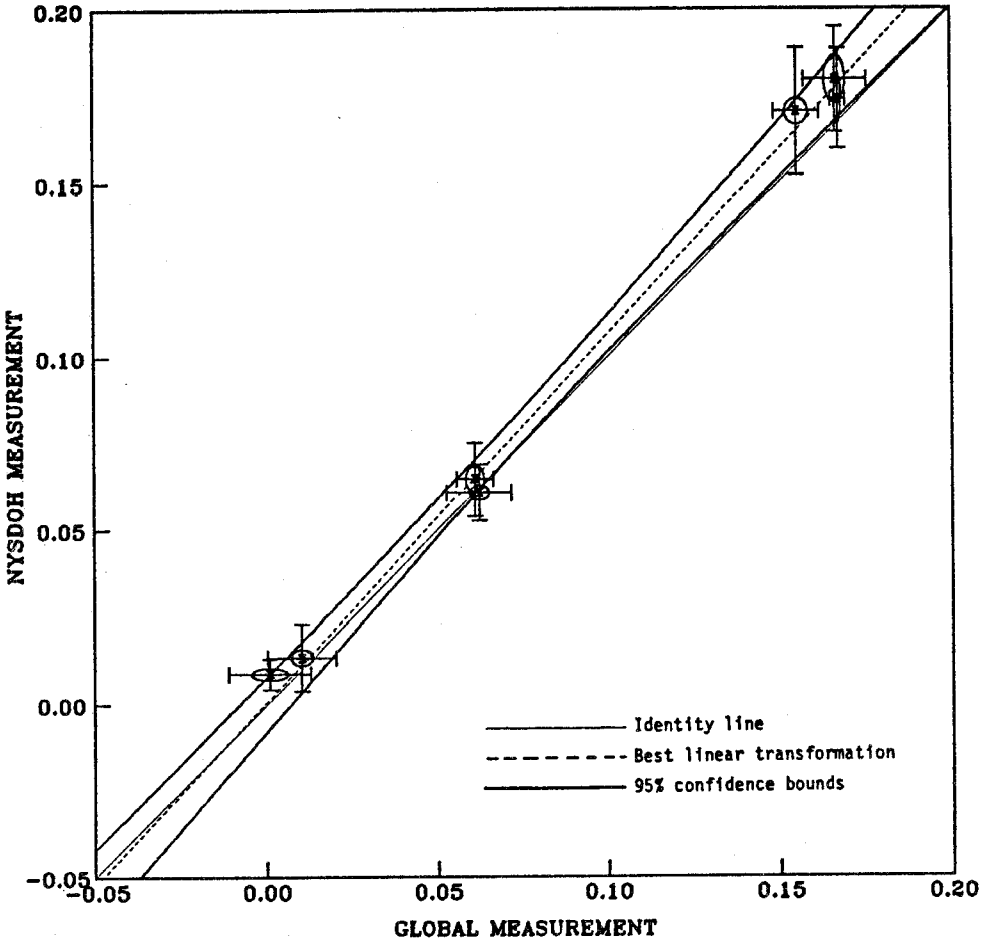
Comparison of means of GLOBAL vs. NYSDOH measurements for MG11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



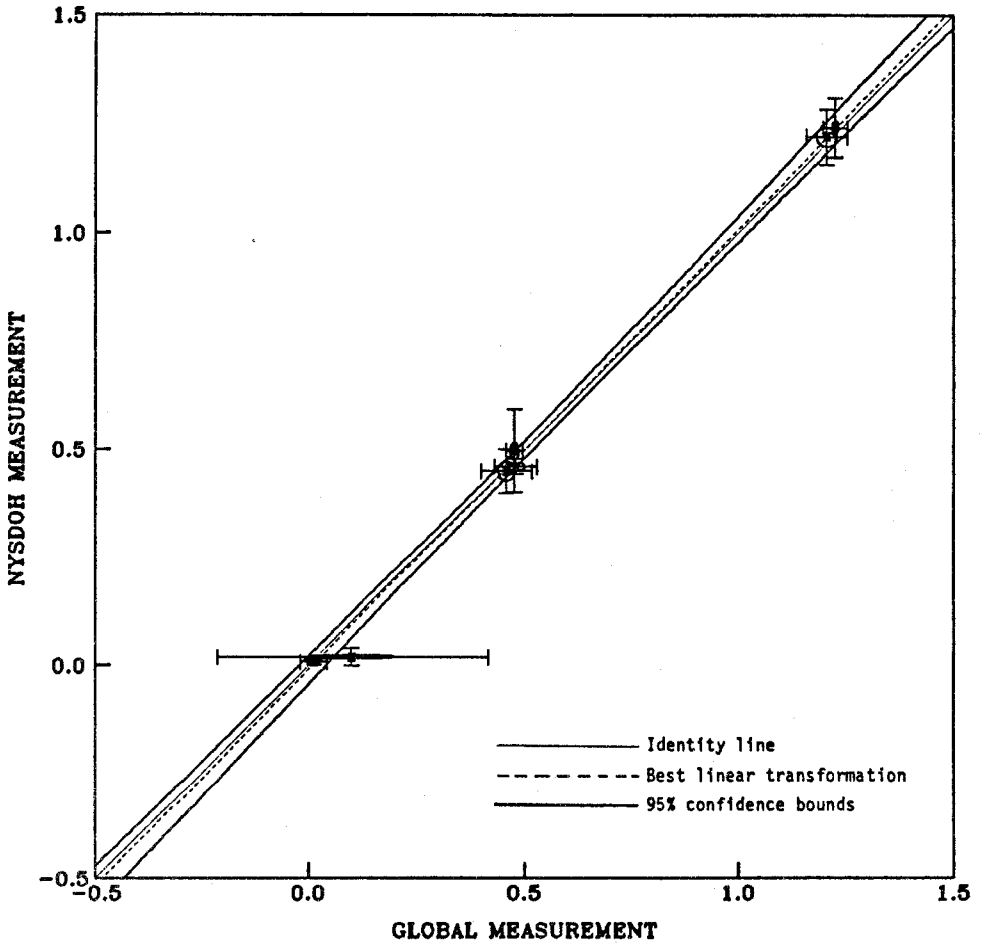
Comparison of means of GLOBAL vs. NYSDOH measurements for MN11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



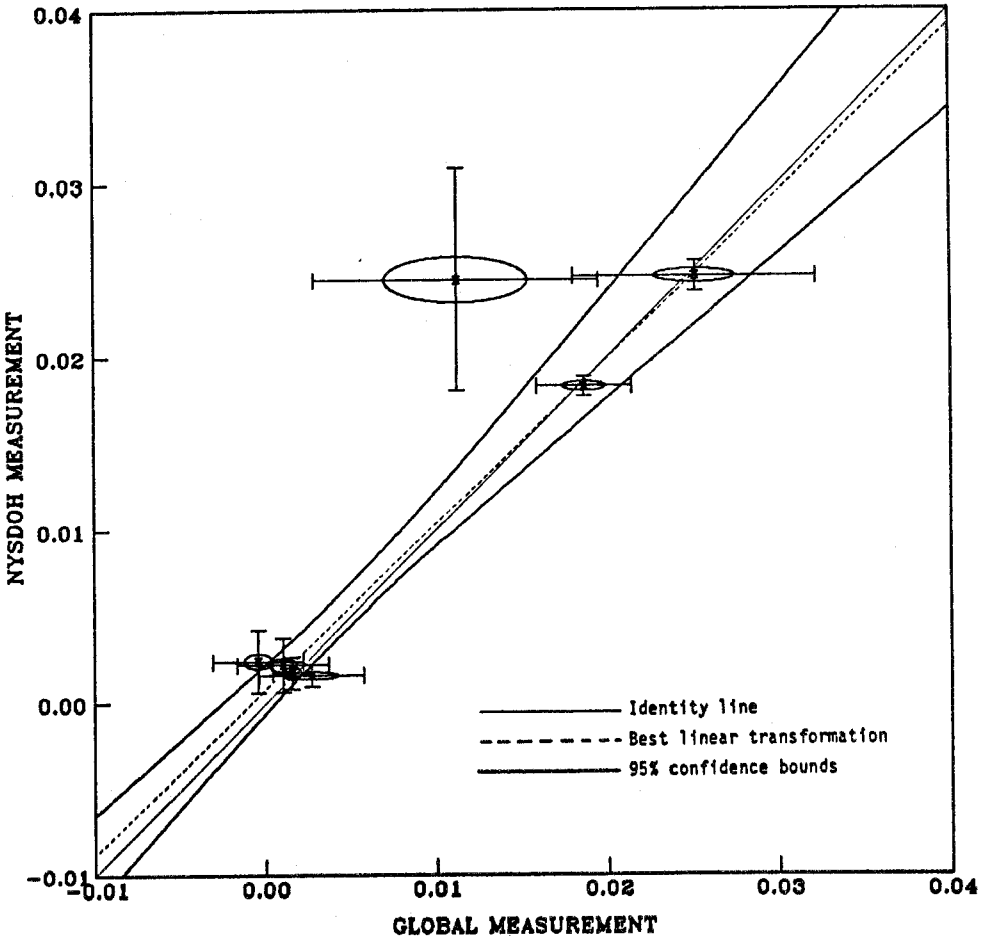
Comparison of means of GLOBAL vs. NYSDOH measurements for NA11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



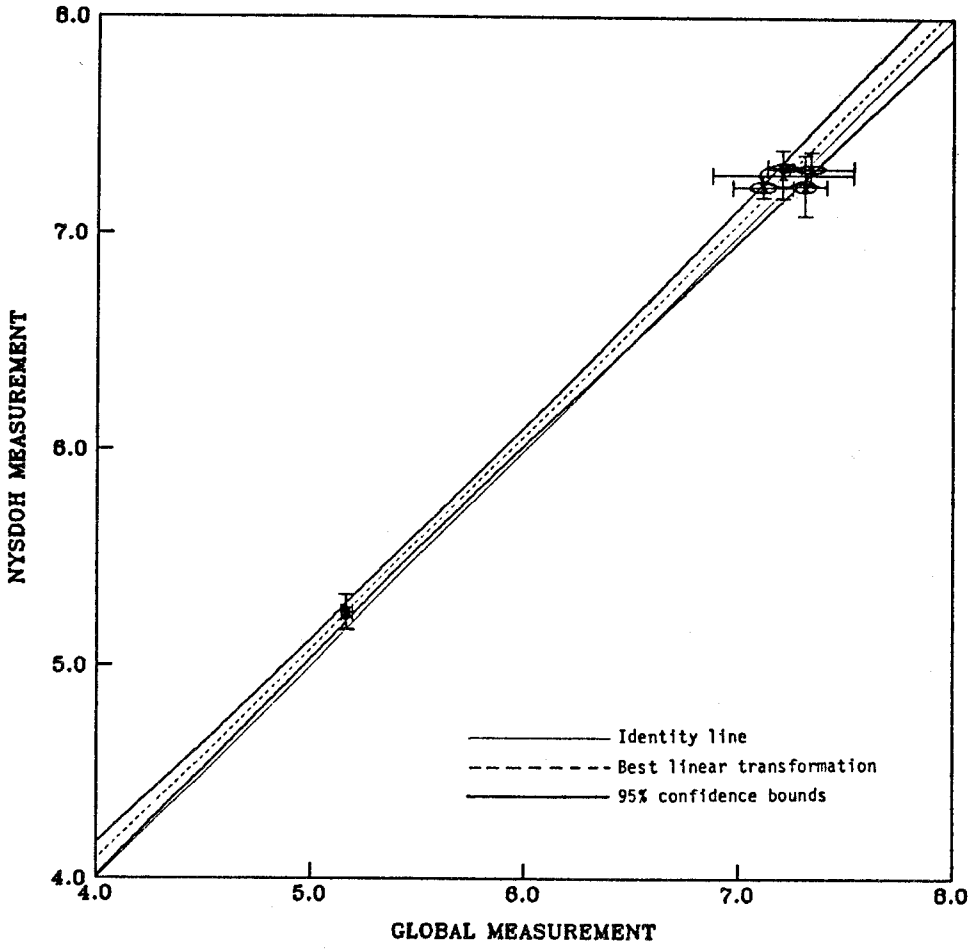
Comparison of means of GLOBAL vs. NYSDOH measurements for NH411. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



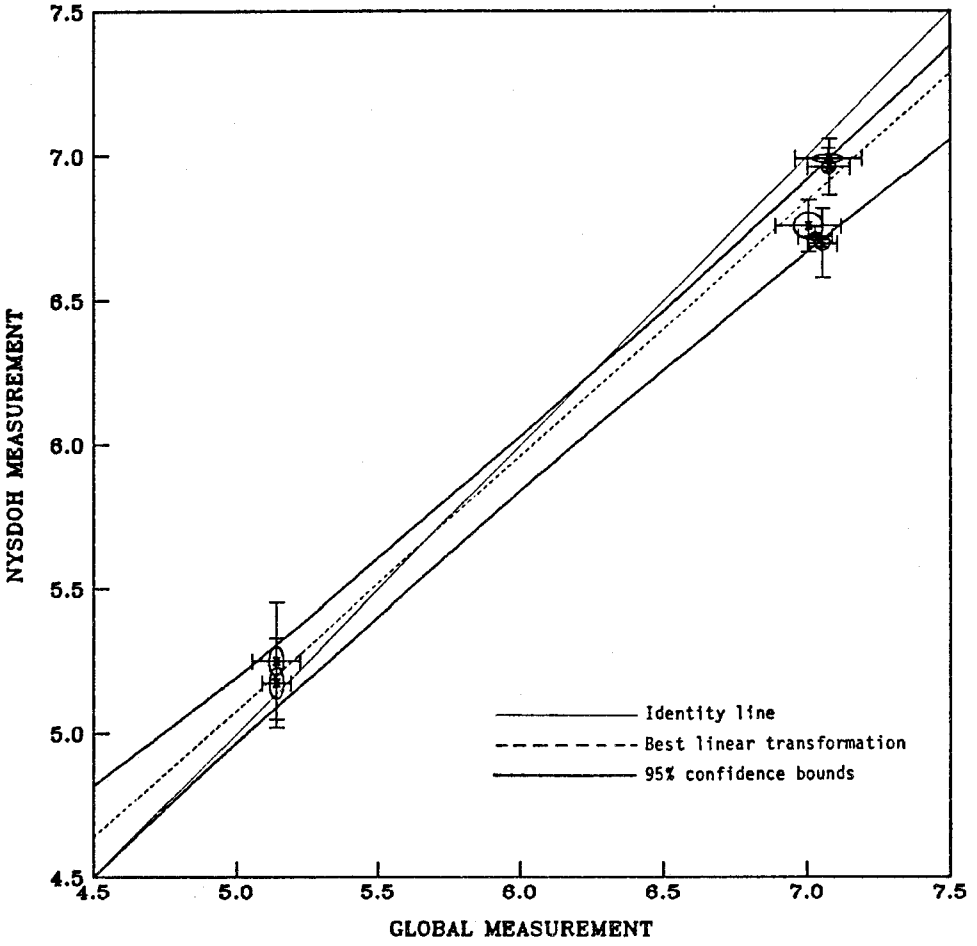
Comparison of means of GLOBAL vs. NYSDOH measurements for NO311. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



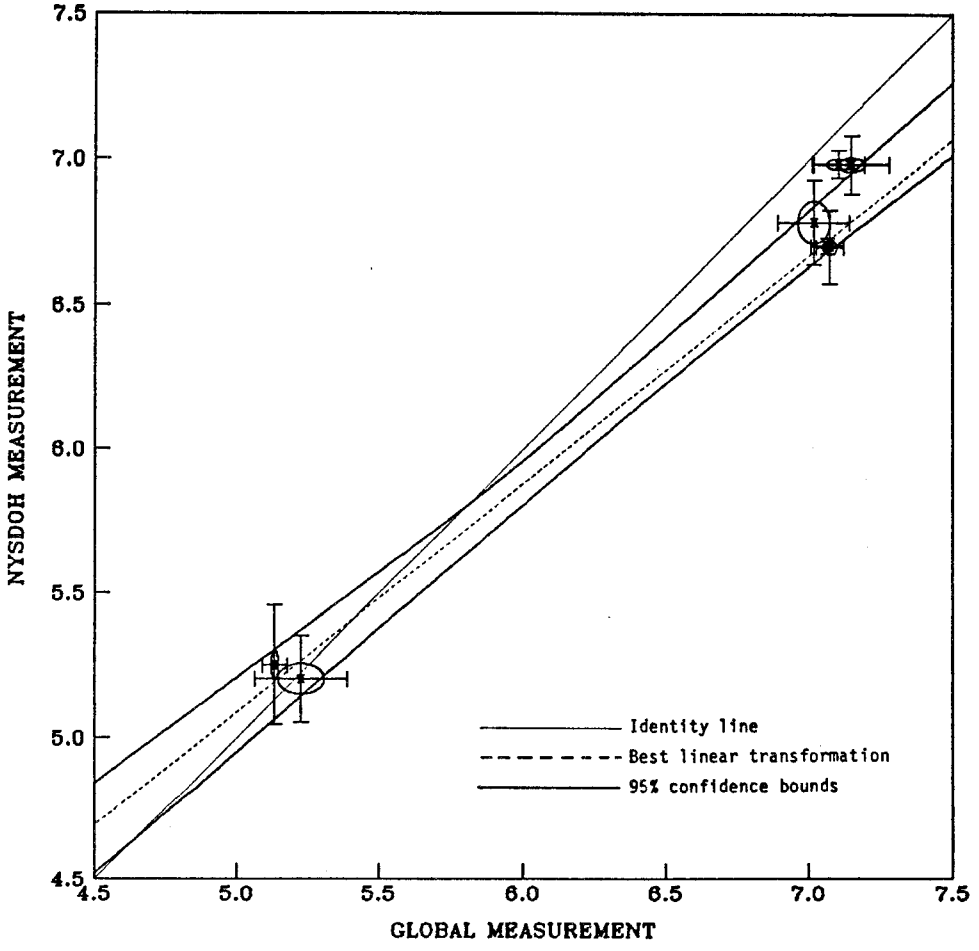
Comparison of means of GLOBAL vs. NYSDOH measurements for PTL11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



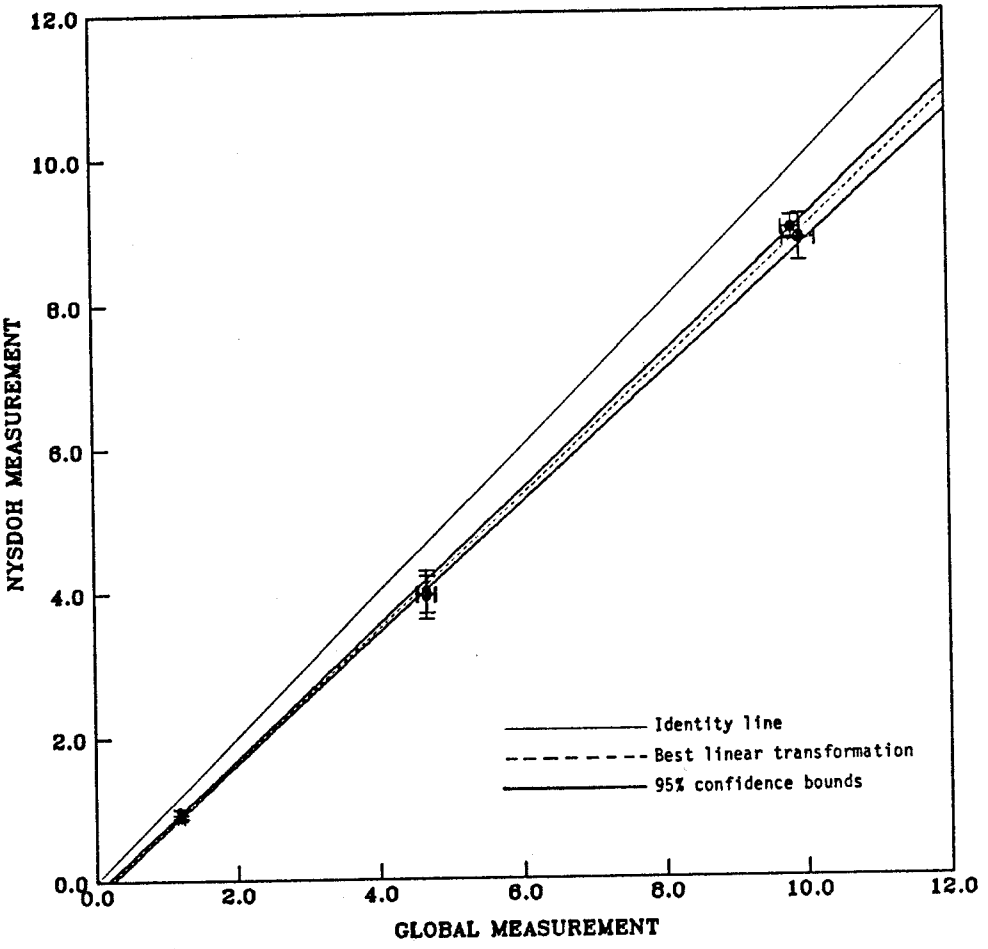
Comparison of means of GLOBAL vs. NYSDOH measurements for PHEQ11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



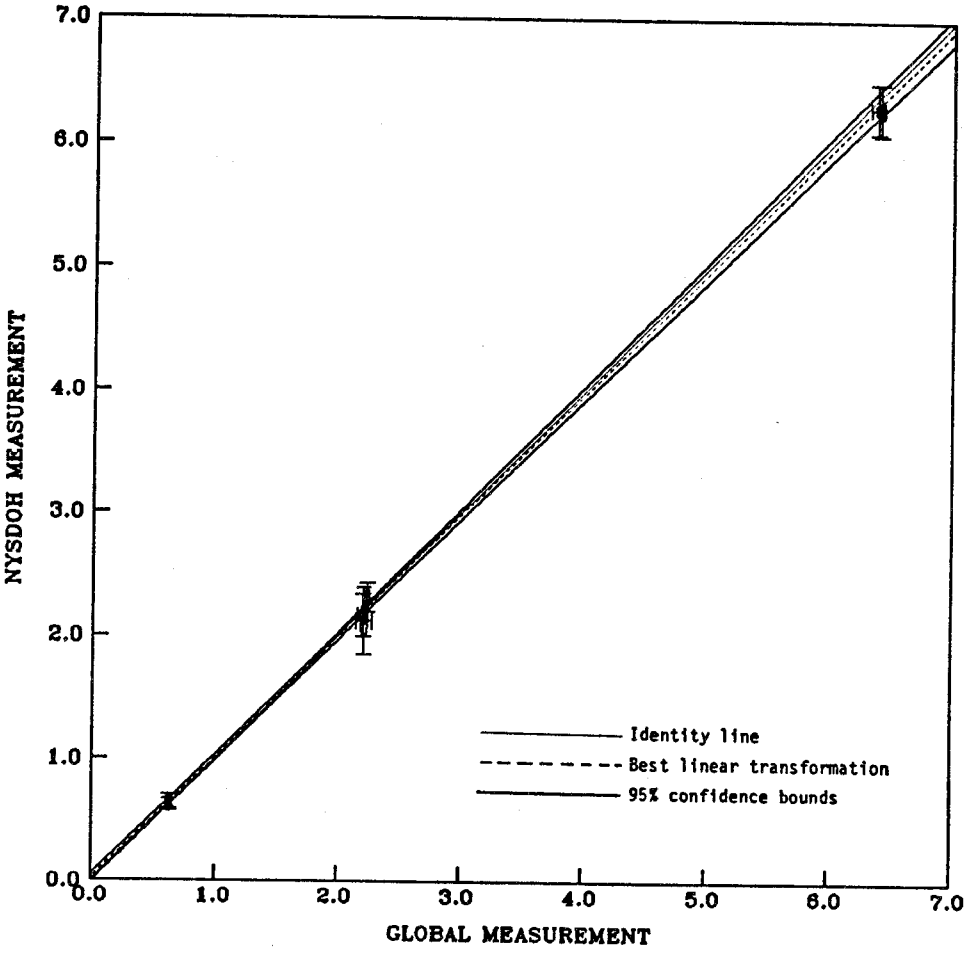
Comparison of means of GLOBAL vs. NYSDOH measurements for PHAL11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



Comparison of means of GLOBAL vs. NYSDOH measurements for PHAC11. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



Comparison of means of GLOBAL vs. NYSDOH measurements for SI0211. Uncertainty shown by bars (standard deviation) and ellipses (standard error).



Comparison of means of GLOBAL vs. NYSDOH measurements for S0411. Uncertainty shown by bars (standard deviation) and ellipses (standard error).

Appendix B

MODELS CONSIDERED AND DERIVATION OF MAXIMUM LIKELIHOOD ESTIMATES

MODELS CONSIDERED

Formally stated, the four linear models and the general bias model outlined in Section 2 are as follows.

If

$$X_{2i-1,j} = \mu_i + \delta_{ij} \ ,$$

then the models are

$$(1) \quad X_{2i,j} = \mu_i + \varepsilon_{ij}$$

$$(2) \quad X_{2i,j} = \mu_i + \alpha + \varepsilon_{ij}$$

$$(3) \quad X_{2i,j} = \mu_i + \beta\mu_i + \varepsilon_{ij}$$

$$(4) \quad X_{2i,j} = \mu_i + \alpha + \beta\mu_i + \varepsilon_{ij}$$

$$(5) \quad X_{2i,j} = \mu_i + \gamma_i + \varepsilon_{ij};$$

where

$\delta_{ij} \sim N(0, \sigma^2_{2i-1})$ are the errors in measurement by Global;
$\varepsilon_{ij} \sim N(0, \sigma^2_{2i})$ are the errors in measurement by NYSDOH; $X_{2i-1,j}$

is the j^{th} measurement by Global on the i^{th} group; $X_{2i,j}$ is the j^{th} measurement by NYSDOH on the i^{th} group; and μ_i is the long-run mean concentration measured by Global. The γ_i in the fifth model are the long-run mean differences in concentration by each laboratory for each audit group. The measurement error terms δ and ε are independent and normally distributed, with variances that may vary from audit group to audit group as well as from laboratory to laboratory.

DERIVATION OF MAXIMUM LIKELIHOOD ESTIMATES

As above, we use the assumption of random normal errors and assign the parameters μ_i ($i = 1, 2, \ldots, 7$) to the means of Global's measurements of each audit group. We get the following density functions for the observations at each of the audit groups for the linear bias model.

For Global:

$$f_{2i-1} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_{2i-1}} \left(\frac{1}{\sigma_{2i-1}}\right)^{n_{2i-1}} \exp\left[-1/2 \sum_{j=1}^{n_{2i-1}} \left(\frac{X_{2i-1,j} - \mu_i}{\sigma_{2i-1}}\right)^2\right]$$

For NYSDOH:

$$f_{2i} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_{2i}} \left(\frac{1}{\sigma_{2i}}\right)^{n_{2i}} \exp\left[-1/2 \sum_{j=1}^{n_{2i}} \left(\frac{X_{2i,j} - ((1+\beta)\,\mu_i + \alpha)}{\sigma_{2i}}\right)^2\right]$$

where

n_{2i-1} = audit subgroup sample size corresponding to Global

σ^2_{2i-1} = variances corresponding to Global

n_{2i} = audit subgroup sample size corresponding to NYSDOH

σ_{2i}^2 = variances corresponding to NYSDOH

$i = 1, 2, \ldots, 7.$

The density function for the linear bias model is then:

$$L_4 = \prod_{j=1}^{14} f_j$$

Maximum likelihood estimates (m.l.e.) for the linear bias model are obtained by maximizing this equation as a function of the parameters $\mu, \ldots, \mu_7$; $\sigma_1^2, \ldots, \sigma_{14}^2$; α; and β. Or, more conveniently, we could minimize -2 log L_4.

Many computer algorithms are available for solving minimization problems such as this. A "grid search" type of algorithm, however, provides the most consistently reliable answer. A grid search involves literally evaluating the likelihood function at every possible value of the parameter, or at least on a reasonably fine mesh over all the possible values, and empirically determining which parameter values maximize the likelihood function. This approach has the additional advantage of providing the range of α and β that decides acceptance of the $\alpha = \beta = 0$ hypothesis (hypothesis number 1 in Section 3), i.e., the subset of all α and β for which the (logged) general linear bias likelihood equation is within one-half (5.99) of its (logged) value at the m.l.e. solution. This range provides, by definition, the 95 percent confidence intervals for the best fit m.l.e.'s of linear interlaboratory bias.

Our grid search algorithm takes advantage of the fact that μ and σ can be defined implicitly in terms of each other and by α and β at the m.l.e. solution as follows:

$$\mu_i = \frac{\dfrac{X_{2i-1}}{\sigma^2_{2i-1}/n_{2i-1}} + \dfrac{(1+\beta)(X_{2i}-\alpha)}{\sigma^2_{2i}/n_{2i}}}{\left[\dfrac{1}{\sigma^2_{2i-1}/n_{2i-1}} + \dfrac{(1+\beta)^2}{\sigma^2_{2i}/n_{2i}}\right]} \quad ,$$

$$\sigma^2_{2i-1} = \sum_{j=1}^{n_{2i-1}} \left(X_{2i-1,j} - \bar{X}_{2i-1}\right)^2 \Big/ n_{2i-1} + \left(\bar{X}_{2i-1} - \mu_i\right)^2 \quad ,$$

$$\sigma^2_{2i} = \sum_{j=1}^{n_{2i}} \left(X_{2i,j} - \bar{X}_{2i}\right)^2 \Big/ n_{2i-1} + \left\{\bar{X}_{2i-1} - [(1+\beta)\mu_i + \alpha]\right\}^2 \quad ,$$

for $i = 1, 2, \ldots, 7$.

This system of equations quickly converges to a single answer by iteratively recalculating the first in terms of the next two and the last two in terms of the first. These equations can be obtained by taking first partial derivatives of -2 log L_4, setting them equal to 0, and solving.

This is saying that for any given line drawn through the data there is just one set of μ and σ that maximizes the likelihood function. By taking advantage of this, the problem of maximizing L_4 is reduced to that of a simple grid search over the possible range of α and β.

The m.l.e for L_1, L_2, and L_3 in Section 2 can be found by simply replacing α and/or β with zero as appropriate in each of the above equations.

For the general bias model the density functions for the Global audit groups are unchanged, but for NYSDOH they become:

$$f_{2i} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_{2i}} \left(\frac{1}{\sigma_{2i}}\right)^{n_{2i}} \exp\left[-1/2 \sum_{j=1}^{n_{2i}} \left(\frac{x_{2i,j} - (\mu_i + \gamma_i)}{\sigma_{2i}}\right)^2\right]$$

where $\gamma_i = 1, 2, \ldots, 7$ are the seven bias terms used to account for the deviations between the NYSDOH and Global means. Maximum likelihood estimates for this density are obtained simply as:

μ_i = observed Global means,

γ_j = observed difference between the Global and NYSDOH means,

$$\sigma^2_{2i-1} = \sum_{j=1}^{n_{2i-1}} (X_{2i-1,j} - \bar{X}_{2i-1})^2 \; / \; n_{2i-1}$$

and

$$\sigma^2_{2i} = \sum_{j=1}^{n_{2i}} (X_{2i,j} - X_{2i}) \; / \; n_{2i}$$

GLOSSARY

absolute bias	The difference between a measured value and the true value. (See "accuracy.")
acceptance criteria	The range in which the analytical measurement of a quality assurance or quality control sample is expected to be; measurements outside that range are considered suspect.
accuracy	The closeness of a measured value to the true value of an analyte. For this report, accuracy is calculated as: $\frac{\bar{X} - R}{R} 100$ where the $\bar{X}$ is the mean of all measured values and R = the theoretical or index value.
acid-neutralizing capacity (ANC)	Total acid-combining capacity of a water sample determined by titration with a strong acid. Acid-neutralizing capacity includes alkalinity (carbonate species) as well as other basic species (e.g., borates, dissociated organic acids, alumino-hydroxy complexes).
air equilibration	The process of bringing a sample aliquot to equilibrium with standard air (300 ppm CO_2) before analysis; used with some pH and dissolved inorganic carbon measurements.
aliquot	Fraction of a sample prepared for the analysis of particular constituents; sent in a separate container to the analytical laboratory.
among-batch precision	The estimate of precision that includes effects of different laboratories and day-to-day difference within a single laboratory, calculated from audit sample data.
analyte	A chemical species that is measured in a water sample.
analytical laboratory	In this report, a laboratory under contract with the U.S. Environmental Protection Agency to analyze water samples shipped from the processing laboratory.
analytical laboratory duplicate	Aliquot of a sample that is split in the analytical laboratory. The aliquots are analyzed in the same batch.
anion	A negatively charged ion.

anion-cation balance

In an electrically neutral solution such as water, the total charge of positive ions (cations) equals the total charge of negative ions (anions). In this report, anion-cation balance is expressed as percent ion balance difference and is calculated as follows:

$$\left(\frac{\Sigma \text{ anions} - \Sigma \text{ cations} + \text{ANC}}{\Sigma \text{ anions} + \Sigma \text{ cations} + \text{ANC} + 2[H^+]} \right) 100$$

where:

$$\Sigma \text{ anions} = [Cl^-] + [F^-] + [NO_3^-] + [SO_4^{2-}],$$

$$\Sigma \text{ cations} = [Na^+] + [K^+] + [Ca^{2+}] + [Mg^{2+}] + [NH_4^+],$$

$$\text{ANC} \simeq$$ alkalinity (the ANC value is included in the calculation to account for the presence of unmeasured ions such as organic ions), and

$$[H^+] = (10^{-pH}) \times 106 \ \mu\text{eq/L}.$$

anion deficit

The concentration (in microequivalents per liter) of measured anions less the measured cations.

ASTM Type I reagent-grade water

Deionized water that meets American Society for Testing and Materials (ASTM) specifications for Type I reagent-grade water (ASTM, 1984) and that has a measured conductance of less than 1 μS/cm at 25 °C. This water is used to prepare blank samples and reagents.

audit sample

In this survey, a standardized water sample submitted to an analytical laboratory for the purpose of checking overall performance in sample analysis. Natural audit samples in the NSS-I were lake water; synthetic audit samples were prepared by diluting concentrates of known chemical composition in ASTM Type I reagent-grade water.

base site

A location providing support for sampling personnel during field sampling operations.

base-neutralizing capacity (BNC)

Total base-combining capacity of a water sample due to dissolved CO_2, hydronium, and hydroxide; determined by titration with a strong base.

batch

A group of samples processed and analyzed together. A batch consists of all samples (including quality assurance and quality control samples) that are assigned a unique batch identification number and that are processed and sent to one analytical laboratory in one day. In the NSS-I, a batch did not exceed 40 samples.

batch ID

The numeric identifier for each batch.

bias	The systematic difference between values or sets of values.
blank sample	A sample of ASTM Type I reagent-grade water analyzed as a quality assurance or quality control sample during the NSS-I (see calibration, reagent, processing laboratory, and field blanks).
calculated conductance	The sum (as microsiemens per centimeter) of the theoretical specific conductances of all measured ions in a sample.
calibration blank sample	A sample of ASTM Type I reagent-grade water defined as a 0 mg/L standard used in standardizing or checking the calibration of analytical instruments; also used to determine instrument detection limits.
calibration curve	The linear regression of the analytical instrument response to a set of calibration standards (varying in concentrations) from which the linear dynamic range is determined.
carryover	An artifact of the analyte carried from a sample of high concentration to a subsequent sample or samples as a result of incomplete rinsing of an instrument or apparatus.
cation	A positively charged ion.
circumneutral	Close to neutrality in pH (near pH 7).
closed system	Method of measurement in which a water sample is collected and analyzed for pH and dissolved inorganic carbon without exposure to the atmosphere. These samples were collected in syringes directly from the sampling apparatus and were analyzed in the processing laboratory.
comparability	A measure of data quality that allows the similarity within and among data sets to be established confidently.
completeness	A measure of data quality that is the quantity of acceptable data actually collected relative to the total quantity that was expected to be collected.
component (of a sampling system)	For this report, any of the sets of procedures used to get a sample from the stream to analysis. Major components include sample collection, sample processing, and sample analysis. Other components include sample transport, sample shipment, and data reporting. Together, these components are the system.

conductance balance

A comparison of the measured conductance of a water sample (in microsiemens per centimeter) to the equivalent conductances (in microsiemens per centimeter) of each ion measured in that water sample at infinite dilution. In this report, conductance balance is expressed as percent conductance difference and is calculated as follows:

$$\left(\frac{\text{calculated conductance - measured conductance}}{\text{measured conductance}}\right)100$$

The ions used to calculate conductance are Ca, Cl^-, CO_3^{2-}, H^+, HCO_3^-, K, Mg, Na, NO_3^-, OH^-, and SO_4^{2-}.

confidence interval (95% and 99%)

The range of values, calculated from an estimate of the mean and standard deviation, between the confidence limits. This interval has a high probability (a 95 or 99 percent level of confidence) of containing the true population value.

confidence limits

Two statistically derived values or points, one below and one above a statistic, that provide a given degree of confidence that a population parameter falls between them.

control limits

Two values between which the analytical measurement of a quality assurance or quality control sample is expected to be; measurements outside these limits are suspect.

Cubitainer

A 3.8-L container made of semirigid polyethylene used to transport field samples (routine, duplicate, blank) from the stream site to the processing laboratory.

data base

All computerized results of the survey, which include the raw, verified, validated, and enhanced data sets as well as back-up and historical data sets.

data base audit

An accounting of the data and of the data changes in the data base; includes changes made within a data set and among all data sets.

data package

A report, generated by an analytical laboratory for each batch of samples analyzed, that includes analytical results, acid-neutralizing capacity titration data, ion chromatography specifications, analysis dates, calibration and reagent blank data, quality control check sample results, and analytical laboratory duplicate results.

data qualifier

Annotation applied to a field or analytical measurement related to possible effects of the quality of the datum. (See definitions for "flags" and "tags".)

data quality objectives

Accuracy, detectability, and precision limits established before a sampling effort. Also include comparability, completeness, and representativeness.

Data Set 1	Set of files containing raw data. (See definition for "raw data set").
Data Set 2	Set of files containing verified data. (See definition for "verified data set.")
Data Set 3	Set of files containing validated data. (See definition for "validated data set.")
Data Set 4	Set of files containing final, enhanced stream data.(See definition for "enhanced data set.")
detectability	The capability of an instrument or method to determine a measured value for an analyte above either zero or background levels.
detection limit quality control check sample	A quality control check sample that has a specified theoretical concentration and that is designed to check instrument calibration at the low end of the linear dynamic range.
dissolved inorganic carbon (DIC)	A measure of the dissolved carbon dioxide, carbonic acid, bicarbonate and carbonate anions that constitute the major part of ANC in a stream.
dissolved organic carbon (DOC)	In a water sample, the organic fraction of carbon that is dissolved or unfilterable (for this report, the fraction that will pass a filter of 0.45-μ m pore size).
enhanced data set	Data Set 4. Missing values or errors in the validated data set are replaced by substitution values; duplicate values are averaged; negative values (except for ANC and BNC) are set to zero. Values for field blank, field duplicate, and performance audit samples are not included in this data set.
equivalent	Unit of ionic charge; the quantity of a substance that either gains or loses one mole of protons or electrons.
exception	An analytical result that does not meet the expected QA or QC criteria and for which a data flag is generated.
exception program	A computer program in AQUARIUS-II that identifies or flags analytical results classified as exceptions.
extractable aluminum	Operationally defined aluminum fraction that is extracted by the procedure used during the NSS-I; this measurement is intended to provide an indication of the concentration of the aluminum species that may be available in a form toxic to fish.
field audit sample	A standardized water sample submitted to field laboratories to check overall performance in sample analysis by field laboratories and by analytical laboratories. Field natural audit samples were lake water; field synthetic audit samples were prepared by diluting concentrates of known chemical composition into ASTM Type I reagent-grade water.

field blank sample	A sample prepared at the processing laboratory consisting of ASTM Type I reagent-grade water and transported to the stream site by the field sampling crews. At the stream site, the blank was processed through the sampling apparatus. These samples were analyzed at the processing laboratory (except for pH and DIC) and analytical laboratories and were employed in the calculation of system decision and system detection limits and instrument detection limits.
field duplicate sample	The second sample of stream water collected by the sampling crew at the same location and depth at the stream site immediately after the routine sample, in accordance with standardized protocols.
field duplicate pair	A routine stream water sample and a second sample (field duplicate sample) collected from the same stream, by the same sampling crew, during the same visit, and according to the same procedure.
field natural audit sample	See field audit sample.
field synthetic audit sample	See field audit sample.
flag	Qualifier of a data point that did not meet established acceptance criteria or that was unusual. Flags were assigned during the verification and validation procedures.
Gran analysis	A mathematical procedure used to identify the equivalence point or points of the titration of a carbonate system and subsequently for ANC and BNC of that system (Hillman et al, 1987).
holding time	(1) In the processing laboratory, the elapsed time between sample collection and sample preservation. (2) In the analytical laboratory, the elapsed time between sample processing in the processing laboratory and final sample analysis or reanalysis.
imprecision	For a particular analyte, the degree of irreproducibility of or deviation of a measurement from the average of a set of measurements; the variation about the mean.
index value	A mean value for measurements of a performance audit sample made at either the support laboratory (synthetic audit samples) or by a number of analytical laboratories (natural audit samples).
in situ	For this survey, any measurements made within the water column of a stream.
initial dissolved inorganic carbon (DIC)	A measurement of dissolved inorganic carbon made on an aliquot immediately before it is titrated for ANC.

instrument detection limit	For each chemical variable, a value calculated from laboratory calibration blank, reagent blank, or field blank samples that indicates the minimum concentration reliably detectable by the instrument(s) used; calculated as three times the standard deviation of 10 or more nonconsecutive (i.e., from different calibrations) blank sample analyses.
interlaboratory bias	Systematic differences in performance between laboratories estimated from analysis of the same type of samples.
ionic strength	A measure of the interionic effect resulting from the electrical attraction and repulsion between different ions. In very dilute solutions, ions behave independently, and the ionic strength can be recalculated from the measured concentrations of anions and cations present in the solution.
laboratory blank sample	A sample of ASTM Type I reagent-grade water prepared and analyzed by analytical laboratories. (See calibration blank, reagent blank.)
laboratory duplicate sample	Sample aliquot that is split and prepared at the analytical laboratories and that is analyzed once in a batch.
linear dynamic range	The range of analyte concentration for which the calibration curve is best fitted by a straight line.
management team	EPA personnel responsible for overseeing the NSS-I sampling and QA design and the subsequent interpretation of stream data results.
matrix	The physical and chemical composition of a sample being analyzed.
matrix spike	A QC sample, analyzed at an analytical laboratory, that was prepared by adding a known concentration of analyte to a sample. Matrix spike samples can be used to determine possible chemical interferences within a sample that might affect the analytical result.
method-level precision	Precision estimates based on pooled standard deviations and pooled relative standard deviations of processing and analytical laboratory duplicate samples.
nephelometric turbidity unit (NTU)	A measure of light scatter by a solution of suspended materials detected at an angle of 90 degrees to an incident light source.
on-site evaluation	A formal on-site review of field sampling, field laboratory, or analytical laboratory activities to verify that standardized protocols are being followed.
open system	A measurement of pH or dissolved inorganic carbon obtained from a sample collected in a beaker and exposed to the atmosphere during collection, processing, and preparation before measurement.

outlier	Observation not typical of the population from which the sample is drawn.
P_{95}	The 95th percentile of a distribution of blank sample measurements.
percent conductance difference calculation	A QA procedure used to check that the measured specific conductance does not differ significantly (outside the acceptance criteria) from the calculated specific conductance value.
percent ion balance difference calculation	A QA procedure used to check that the sum of the anion equivalents equals the sum of the cation equivalents (see anion-cation balance).
percent relative standard deviation (%RSD)	The standard deviation divided by the mean, multiplied by 100, expressed as percent (sometimes referred to as the coefficient of variation).
pH	The negative logarithm of the hydrogen-ion activity. The pH scale runs from 1 (most acidic) to 14 (most alkaline); the difference of 1 pH unit indicates a tenfold change in hydrogen-ion activity.
pH-ANC	A measurement of pH made in the analytical laboratory immediately before the ANC titration procedure and before the potassium chloride spike has been added.
pH-BNC	A measurement of pH made in the analytical laboratory immediately before the BNC titration procedure and before the potassium chloride spike has been added.
platinum cobalt unit (PCU)	A measure of the color of a water sample defined by a potassium hexachloroplatinate and cobalt chloride standard color series.
population estimate	A statistical estimate of the number of streams (target streams) that have a particular set of chemical characteristics (i.e., alkalinity class within a subregion) extrapolated from the number of streams sampled (probability sample).
precision	A measure of the capability of a method to provide reproducible measurements of a particular analyte.
processing laboratory	The laboratory that processed samples and measured selected variables. The NSS-I processing laboratory was located in Las Vegas, Nevada.
processing laboratory blank	An ASTM Type I reagent-grade water sample prepared and processed at the processing laboratory but analyzed at an analytical laboratory.
processing laboratory duplicate	A split sample prepared and analyzed at the processing laboratory.

protolyte	That portion of a molecule that reacts with either H^+ or OH^- in solution.
protolyte analysis program	An exception-generating computer program of AQUARIUS II that evaluates in situ, processing laboratory, and analytical laboratory measurements of pH, DIC, ANC, BNC, and DOC in light of known characteristics of carbonate equilibria.
QC chart	A graphical plot of test results with respect to time or sequence of measurement, together with limits within which the results are expected to lie when the system is in a state of statistical control (Taylor, 1987).
quality assurance (QA)	Steps taken to ensure that a study is adequately planned and implemented to provide data of the highest quality and that adequate information is provided to determine the quality of the data base resulting from the study.
quality assurance sample	A sample (other than the routine stream sample) that is analyzed in the analytical laboratory and that has an origin and composition unknown to the analyst.
quality control (QC)	Steps taken during sample collection, processing, and analysis to ensure that the data quality meets the minimum standards established by the QA plan.
quality control check sample (QCCS)	A sample of known concentration used to verify continued calibration of an instrument.
quality control sample	Any sample used by the analyst to check immediate instrument calibration or response; the measurement obtained from a quality control sample is expected to fall within specific acceptance criteria or control limits.
raw data set	Data Set 1. The initial data set that received a cursory review to confirm that data are provided in proper format and are complete and legible.
reagent	A substance (because of its chemical reactivity) added to water to determine the concentration of a specific analyte.
reagent blank sample	A laboratory blank sample that contained all the reagents required to prepare a sample for analysis.
representativeness	The degree to which sample data accurately and precisely reflect the characteristics of a population.
required detection limit	For each chemical variable, the highest instrument detection limit based on analyses of laboratory blanks or detection limit check standards allowable in the analytical laboratory contract.

routine sample	The first stream sample collected at a site in accordance with standardized protocols.
sample ID	The numeric identifier given to each stream sample and to each QA sample in each batch.
sampling crew	A team of stream sampling personnel who gained access to the stream site on foot or by vehicle.
SAS	Statistical Analysis System, Inc. A statistical data file manipulation package that has data management, statistical, and graphical analysis abilities. The NSS-I data base was developed and analyzed primarily using SAS software and is distributed in SAS format.
specific conductance	A measure of the electrical conductance (the reciprocal of the electrical resistance) or total ionic strength of a water sample expressed as microsiemens per centimeter at 25 °C.
spike	A known concentration of an analyte introduced into a sample or aliquot.
split sample	A replicate portion or subsample of a total sample obtained in such a manner that it is not believed to differ significantly from other portions of the same sample (Taylor, 1987).
standard deviation	The square root of the variance of a given statistic, calculated by the equation: $$\text{standard deviation} = \sqrt{\Sigma(x - \bar{x})^2/(n - 1)}$$
statistical (significant) difference	A conclusion based on a stated probability that two sets of measurements did not come from the same population of measurements.
stream ID	An identification code, assigned to each stream in the survey, which indicates subregion, alkalinity characteristics, and map coordinates.
synoptic	Relating to or displaying conditions as they exist simultaneously over a broad area.
system decision limit	For each chemical variable except pH, a value that reliably indicates a concentration above background, estimated as either the 95th percentile (P_{95}) or as 1.65 times the standard deviation of the field blank sample measurements.
system-level precision	Cumulative variability associated with sample collection, transport, processing, preservation, shipment, analysis, and data reporting. An estimate of data variability for each analyte associated with analyte concentration; the estimate is based on the statistical evaluation of field routine-duplicate pairs.

systematic error	A consistent difference introduced in the measuring process. Such differences commonly result in biased estimations.
tag	Code on a data point that is added at the time of sample collection or analysis to qualify the datum.
target population	In this survey, the stream population of interest that was sampled. This population was defined by the sampling protocol.
target stream	A stream of interest in the target population.
theoretical value	The expected value of the synthetic audit sample assuming no preparation error and no external effect.
titration data	Individual data points from the Gran analysis of ANC and BNC.
true color	The color of water that has been filtered or centrifuged to remove particles that may impart an apparent color; true color ranges from clear blue to blackish-brown.
turbidity	A measure of light scattering by suspended particles in an unfiltered water sample.
validated data set	Data Set 3. Final product of the validation process in which sample data are examined in the context of a subregional set of samples, rather than at the batch and sample level. Outliers are identified and flagged. Data for field blank, field duplicate, and performance audit samples are included in this data set.
validation	Process by which data are evaluated for quality with reference to the intended data use; includes identification of outliers and evaluation of potential systematic error after data verification.
verification	Process of ascertaining the quality of the data in accordance with the minimum standards established by the QA plan.
verified data set	Data Set 2. Final product of the verification process in which each sample batch and each sample value has been reviewed individually and all questionable values are corrected or identified with an appropriate flag.
within-batch precision	The estimate of precision expected in the analysis of samples in a batch by the same laboratory on any single day. In this report, overall within-batch precision includes the effects of sample collection, processing, and analysis; analytical within-batch precision includes the effects of sample analysis within the analytical laboratories.
withheld sample	An additional duplicate sample collected from a stream by the sampling crew as part of a holding-time experiment. This sample was held in the dark at 4 °C for a 24-hour period prior to processing and preservation.

within-laboratory precision goal	A precision goal based on the data quality objectives for the analysis of laboratory duplicate samples within a single laboratory.

☆U.S. GOVERNMENT PRINTING OFFICE: 1988-548-158/87035

SUBREGIONS OF THE NATIONAL STREAM SURVEY - PHASE I