Quality Assurance in Water Quality Monitoring

Note: The Water Quality Branch, Inland Waters Directorate, became Eco-Health Branch, Ecosystem Sciences and Evaluation Directorate, on April 1, 1992

Water Quality Branch
Inland Waters Directorate
Conservation and Protection
Ottawa, Ontario

628
.161
0971
Gas

© Minister of Supply and Services Canada 1993

TD
226
.G37
1993

Available in Canada through
your local bookseller
or by mail from
Canada Communication Group — Publishing
Ottawa, Canada K1A 0S9
Catalogue No. EN40-448/1993E
ISBN 0-660-15188-9

CANADIAN CATALOGUING IN PUBLICATION DATA

Gaskin, James E. (James Emerson)

Quality assurance in water quality monitoring

Issued also in French under title: Assurance de
la qualité dans le contrôle de la qualité des eaux.
ISBN 0-660-94280-1
DSS cat. no. En40-448/1993E

1. Water quality management -- Canada.
2. Water quality -- Canada -- Measurement.
I. Canada. Ecosystem Sciences and Evaluation
Directorate. II. Title.

TD226.G37 1993 628.161'0971 C93-099473-6

Canada	Groupe
Communication	Communication
Group	Canada
Publishing	Édition

Contents

Page

Foreword v

Part I--QUALITY ASSURANCE MANAGEMENT PLAN

1.0 QA Management Policy I-1
2.0 Documentation I-6
3.0 Responsibility Centres I-7
4.0 QA/QC Operational Strategies I-11
5.0 QA/QC Implementation Considerations I-13
6.0 Addendum I-16

Part II--QUALITY ASSURANCE PRINCIPLES AND GUIDELINES FOR WATER QUALITY SAMPLING

1.0 Introduction II-1
2.0 Designing a Water Quality Monitoring Plan II-5
3.0 Training of Field Personnel II-21
4.0 Field Trip Preparation II-23
5.0 Accessibility to Sampling Sites II-26
6.0 Water Quality Sampling II-28
7.0 Sampling for Specific Parameters in Specific Matrices II-32
8.0 Sequential Triplicate Sampling II-44
9.0 Sample Handling, Preservation, Storage, and Transportation II-47
10.0 Maintenance and Calibration of Instrumentation II-49
11.0 Field Testing II-52
12.0 Data Collection II-59
13.0 Chain of Custody II-63
14.0 Field Safety II-64
15.0 Field Audit Program II-67
16.0 QA/QC Technical and Advisory Committee II-68

Part III--QUALITY ASSURANCE IN THE NATIONAL WATER QUALITY LABORATORY

1.0 Introduction III-1
2.0 Intralaboratory Quality Assurance Plan III-3
3.0 Definitions of Analytical Terminology III-7
4.0 Protocols III-15
5.0 Quality Control Procedures III-19
6.0 Quality Control Practices in the National Water Quality Laboratory III-30

Contents (Continued)

Page

7.0 Guidelines for Good Laboratory Practice III-50
8.0 Guidelines for Instrument Performance Control Checks III-58
9.0 Quality Assessment Forms III-70

Part IV--A MANUAL FOR EFFECTIVE INTERLABORATORY QUALITY ASSURANCE

1.0 The Administration of Quality IV-1
2.0 Prerequisites for an Interlaboratory QA Study IV-9
3.0 Designing an Interlaboratory QA Study IV-11
4.0 Preparation of Containers IV-13
5.0 The Role of Certified Reference Materials and Standards in External QA Programs . . IV-16
6.0 Preparation of Test Samples IV-24
7.0 Distribution of the Interlaboratory QA Study IV-29
8.0 The NWRI AQC Data Base Management System IV-33
9.0 The Federal—Provincial and Prairie Provinces Water Board QA Programs IV-59
10.0 Other Evaluation Techniques IV-69
11.0 The Reporting Process (Feedback) IV-73
12.0 Remedial Action to QA Report IV-75

Part V--QUALITY ASSURANCE GUIDELINES AND PRINCIPLES FOR THE HANDLING AND MANAGEMENT OF WATER QUALITY DATA

1.0 Introduction V-1
2.0 Water Quality Data Collection, Recording, Documentation, and Handling V-8
3.0 Data Screening and Validation V-13
4.0 Data Verification V-44
5.0 Data Quality Assessment V-47
6.0 Statistical Techniques for Water Quality Data Analysis V-63
7.0 Data Presentation and Data Reporting V-72
8.0 Data Custody and Data Transfer V-80
9.0 Management of Water Quality Data Bases V-82
10.0 Training in Data Handling V-88
11.0 Data Audits V-89

PART VI--GENERAL GLOSSARY

Glossary of QA/QC terms for environmental monitoring VI-1

Foreword

The Water Quality Branch has long recognized the importance of having a quality assurance program that governs the three major stages of its water quality monitoring program: sampling, laboratory analytical measurement, and data collection. This quality assurance manual deals with each of these stages in detail.

Part I outlines the Water Quality Branch's quality assurance management plan. Compliance with the policies and guidelines of the management plan contributes significantly to the credibility of the research and monitoring performed by the Water Quality Branch and to the efficient use of available resources.

A fundamental consideration in the reliability of any analytical measurement is that of sample quality. Assurance of acceptable data quality requires a system that ensures that all activities associated with the sampling program are well defined and supported by accepted and standardized practices. Part II, Quality Assurance Principles and Guidelines for Water Quality Sampling, is a manual written to provide guidance for the development of a comprehensive and workable field quality assurance/quality control program within the Water Quality Branch.

Part III of this manual, Quality Assurance in the National Water Quality Laboratory, describes the steps that must be taken to ensure that all measurement and data processing activities that affect the quality of the data produced in a modern laboratory satisfy pre-established standards. It is intended as a reference manual for laboratory staff and data users.

A key element in the quality management strategies of environmental programs is external quality assurance, the principal element of which is the interlaboratory comparison study. Part IV, A Manual for Effective Interlaboratory Quality Assurance, focuses on the interlaboratory study, providing insight on how such studies relate to quality management and how they are designed, prepared, distributed, and interpreted.

The data quality assurance activity has been entrenched as one of the basic components of the monitoring program in support of the quality assurance and quality control activities in the sampling and analytical measurement operations. Part V of this manual is principally intended as a quality assurance manual for the handling and management of water quality data and can be used concurrently with the QA manuals on field sampling (Part II) and intralaboratory/interlaboratory analytical work (Parts III and IV, respectively).

Finally, a general glossary provides definitions for a number of terms used in Quality Assurance activities associated with water quality monitoring.

Part I

QUALITY ASSURANCE MANAGEMENT PLAN

James E. Gaskin

Contents

Page

ACKNOWLEDGMENTS v

EXECUTIVE SUMMARY vi

1.0 QA MANAGEMENT POLICY I-1
- 1.1 Purpose and scope I-1
- 1.2 Significance I-1
- 1.3 Program planning I-1
- 1.4 Managerial accountability I-1
- 1.5 Quality management goals I-3
- 1.6 Performance indicators I-3
- 1.7 Operational control I-4
- 1.8 Operational cost I-5

2.0 DOCUMENTATION I-6

3.0 RESPONSIBILITY CENTRES I-7
- 3.1 Responsibilities of WQB and NWRI I-7
- 3.2 Responsibilities of WQB directors I-7
- 3.3 Responsibilities of WQB chiefs I-8
- 3.4 Responsibilities of section heads and project leaders I-8
- 3.5 Responsibilities of the QA officer I-8
- 3.6 Responsibilities of the QA/QC Technical and Advisory Committee I-9
- 3.7 Responsibilities of other staff I-10

4.0 QA/QC OPERATIONAL STRATEGIES I-11
- 4.1 Strategies for water quality sampling I-11
- 4.2 Strategies for laboratory analyses I-11
- 4.3 Strategies for data management I-12
- 4.4 Recommendation for implementation of strategies I-12

5.0 QA/QC IMPLEMENTATION CONSIDERATIONS I-13
- 5.1 Available resources I-13
- 5.2 Human resource management I-13
- 5.3 Cooperative arrangements I-13
- 5.4 Issues and opportunities I-13
- 5.5 Workload forecast and impact on operations I-14
- 5.6 Reporting against the QA management plan I-14

Contents (Continued)

Page

6.0 ADDENDUM: QUALITY ASSURANCE GUIDELINES FOR THE SELECTION AND MONITORING OF CONTRACT LABORATORIES FOR CHEMICAL ANALYSIS OF ENVIRONMENTAL SAMPLES I-16
6.1 Phase One: Pre-Contract Period I-17
6.2 Phase Two: During the Contract Period I-18
6.3 Phase Three: Post-Contract Period I-19

APPENDIX 1. Quality control clauses for analytical contracts 1-20

Figures

1. Objectives of the Water Quality Branch's QA program I-2

Acknowledgments

Editorial comments made by R. Arseneault, P.E. Brooksbank, A. Germain, R. John, J. Merriman, F. Philbert, and D. Warry are gratefully acknowledged. Additional thanks are owed to Dr. J. Lawrence for composing the Executive Summary.

Executive Summary

Quality assurance (QA) impacts on all phases of research and data-gathering activities. It ensures that the study design is optimum prior to implementation and that studies are implemented as designed. The existence of quality control (QC) guidelines and procedures ensures that the data produced are suitable for the intended use.

The Water Quality Branch's QA management plan assigns formal responsibility and accountability to managers and staff to develop and implement a comprehensive and verifiable QA program for all water quality monitoring programs, including federal–provincial water quality monitoring activities.

The QA management plan requires the preparation of individual QA plans that provide detailed guidelines for the implementation of the aforementioned policies. Compliance with these policies and guidelines contributes significantly to the credibility of the research and monitoring performed and to the efficient use of available resources.

Quality Assurance Management Plan

James E. Gaskin

1.0 QA MANAGEMENT POLICY

1.1 Purpose and Scope

The purpose of quality assurance (QA) planning by managers in the Water Quality Branch (WQB) is to assure the credibility of water quality monitoring activities of the WQB using concepts of quality and improvements in quality measurements. This assures a flexibility in meeting future goals and objectives successfully. The QA management policy encourages excellence and establishes formal responsibility and accountability on the part of WQB managers and staff to develop and implement a comprehensive and verifiable QA program for all water quality monitoring activities.

1.2 Significance

The establishment and adoption of a QA policy by management are necessary for an effective QA program. Success of the QA program will depend heavily on management's commitment to quality outputs as well as the resources that will be made available for conducting the program. Mechanisms must be in place for assuring the formal assignment of responsibilities and tasks among all administrative and operational staff.

1.3 Program Planning

The identification of objectives is the first step in planning and requires careful consideration. Objectives must be defined in such a way that ultimate success or failure can be determined. Sound operational objectives established in the planning process will provide the basis for further breakdown and selection of intended results (see Fig. 1).

To achieve the QA objectives set out for the monitoring program, there must be clear and specific guidelines that are formulated by mutual agreement between concerned parties. In addition, the QA objectives must fit logically and consistently within the WQB's overall hierarchy of objectives.

1.4 Managerial Accountability

The managerial accountability process requires that managers:

- use appropriate information to monitor the ongoing performance of their operations;
- establish and maintain adequate operational control;
- develop and implement means of improving the performance of their operations;

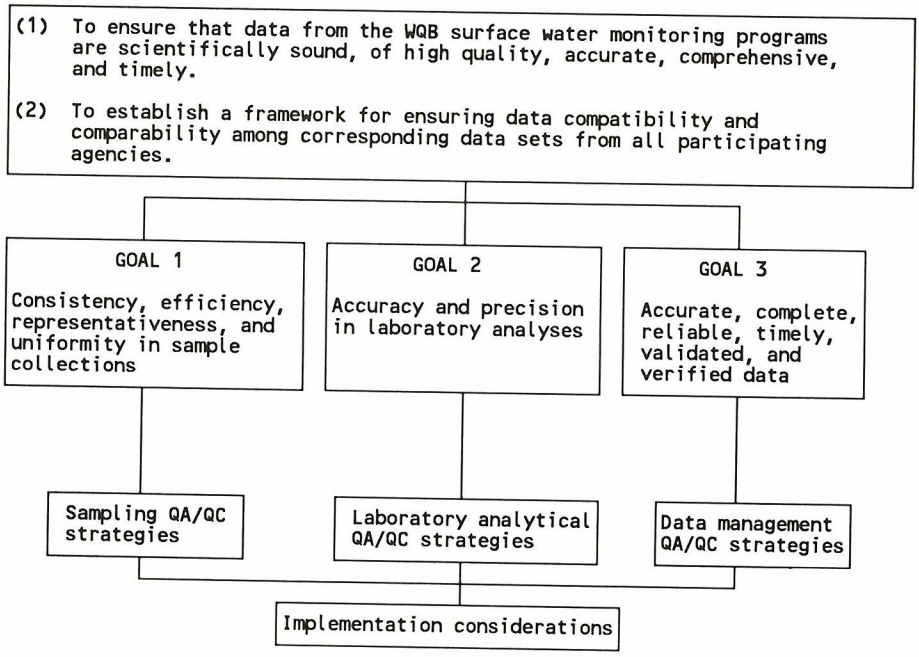


Figure 1. Objectives of the Water Quality Branch's QA program.

- account for the resources that have been used in their operations and for the extent to which goals and objectives have been achieved; and

- ensure that QA management plans are in place for each monitoring program.

1.5 Quality Management Goals

Quality management goals can be summarized as follows:

- formulating work plans and budget proposals that identify projects, work schedules, and resource levels necessary for the effective operation of the WQB monitoring programs;

- ensuring the maintenance of up-to-date manuals in which methodologies and procedures are kept current;

- ensuring that current procedures are used for the collection of representative aquatic samples to yield data and information for water resource managers and the public;

- ensuring a comprehensive quality assurance/quality control (QA/QC) monitoring program that leads to the acquisition of data that have a known quality and meet program needs;

- ensuring that the most appropriate interpretive techniques will be used in examining data;

- ensuring that data validation and verification procedures will be used to examine and certify the quality of data before release to data users;

- establishing an audit program encompassing sampling, laboratory analysis, and data management; and

- encouraging training of operational personnel on a continuous basis to provide consistency and efficiency in all operations.

1.6 Performance Indicators

Performance indicators set by WQB management must be yardsticks by which to assess the validity of the collected water quality data; they must be measurable and consistent over time, so that they can be monitored on a period-to-period basis.

In selecting and developing performance indicators, WQB managers must consider:

- the nature of the WQB monitoring operations;

- the use to which indicators will be put for operational control purposes;

- the relative significance of the three components (sampling, laboratory analysis, and data management) of the water quality monitoring programs;

- the level of participation of other water quality agencies (e.g., provincial and territorial agencies) in monitoring programs;

- the cost and feasibility of systematically gathering data and preparing data and information reports; and

- estimated turnaround times for goals and objectives within the operational activities.

Estimates of data precision, data accuracy, data completeness, data representativeness, and the validated detection limit will be used to judge the reliability of water quality data. Additional determinations of ionic and mass balances will be made, when required, to determine the level of quality and acceptability of the data.

Comparative techniques and methods will also be used to assess the correctness of the data. The greater the direct contribution of performance indicators to the control and improvement of operations at least cost, the more efficient or cost-beneficial these indicators will be.

1.7 Operational Control

Operational control is the process by which managers assure resources are used effectively and efficiently in the accomplishment of WQB objectives. Before operations take place, managers planning their operations and formulating operational controls, must:

- ascertain the objectives of the organization;

- determine the best means of achieving the targeted objectives; and

- determine the resources required for the organization's use (i.e., its planned inputs).

In controlling their operations, WQB managers must try to make the results produced by their operations conform as closely as possible to plans. To achieve this, certain items of key information about their operations must be used to analyze and measure performance. From such analyses, managers must take corrective actions and guide, supervise, and motivate subordinates towards improved performance.

For effective operational control, performance reports should be timely, accurate, forward-looking to detect deviation, relevant to the manager, objective, cost-effective, and understandable.

1.8 Operational Cost

The feasibility of obtaining performance information as simply and as cheaply as possible should be considered at the outset.

The time and effort devoted to the development of a formal QA program should be more than compensated by cost-effective and reliable analytical data.

The cost of the QA program should not exceed 15 to 20% of the cost of the water quality monitoring operation, except in cases where an expenditure of 30 to 40% may be required for gathering reliable data from studies on toxic chemicals.

2.0 DOCUMENTATION

In support of the QA management plan, the following manuals must be prepared and maintained:

(1) *Field Sampling QA Manual.* This document (see Part II) outlines the QA/QC program for field operations to ensure the integrity and representativeness of samples, including sampling design, field testing, sample preservation, transportation, and handling.

(2) *Intralaboratory QA Manual.* This document (see Part III) specifies the internal QC practices and protocols at the technical level to ensure a quality product. Quality assessment criteria should also be included to verify the effectiveness of the QC practices.

(3) *Interlaboratory QA Manual.* This document (see Part IV) describes the interlaboratory activities required to ensure comparability of data generated by different laboratories. The role and applications of natural reference materials are also included.

(4) *Data Management QA Manual.* This document (see Part V) specifies the protocols and guidelines for data collection, data recording, data screening, data validation, data verification, data quality assessment, data presentation, data reporting, data custody, data transfer, management of data bases, training in data handling, techniques for water quality data analysis, and data audits.

3.0 RESPONSIBILITY CENTRES

The QA management plan for the WQB involves interaction between WQB directors, chiefs, section heads, project leaders, laboratory and field managers, QA officers, and the QA/QC Technical and Advisory Committee in establishing and managing the QA program. These individuals are responsible for defining and documenting the duties of all staff relating to QA, QC, and quality assessment and for ensuring that the QA processes are easily understood and implemented.

3.1 Responsibilities of WQB and NWRI

The WQB Headquarters is responsible for coordinating the national QA/QC program for all water quality monitoring activities, including the Federal–Provincial Water Quality Monitoring Agreement.

The National Water Research Institute (NWRI) is responsible for interlaboratory QA/QC programs and analytical reference materials. In addition, the NWRI assists and collaborates with the WQB in:

- planning, designing, and implementing QA/QC programs;
- evaluating field and laboratory data; and
- performing field, laboratory, and data audits.

3.2 Responsibilities of WQB Directors

The QA-related responsibilities of WQB directors are as follows:

- to provide the necessary resources for conducting the required QA programs;
- to ensure that the WQB's QA management plan is implemented by the respective responsibility centres;
- to ensure that WQB protocols and procedures relating to data quality management are exposed to peer review and will meet accepted international standards; and
- to ensure that QA/QC practices and documentation are implemented on a standardized basis on a national scale within the WQB.

3.3 Responsibilities of WQB Chiefs

The QA-related responsibilities of WQB chiefs are as follows:

- to act as a lobby group and connecting link in communicating with WQB directors to convey the essentiality of having adequate resources for conducting the required QA/QC program(s);
- to ensure that there is a QA management plan for each monitoring program;
- to ensure that adequate resources are available for QA programs;
- to ensure that staff are provided with suitable QA training;
- to ensure that the quality of the data produced meets the objectives of the monitoring program; and
- to serve as a review panel for documentation, recommendations, and suggestions originating from the QA/QC Technical and Advisory Committee.

3.4 Responsibilities of Section Heads and Project Leaders

The QA-related responsibilities of section heads and project leaders are as follows:

- to conduct their part of the QA program and to implement appropriate QA procedures;
- to implement QA policies in the most efficient and economical manner, and to establish an understanding of the quality of the data;
- to administer intralaboratory and interlaboratory QA programs with the aim of enhancing the overall capability and performance of monitoring activities; and
- to ensure that staff participate in scheduled system and performance audits.

3.5 Responsibilities of the QA Officer

The responsibilities of the QA officer can be summarized as follows:

- to provide a focal point for WQB QA/QC activities, including revising and updating QA/QC documents to keep abreast of current QA/QC practices;
- to coordinate and participate in field QA/QC exercises and in intralaboratory, interlaboratory, and data handling QA/QC activities;

- to evaluate and assess QA/QC data against predefined criteria (detection limit, precision, accuracy, bias, data completeness, and data representativeness), and to aid in the interpretation of data originating from the QA/QC programs;

- to assist in audit exercises and recommend remedial action to correct problems;

- to liaise with other agencies, particularly water quality agencies, in relation to QA/QC programs, and to participate in the exchange of QA/QC information with other groups (including provincial, federal, and international agencies); and

- to compose annual summaries of QA/QC programs, paying attention to the strengths and weaknesses of the programs, and to suggest improvements, modifications, and possible new approaches to existing programs.

3.6 Responsibilities of the QA/QC Technical and Advisory Committee

The responsibilities of the QA/QC Technical and Advisory Committee are as follows:

- to advise on and recommend measures to ensure the fulfillment of the QA/QC objectives of water quality monitoring programs, including Federal–Provincial Water Quality Monitoring Agreement programs;

- to assist in improving regional QA/QC programs by identifying quality defects and by recommending appropriate corrective actions;

- to ensure that personnel within the WQB are properly trained and able to carry out their QA/QC responsibilities;

- to ensure that documentation of QA/QC guidelines, principles, and protocols is implemented by all participating groups;

- to provide information, advice, and recommendations to operational managers within the WQB concerning all aspects of its QA/QC programs;

- to provide technical assistance and coordination in communicating the WQB QA/QC protocols to provincial and U.S. agencies, and to assist them in implementing compatible methodologies, so that the groundwork for the generation of compatible data can be established;

- to assist in the implementation and monitoring of the performance of QA/QC programs, so that collected data meet established standards of precision, accuracy, and completeness; and

- to organize periodic QA/QC workshops and advanced seminars for WQB personnel, and to identify suitable training courses in interdisciplinary areas of water quality monitoring and water quality technology.

3.7 Responsibilities of Other Staff

All staff are responsible for carrying out the QA/QC tasks assigned to them, consistently adhering to the documented QA/QC procedures and protocols, and recording the results of all QC operations and checks in the required format.

Staff personnel are required to inform supervisors and superiors of operational problems and difficulties that may beset any part of the QA/QC program.

4.0 QA/QC OPERATIONAL STRATEGIES

4.1 Strategies for Water Quality Sampling

(1) Produce a work plan and budget proposals to identify projects, work schedules, and resource levels necessary for the effective operation of the field QA/QC program.

(2) Use current standard and accepted procedures for the collection of representative aquatic samples to yield data and information for water resource managers and the public.

(3) Acquire knowledge about the precision and integrity associated with samples collected from aquatic compartments for the scientific interpretation of data.

(4) Establish the maintenance of minimum sample transit times and proper and appropriate procedures to preserve the integrity of the collected samples prior to analysis.

(5) Provide an effective response capability to advances and changes in water quality sampling technology.

(6) Provide appropriate and enforceable safety procedures for field sampling personnel.

4.2 Strategies for Laboratory Analyses

(1) Produce a work plan and budget proposals to identify projects, work schedules, and resource levels necessary for the effective operation of the laboratory QA/QC program.

(2) Maintain up-to-date laboratory methods and QA/QC manuals.

(3) Achieve the production of reliable data that describe the quality characteristics of the samples submitted by the laboratory's clients.

(4) Conduct analytical measurements as soon as possible.

(5) Provide effective training for laboratory personnel to improve or increase their analytical skills.

(6) Encourage the ongoing process of method adaptation and evaluation to continually test the effectiveness of analytical procedures.

(7) Continue internal and external QA/QC programs to ensure the production of analytical data of known quality and validity, and provide water quality data for interlaboratory comparisons.

4.3 Strategies for Data Management

(1) Provide data bases that are accurate, well documented, and complete.

(2) Ascertain data quality after the data have been recorded, to provide timely feedback to the laboratory analysts so that analytical and recording errors can be corrected.

(3) Ensure that the best interpretive techniques are used in examining the laboratory-generated data with project objectives in focus, and ensure that the data reported are of known and acceptable quality.

(4) Promote and develop uniform approaches to data storage and retrieval.

(5) Establish an evaluation system for ascertaining the quality and reliability of the data produced, ensuring that all data are accompanied by estimates of their precision and accuracy.

4.4 Recommendation for Implementation of Strategies

Each project leader should prepare a project description outlining what has to be done to address and implement the proposed strategies. This record should then become part of the WQB's project plan.

5.0 QA/QC IMPLEMENTATION CONSIDERATIONS

To ensure success of the QA program, there must be an implementation plan that addresses those strategies that have been assembled to provide guidance within the QA/QC activities. The elements of the implementation plan should focus on available resources, human resource management, cooperative arrangements, issues and opportunities, workload forecast and impact on operations, and reporting against the QA management plan.

5.1 Available Resources

Any scheme for judging or formulating the actual resources required for the overall QA/QC program of the WQB must take into account the resource implications for the QA/QC program. Available resources for the QA/QC program will be dependent on budget constraints, salary levels, and the availability of competent operational personnel.

5.2 Human Resource Management

Personnel must have the education, training, experience, and motivation to carry out the requirements of a QA program. In addition to the technical aspects of the QA program, a primary requisite is a good personnel management program. This program will involve selecting, developing, and maintaining a staff capable of economically and effectively meeting the objectives of the three operational activities and serving the needs of those who use their resources.

5.3 Cooperative Arrangements

The WQB will continue to enter into cooperative agreements with the provinces and territories as well as other agencies and groups (local, national, and international water quality agencies) to achieve specific water quality objectives and better management of surface waters in Canada.

5.4 Issues and Opportunities

(1) Advances and changes in information and technology in the chemical, biological, physical, hydrological, and data handling areas of aquatic monitoring and improved QA/QC practices will have an impact on the QA/QC program of the WQB.

(2) Ongoing changes in the emphasis of a number of governmental programs relating to such topics as water pollution, toxic chemicals, and toxic rain will require continuous effective and efficient response capabilities that are geared to sudden or gradual changes in the occurrence or buildup of water pollutants.

(3) The increasing cost of data collection and data handling will require a greater emphasis on more precise and accurate measurements and on the use of the most cost-effective techniques and procedures for collecting data at each stage of the monitoring program.

(4) Revised procedures for field sampling, laboratory analysis, and data management not only should lead to an increase in the productivity of the technical and professional staff, but should also improve the level of consistency and efficiency in the application of the sampling and analytical methods.

5.5 Workload Forecast and Impact on Operations

(1) QA/QC exercises will be performed according to a number of fixed schedules and, in some cases, on an ad hoc basis. In addition, precision and accuracy studies will be required on a regularly scheduled basis.

(2) Methods of analysis will be adapted for use in the laboratory after they are tested by the NWRI or by the National Water Quality Laboratory (NWQL).

(3) Additional statistical techniques will be put in place to provide improved analysis of data sets.

(4) Quality assessments of the QA/QC programs will be made at the end of each fiscal year, and periodic audits will be conducted for each operational activity.

(5) The level of acquisition of new instrumentation, methodologies, procedures, and techniques will partly determine the extent of training that will become necessary.

(6) Increased QA/QC cooperation between regional laboratories and increased emphasis on field sampling and laboratory analytical problem areas may force the WQB to priorize both its internally and externally oriented QA/QC programs.

(7) Closer linkage and cooperation between regional data centres will help to provide more efficient and effective data transfer between regions.

(8) Implementation of specific QA/QC projects should help to achieve the objective of minimizing inaccuracies in sampling and in laboratory analytical measurements and contribute to the production of data of known quality.

5.6 Reporting Against the QA Management Plan

One of the features of a good QA/QC program is an appropriate reporting system. Reporting against the QA management plan has a number of basic purposes, such as:

- documenting the year-to-date difference in actual performance in relation to the planned performance as set out in the plan;

- providing a year-end forecast of the variance in actual operational performance in relation to expected performance; and

- documenting new initiatives.

Reports against the QA management plan should also include such items as:

- reports of achievements of priorities, objectives, and goals;
- new priorities;
- changes in plans; and
- summary reports of overall operational resources, results, and performance indicators.

6.0 ADDENDUM: QUALITY ASSURANCE GUIDELINES FOR THE SELECTION AND MONITORING OF CONTRACT LABORATORIES FOR CHEMICAL ANALYSIS OF ENVIRONMENTAL SAMPLES*

Environment Canada generates a considerable amount of chemical analytical data under contract with private and university laboratories. One of the responsibilities of contract managers must be to ensure that the data collected are reliable and compatible on a regional, national, and international basis. The analysis of environmental samples is more complex than many other chemical analyses due to the low concentrations of contaminants normally encountered. Frequently, the parameters of interest are present in the low parts per billion or parts per trillion range. At these concentrations, many factors that are of little concern for normal analyses become critically important and can severely affect the reliability of the data.

Stringent QA and QC procedures and protocols must become an integral part of all analytical contracts if reliable, traceable, and compatible data are to be generated. These procedures must ensure, and clearly demonstrate, the integrity of samples throughout collection, handling, and analysis. Only by collecting and using data of known quality can we be sure that spatial and temporal trends are statistically significant or that data from different laboratories are comparable. Data of mediocre quality can often be meaningfully interpreted providing the limitations are known, whereas data without associated error limits are open to misinterpretation.

Project managers are responsible for ensuring that contracts contain sufficient QC measures to define the precision and accuracy of the data generated to meet the project objectives. In general terms, intralaboratory QC defines laboratory precision, interlaboratory round robins determine between-laboratory comparability, and the use of certified reference materials (CRMS) determines the degree of accuracy. Without the knowledge of precision and accuracy, the quality of data is difficult to define. Hence, provision for all three activities should be included in all analytical contracts.

The purpose of this short document is to provide project managers with a set of QA protocols for use in preparing, selecting, and managing analytical contracts. The protocols are as applicable to departmental laboratories as they are to those in the private sector. They are written in general terms, and project managers should modify them to meet their specific requirements. Staff of the Quality Assurance and Methods Section, NWRI, are willing to assist project managers with this task should this be necessary. Recommended QA clauses for inclusion in contracts are provided for reference.

---

* Modified from Chau, A.S.Y., H.B. Lee, and J. Lawrence. 1986. Quality Assurance Guidelines for the Selection and Monitoring of Contract Laboratories for Chemical Analysis of Environmental Samples. NWRI Contribution 87-66, National Water Research Institute, Canada Centre for Inland Waters, Burlington, Ont.

6.1 Phase One: Pre-Contract Period

At the outset, the project manager must clearly define the project objectives and the data quality required to meet these objectives. Adequate QA clauses must be included in all contracts (examples are provided in Appendix 1).

Prior to any consideration of a contract, a laboratory must be able to provide the following evidence or information for review:

(1) Clear documentation of analytical and sample handling methodologies (see checklist below), including, wherever applicable, extraction/digestion, cleanup, derivatization, evaporation, and quantitation procedures used. Such methodologies must be available for review at any time.

(2) Statements of method detection limits (MDLs) and laboratory performance (i.e., precision and accuracy) on replicate analyses of samples fortified at or close to MDLs as well as at higher levels.

(3) The availability of relevant laboratory instrumentation and analytical standards.

(4) Records of instrumentation performance, e.g., calibration curves, response factors, detector linearity, resolution, instrument detection limit, etc.

(5) Documented in-house QA protocols and, where appropriate, records of in-house QA data for previous contracts on the same parameters in the same matrices at similar concentration levels.

(6) Performance on previous interlaboratory comparison studies or participation in a qualifying pre-contract QC study.

CHECKLIST FOR SAMPLE HANDLING AND ANALYTICAL METHODOLOGY

(A) *Sample*

(1) Sample handling procedure.
(2) Sample holding time until analysis.
(3) Sample preservation procedure (if needed).
(4) Sample storage conditions.

(B) *Analytical Method*

(1) Method documentation.
(2) Ruggedness*.
(3) Application*.

(4) Specificity*.
(5) Sensitivity*.
(6) Detection limit (definition and statement).
(7) Precision data at min. 2 levels*.
(8) Accuracy data (or recovery) at min. 2 levels*.
(9) Description of how the above method specifications were generated.

* Demonstration at realistic levels.

6.2 Phase Two: During the Contract Period

During the contract period, the project manager must ensure that the following activities are carried out by the contractor:

(1) To demonstrate the precision of data generated, the contract laboratory should perform duplicate analyses of every 10th, 15th, or 20th sample, depending on the situation.

(2) To demonstrate the accuracy of data generated, the contract laboratory should analyze a CRM (if available) or a check sample provided by the scientific authority once every 10, 15, or 20 samples. The agreement with the "true value" must be ± 25% or better, or must meet the defined objectives of the Scientific Authority.

(3) For big contracts, the contractor should provide preliminary results or data sheets to the Scientific Authority at regular intervals (e.g., monthly) rather than just a final report at the end of the contract. If obvious analytical errors are found, the Scientific Authority has the right to reject all or part of a batch of analyses and request re-analysis, in whole or in part, of the batch of samples.

(4) Blind samples in the form of CRMs, reference materials (RMs), field split samples, sample extracts, etc. (incorporated in the sample set by the Scientific Authority) must be analyzed and reported with the data set.

(5) All sample data and chromatograms (or digitally stored data sufficient to regenerate the original chromatograms) should be retained by the contractor for all analyses unless otherwise authorized in writing by the Scientific Authority.

(6) The contractor should participate in relevant interlaboratory QC studies whenever possible during the contract.

(7) The contract laboratory should not change analytical methodologies in the middle of the contract unless authorized in writing by the Scientific Authority.

6.3 Phase Three: Post-Contract Period

(1) The Scientific Authority shall have the right to take possession of all raw data and sample chromatograms that the contractor wishes to discard after the contract.

(2) Report review: the final report must contain all data, including QA data.

APPENDIX 1: QUALITY CONTROL CLAUSES FOR ANALYTICAL CONTRACTS

The following clauses should be included in all analytical contracts issued by departmental project managers:

(1) "A sound quality control program must be developed and documented by the contractor. All quality control data must be made available to the Scientific Authority and the project manager upon request."

(2) "The proposed sample collection, handling, storage, preservation procedures and analytical methodologies must be documented and approved by the Scientific Authority before work is initiated."

(3) "The contractor must participate in, and perform satisfactorily in, pertinent quality control round robins in a timely manner under the guidance of a Departmental Quality Assurance Laboratory designated by the project manager. Failure to comply with this requirement could result in partial withholding of funds and/or cancellation of the contract."

(4) "Contract laboratories are encouraged to participate in appropriate external quality assessment on a continuous basis to establish their credibility."

Part II

QUALITY ASSURANCE PRINCIPLES AND GUIDELINES FOR WATER QUALITY SAMPLING

James E. Gaskin

Contents

Page

PREFACE vii

ACKNOWLEDGMENTS ix

1.0 INTRODUCTION II-1
- 1.1 The need for a field QA/QC program II-1
- 1.2 Objectives of the field QA/QC program II-1
- 1.3 QA/QC documentation II-1
- 1.4 References II-4

2.0 DESIGNING A WATER QUALITY MONITORING PLAN II-5
- 2.1 Sampling program II-5
- 2.2 Site selection II-7
  - 2.2.1 Selection of river sampling stations II-9
  - 2.2.2 Selection of lake sampling stations II-10
  - 2.2.3 Selection of stream sampling stations II-10
  - 2.2.4 General considerations for station location II-11
- 2.3 Site documentation II-11
- 2.4 Establishing the H_0 and H_a hypotheses II-12
- 2.5 Sampling frequency and sample size II-13
- 2.6 Field and laboratory analytical practices II-16
- 2.7 Data quality objectives II-16
- 2.8 The use of statistical analysis II-17
- 2.9 Cost considerations II-18
- 2.10 References II-20

3.0 TRAINING OF FIELD PERSONNEL II-21

4.0 FIELD TRIP PREPARATION II-23
- 4.1 Selection of sampling equipment II-23
- 4.2 Sample bottles II-24
- 4.3 References II-24

5.0 ACCESSIBILITY TO SAMPLING SITES II-26
- 5.1 References II-27

6.0 WATER QUALITY SAMPLING II-28
- 6.1 Sample representativeness II-28
- 6.2 Automatic sampling II-29

Contents (Continued)

Page

6.3 Sampling in winter conditions II-30
6.4 References II-30

7.0 SAMPLING FOR SPECIFIC PARAMETERS IN SPECIFIC MATRICES II-32
7.1 Toxic chemicals in water II-32
7.1.1 Liquid–liquid extraction techniques II-34
7.1.2 Column extraction II-35
7.2 Toxic chemicals in bottom sediment II-35
7.3 Toxic chemicals in biota II-38
7.4 Toxic chemicals in packed snow II-40
7.5 Bacterial sample collection II-41
7.6 Sampling for radioactive parameters II-42
7.7 References II-42

8.0 SEQUENTIAL TRIPLICATE SAMPLING II-44
8.1 References II-45

9.0 SAMPLE HANDLING, PRESERVATION, STORAGE, AND TRANSPORTATION II-47
9.1 References II-48

10.0 MAINTENANCE AND CALIBRATION OF INSTRUMENTATION II-49
10.1 General maintenance and calibration II-49
10.2 Maintenance of specific instrumentation II-50
10.3 References II-51

11.0 FIELD TESTING II-52
11.1 Quantitative analysis II-52
11.1.1 Application and scope II-52
11.1.2 Preparation of standards, blanks, reagents, and sample solutions II-52
11.1.3 Calibration and measurements II-52
11.1.4 Standard addition technique II-53
11.2 Precision and accuracy estimates II-53
11.3 Measurement comparisons II-58
11.4 References II-58

12.0 DATA COLLECTION II-59
12.1 Recording field data II-59

Contents (Continued)

Page

12.2 Support data II-60
12.3 References II-62

13.0 CHAIN OF CUSTODY II-63

14.0 FIELD SAFETY II-64
14.1 References II-65

15.0 FIELD AUDIT PROGRAM II-67
15.1 References II-67

16.0 QA/QC TECHNICAL AND ADVISORY COMMITTEE II-68

APPENDIX 1. Flowchart of QA/QC activities for water quality sampling II-70

Tables

		Page
1.	Total quality cost summary	II-19
2.	Recommended containers and washing procedures for selected water quality parameters	II-25
3.	Procedures used for the preservation of water samples collected for toxic chemical analyses	II-33
4.	Partial results of sequential triplicate sampling of selected stations included in the Nova Scotia LRTAP Lakes Monitoring Program (1984 to 1985)	II-46
5.	Specific conductance (μS/cm) results obtained in the field	II-56
6.	Specific conductance (μS/cm) results obtained in the laboratory	II-56
7.	Precision and accuracy statistics associated with field and laboratory measurements of specific conductance	II-57

Figures

		Page
1.	Related activities in a water quality pilot sampling survey	II-8
2.	Example of standard addition method	II-54
3.	General format for a field sampling sheet	II-61

Preface

A fundamental consideration in the reliability of any analytical measurement is that of sample quality. Sample quality may be appraised on the basis of the quality of the resulting chemical, physical, and biological measurement data.

Assurance of acceptable data quality requires a system that ensures that all activities associated with the sampling program are well defined and supported by accepted and standardized practices.

This manual is written to provide guidance for the development of a comprehensive and workable field quality assurance/quality control (QA/QC) program within the Water Quality Branch (WQB). Chapter 1 serves as an introduction and looks at the need for a field QA/QC program and the objectives of such a program. Chapter 2 discusses different aspects of the design of a water quality monitoring program, beginning at the sampling design stage, progressing through considerations for sample analysis, right through to data quality objectives. Chapter 3 points out the need for and usefulness of a field personnel training program, with suggestions and recommendations on the directions that the program should take. In Chapter 4, the focus is on the field trip preparation, in particular, selection of sampling equipment and sample containers. Chapter 5 deals with the accessibility to sampling sites and is complemented by Chapter 6, which describes a number of general QA features for water quality sampling at the chosen sites. Sampling for specific parameters in specific matrices (e.g., toxic chemicals in water) is given special attention in Chapter 7, and Chapter 8 describes a specific type of QA sampling program, namely, sequential triplicate sampling.

QA approaches to sample handling, preservation, storage, and transportation are outlined in Chapter 9. Chapter 10 lists the maintenance and calibration procedures that must be undertaken before field testing of samples is begun (Chapter 11). Data collection is addressed in Chapter 12, with emphasis being placed on the gathering of support data. Chain of custody protocols that relate to the protection, validity, and integrity of samples and data are outlined in Chapter 13, and safety guidelines are described in Chapter 14. The safety recommendations are based on a number of established health and safety regulations that have been issued by various government departments.

The field audit program is one of the main vehicles used to ensure that the field QA program is working efficiently and effectively, and a number of recommendations are cited in Chapter 15 for establishing and running such a program. One of the groups that assists in the coordination of the audit program is the QA/QC Technical and Advisory Committee. The functions of this committee are spelled out in Chapter 16. Appendix 1 contains a flowchart that depicts the various QA activities that make up the sampling QA program.

Many sections relating to sampling for specific parameters in specific matrices and sequential triplicate sampling are based on the work done by the WQB (Atlantic Region) and have been reported in the Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program Version 2, by R. Arseneault and G. Howell. This work has been quoted freely.

Acknowledgments

We acknowledge the valuable contribution made by R. Arseneault and G. Howell to a number of sections in this manual. The effort made by E. Watt in an earlier version of a field sampling QA manual also contributed significantly to various portions of the text.

We are also appreciative of the constructive comments and suggestions made in the review process by the following Water Quality Branch personnel: H. Agemian, R. Arseneault, P. Brooksbank, L. Désilets, R. John, F. Philbert, C. Thorpe, H. Vaughan, and D. Warry. The patience and efforts demonstrated by Mrs. Francine Goulet in the typing and retyping of the manuscript are also greatly appreciated.

Quality Assurance Principles and Guidelines for Water Quality Sampling

James E. Gaskin

1.0 INTRODUCTION

1.1 The Need for a Field QA/QC Program

A quality assurance/quality control (QA/QC) program will ensure the integrity and representativeness of the samples taken from Canadian inland waters. If nonrepresentative samples are collected, and if sampling and preservation techniques are inadequate, the quantitative results obtained will be invalid. In the past, QA/QC principles and protocols were applied mainly to laboratory analysis, and very little QA/QC activity was devoted to the sampling program. Emphasis is now being placed not only on the collection of samples that are appropriately representative of the aquatic source from which they are taken, but also on the appropriate preservation, transportation, and handling procedures that are used on each sample in its journey between the collection site and the laboratory.

Collection practices in the sampling and preservation activities need checks and balances to foster confidence in the resulting data. These checks and balances are the QA/QC protocols used to create the climate of data integrity and data acceptability.

1.2 Objectives of the Field QA/QC Program

One of the primary objectives of the field QA/QC program is concerned with the ability of the Water Quality Branch (WQB) to collect and deliver appropriately representative water quality samples to the laboratory. A second, equally important objective is to produce data of known quality, sufficient to meet the needs of the network and its data users. There are six major characteristics of data quality: accuracy, precision, completeness, representativeness, data comparability, and data compatibility. Under ideal conditions, it is desirable to specify acceptable levels of these characteristics at the outset of the monitoring program. The QA/QC program should then be designed to ensure that these levels are met.

1.3 QA/QC Documentation

For a field sampling QA/QC program, there are three documents that are required (Stanley and Verner 1985). The first is the QA program plan, which commits management to a QA policy and sets forth the requirements for data needed to support program objectives. The program plan describes the

overall policies, organization, objectives, and functional responsibilities for achieving data quality goals. The five major functions of the plan are as follows:

- a statement of the purpose and importance of a QA plan;
- a description of the procedures that will be used to carry out the QA program;
- a description of the resources committed to perform the QA work;
- an identification of projects that require QA project plans; and
- a description of how QA implementation will be evaluated.

The second document is the QA project plan, a technical document that specifies the QA and QC requirements for each project. The plan specifies any QA/QC activities required to achieve the data quality goals of the project. It describes how all data are assessed for precision, accuracy, representativeness, completeness, comparability, and compatibility. The QA project plan further requires that all data generated be thoroughly documented, and it must address the following items in sufficient detail to permit unambiguous evaluation of project results:

- project description;
- project organization and designated responsibilities;
- QA objectives for the experimental data in terms of precision, accuracy, completeness, ruggedness, and comparability;
- sampling procedures and sample handling;
- sample custody, transportation, preservation, and storage;
- calibration procedures and frequency of calibration;
- experimental design and analytical procedures;
- reference standards and QC standards;
- documentation;
- data reduction, validation, verification, and reporting;
- internal QC checks and frequency;

- preventive maintenance procedures and schedules;
- specific routine procedures to be used to assess data quality;
- corrective action; and
- QA reports to management.

To satisfy the requirements for quality data, the project plan must describe the following activities:

- network design;
- selection of specific sampling sites;
- sampling, analytical methodology, calibration, and standard operating procedures (SOPs);
- sampling devices, storage containers, and preservatives;
- special operating conditions (e.g., heat, light, reactivity);
- reference, equivalent, or alternative test procedures;
- instrument selection and use;
- preventive and remedial maintenance;
- replicate sampling;
- replicate analyses;
- blank and spiked samples;
- intra- and interlaboratory QC procedures;
- documentation; and
- sample custody.

The field sampling QA/QC manual is also a component of the QA project plan and must provide guidance on policy and procedures. This manual will contribute to the quality of the data generated by:

- providing unified information for all participating agencies;
- detailing procedures to be used in the field;
- providing information on project descriptions, project organization, and designated responsibility;
- considering siting criteria for the sampling plan;
- indicating the QA objectives for precision, representativeness, completeness, and comparability;
- providing information for calibrating and maintaining equipment;
- providing information on safety practices in sampling and field testing operations;
- providing accepted procedures designed to control and define errors associated with field measurements;
- defining statistical techniques for assessing the experimental data; and
- ensuring that the collected data have met the measurement program objectives.

The third document that is needed is a program implementation plan. A number of mechanisms must be put in place to ensure maximum coordination and integration of QA efforts within the overall program (covering sampling, laboratory analysis, and data handling). The management committee directing the cooperative water quality arrangements will coordinate the monitoring and survey programs to ensure that data of known and acceptable quality are produced. Resource levels, schedules, turnaround times, responsibility centres, performance indicators, milestones, risk factors, implications, emerging issues, etc. must all be given attention if the programs are to succeed.

1.4 References

Stanley, T.W., and S. Verner. 1985. The U.S. Environmental Protection Agency's Quality Assurance Program. Am. Soc. Test. Mater. Spec. Tech. Publ. 867.

2.0 DESIGNING A WATER QUALITY MONITORING PLAN

2.1 Sampling Program

The field QA/QC strategy includes a sampling plan that serves as the central element for directing the various operational activities. To produce a truly effective sampling plan, an interdisciplinary approach is necessary. The interdisciplinary contributions assure the development of a plan that is specific and has all the requirements for successful sampling operations.

The sampling plan must concern itself with the "why," "what," "where," "when," and "how" questions. The "why" question must address the broad objective(s) of the monitoring program and the data quality objectives associated with the various specified projects. The rationale for each project should include an estimation of the project's utility and cost-effectiveness. The "why" question can be influenced by such considerations as:

- prevailing environmental concerns and problems;
- a combination of socioeconomic, sociopolitical, and environmental issues;
- the need for continuing or expanding existing projects or programs; and
- the need to capture baseline data at new specific sites.

The answer to the "what to sample" question should be a logical fallout from the answer to the "why to sample" question. Specific water quality parameters may be issue-oriented or may result from water quality objectives arising from negotiated water quality monitoring agreements between federal and provincial agencies.

The choice of "where to sample" may be controlled by the decisions taken on "why to sample" and "what to sample." However, once a river basin, lake, or stream (macrolevel choice) has been agreed upon, selecting sampling stations (microlevel choices) is an exercise controlled by a number of exogenous and endogenous factors, including:

- hydrologic data (e.g., flow rates, sediment discharges) for various sections of the aquatic body;
- topography of the macrosite;
- geographical location with regards to interjurisdictional boundaries;
- ease of access to sampling sites;
- runoffs and material infusions (e.g., sediment, water, sludge, sewage) from external sources;

- mixing currents;
- water depth; and
- anthropogenic activity.

The "when to sample" choice is influenced by, for example:

- objectives of the study;
- available resources;
- seasonality, including spring runoff, heavy snowfall, and heavy rainfall;
- required estimated sampling frequency;
- high or low flow rates within the sampling cross section of the aquatic body;
- unusual environmental and ecological occurrences, such as flooding and chemical spills;
- intentional or unintentional toxic waste discharges;
- anthropogenic activities that may be deleterious to the aquatic body; and
- the density of airborne and non-point source pollutants.

The appropriateness of "how to sample" can be addressed by consideration of nine principles (Green 1979):

(1) Preliminary sampling, such as a pilot survey, should be conducted in order to provide a basis for an evaluation of the sampling design and the statistical options.

(2) Replicate samples should be taken within each combination of time, location, and any other controlled variable. Differences "among" can be demonstrated only by comparison with differences "within."

(3) An equal number of randomly allocated replicate samples should be collected for each combination of controlled variables. Collecting samples from "representative" or "typical" places is not random sampling.

(4) It should be verified that the sampling method is correct for the range of sampling conditions to be encountered. Note that variation in efficiency of sampling can bias "among-area" comparisons.

(5) If the area sampled has a large-scale environmental pattern, it should be broken up into relatively homogeneous sub-areas, and sample collections should be proportioned according to the size of each sub-area.

(6) The sample size should be verified as appropriate to the quantity and spatial distribution of the parameter. An estimate should then be made of the number of replicate samples required to obtain the necessary precision.

(7) To test whether a condition has an effect, samples should be collected both where the condition is present and where it is absent, but all else is the same. An effect can be demonstrated only by comparison with a control.

(8) Ongoing site evaluations for parameters such as pH, conductivity, turbidity, and dissolved oxygen should be incorporated into the sampling plan, with some consideration given to the possible need for more extensive studies at sites where specific problems have been identified.

(9) The data from the pilot survey(s) should be tested and analyzed in order to determine whether the error variation is homogeneous, normally distributed, and independent of the mean. If this is not the case (a situation that is highly probable with biological parameters), then the data could be appropriately transformed, or distribution-free (nonparametric) procedures could be used.

Figure 1 illustrates the interconnection between a number of water quality sampling activities in a pilot survey. See Appendix 1 for a complete flowchart of activities and subactivities connected with the sampling operation.

2.2 Site Selection

Site selection must be given detailed consideration during the development of the sampling program plan, because sampling techniques to be used in a given situation will depend not only on the data needs, but also on the nature of the flow and other conditions.

Appropriate siting is critical to the success of the monitoring network. Poor siting could result in low-quality, unrepresentative data. It is therefore important that comprehensive QA/QC procedures be developed to address network siting. These procedures should be designed to meet a specific set of QA/QC objectives, namely:

- to assure that the geographical distribution of sampling locations meets network objectives;
- to ensure that the sampling sites (microlocations) chosen are appropriately representative of the existing aquatic characteristics of the given water body; and

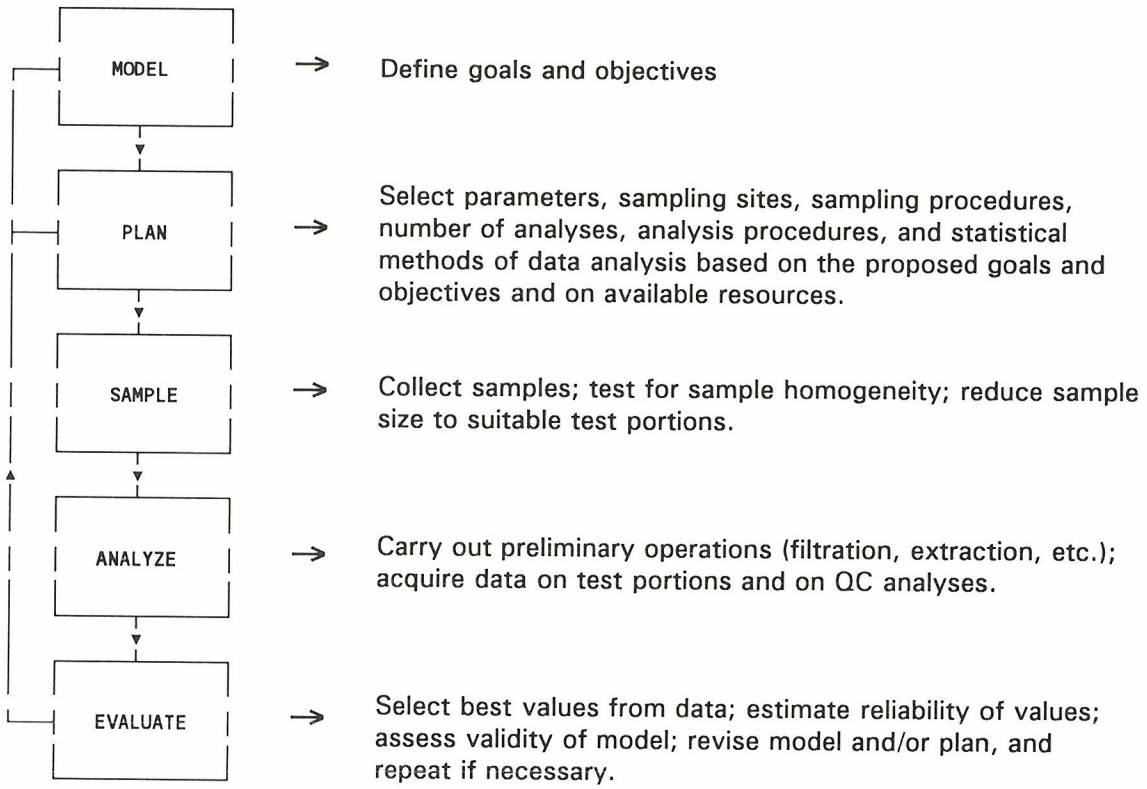


Figure 1. Related activities in a water quality pilot sampling survey.

- to ensure that site documentation is sufficient for data users to assess the suitability of the sites for the required purposes.

Meeting these three QA/QC objectives requires the implementation of the following procedures:

- conducting a preliminary site assessment, followed by a report prepared within the first six months of operations, so that obvious problems are detected early and the site can be changed before a large data set is obtained;

- routine reporting of significant changes in data obtained from the site during period-to-period sampling;

- proper documentation of sampling sites during program operations; and

- independent QA audits of sampling sites. These should be conducted in order to independently assess the accuracy and completeness of current site documentation and the representativeness of each site with respect to its monitoring objectives. Site representativeness could be checked by looking at data for a number of physical parameters (e.g., turbidity, pH, conductivity, dissolved oxygen).

Apart from using objectives to allocate sampling stations, other factors such as basin size and available resources must enter the selection scheme so that a reasonable degree of optimization can be achieved. Resources may be allocated on the basis of basin size, sampling frequency, number of sampling sites within a macrolocation, site accessibility, cost-shared arrangements between monitoring agencies, and distance of site(s) from the office that is coordinating or conducting the sampling operations.

2.2.1 Selection of River Sampling Stations

Three approaches have been defined for selecting the sampling macrolocation within a river basin (Désilets 1986):

(1) The global approach is designed to assess a river basin's hydrochemical balance.

(2) The sub-basin approach is aimed at assessing water quality impairments induced by diffuse sources such as agricultural lands and forestry areas.

(3) The river stretch approach is structured towards assessing the importance of one or various point sources of pollutants and their impact on downstream aquatic environmental quality. Point sources to consider include major cities (wastewaters), hydroelectric dams, industrial sites, landfills, construction sites, commercial fishing areas, and recreational areas.

2.2.2 Selection of Lake Sampling Stations

It has been observed that lake water sampling often shows less temporal variance (but greater spatial variance) than river sampling. This observation favours the use of lakes for long-term trend assessments, as the monitoring costs are potentially reduced. The following criteria have been suggested (Désilets 1986) to assist in the selection of headwater lakes for inclusion in the monitoring network:

(1) Headwater lakes measure the effects of atmospheric deposition and therefore should be located on the most elevated sites of the basin, away from agricultural lands and urban areas, to avoid local climatic effects.

(2) Selected lakes should be easily accessible.

(3) Lake depth should be at least 10 m for stable thermal conditions.

(4) Dystrophic or bog lakes should be avoided.

(5) Lakes that are fed by large inlets should be avoided because of the possible dominance of stream characteristics.

(6) As a general rule, water and sediment sampling stations should be located near the centre of the lake, at the greatest depth, to avoid shoreline effects.

2.2.3 Selection of Stream Sampling Stations

Many factors are involved in the proper selection of sampling stations for streams (Kittrell 1969), including:

- objectives of the stream study;
- water uses;
- access to desirable sampling points;
- entrance and mixing of wastes and tributaries;
- flow velocities and times of water travel;
- marked changes in characteristics of the stream channel;
- types of stream bed, depth, and turbulence;

- artificial and physical structures such as dams, weirs, and wingwalls; and

- resources available for the study.

The ideal sampling station would be a cross section of a stream at which samples from all points on the cross section would yield the same values (e.g., concentrations) for respective constituents, and sequentially taken samples at a given point in time would yield the same values. The former situation should occur when vertical mixing and lateral mixing of any upstream wastes or tributaries are complete at the sampling station. The latter situation should occur only if there is no variation in upstream waste discharges or there is complete mixing longitudinally of any variable waste discharge, and if there are no upstream variations in stream flow, time of water travel, temperature, biological activity, or other factors that contribute to variation in water quality.

Variations in water quality with time require that samples be collected at the proper frequencies and times of day to ensure results that are representative of the variations.

2.2.4 General Considerations for Station Location

As a general rule, past or present monitored sites should be taken into account when locating a water quality sampling station, to ensure a long-term continuity in data set or to identify seasonal cycles. Water quality monitoring stations should be located close to water quantity stations when these already exist, in order to facilitate the study of water quality–quantity relationships. The distance between the collection site and a gauge recorder should not be more than a few hundred feet, unless it has been ascertained that there is no appreciable inflow between sites. If there is no gauging station at or near the sampling site, a discharge measurement should be made at the time of sampling (Kister and Garrett 1983).

When locating a sampling station within a given jurisdiction (province, territory, international border), cooperation between the respective water quality agencies is essential.

2.3 Site Documentation

Documentation of each site should include the following:

- name of lake, river, or stream, county and township, province or territory;

- boundaries of segment sampled, basin, and sub-basin (using NAQUADAT codes);

- Universal Transverse Mercator (UTM) or latitude–longitude coordinates of site;

- elevation of site above sea level;

- area of watershed;

- morphometric measurements for lakes (area, volume, mean depth, and maximum depth);
- topographic map showing site location;
- for streams and rivers, location of nearest flow recorder; also stream classification;
- exact location of sampling point (e.g., distance from bank, whether midstream, location in lake);
- distance of sampling site from point source discharges;
- a general description of the area, including land use practices upstream of sampling locations and ease of access;
- water quality objectives;
- availability of sediment surveys;
- data from other agencies (federal, provincial, territorial, interjurisdictional, international, and private);
- hydrologic data; and
- major issues and concerns.

2.4 Establishing the H_0 and H_a Hypotheses

Having decided on the "why," "what," "where," "when," and "how" to sample, the next step is to formulate a hypothesis that is simple, rational, and consistent with the information gathered from a pilot study or survey regarding the aquatic system that is to be monitored.

All models for testing statistical hypotheses provide a set of decision rules that specify the conditions under which the hypothesis will be rejected or accepted. Often the statistical hypothesis is stated in null form (e.g., "there is no association between variables X and Y"). Hypotheses such as this are referred to as null hypotheses (H_0).

The formulation of hypotheses establishes a bridge between the objectives of the water quality project and the studied phenomenon. It can also help to reduce the scope of the project and limit the collection of information.

In establishing the null hypothesis (H_0), the alternative hypothesis (H_a) must also be considered, and the assembly of both hypotheses is a prerequisite to the later application of statistical tests to the experimental data.

The sampling design should reflect the H_0 model as closely as possible. If, for example, H_0, given as "An impact (resulting from the presence of increased total cyanide concentration) has not led to fish kill effects," contrasts with an H_a that it has, then the sampling study should be designed so that as many as possible of the environmental variables unrelated to impact are controlled; perhaps this can be achieved by choosing control and impact areas that are similar except for the fact that one will be impacted and the other will not.

To assist in the acceptance or rejection of the null hypothesis at a given significance level, careful thought must also be given to the choice of predictor variables (e.g., spatial and temporal patterns) and criterion variables (e.g., sample representativeness, precision measurements, data completeness).

2.5 Sampling Frequency and Sample Size

Most water quality sampling for the WQB is conducted monthly, but the day of the month and the time of day may vary randomly. In many cases, sampling frequency is related to the estimated variance in the measurements of the water quality constituents of concern.

The simplest design case on a single station–single variable basis would be to select the sampling frequency that results in the desired confidence interval width about the annual mean (or geometric mean) for the specified water quality variable at the specified site (Sanders *et al.* 1983). If the estimate of the variance (obtained from historical data or a pilot program) is assumed to be the population variance, σ^2, then the number of samples N (per period) required to obtain a given degree of confidence in estimating a mean value for the parameter of interest can be derived from the following equation:

$$N \geq \frac{Z^2_{\alpha/2} \times \sigma^2}{E^2} \qquad (1)$$

where: $Z_{\alpha/2}$ is the standard normal deviate corresponding to a probability of $\alpha/2$; and
E is the error term, given by μ-x.

If S is used instead of σ, $t_{\alpha/2}$ (Student's t) is used in place of $Z_{\alpha/2}$, and N is given as:

$$N \geq \frac{t^2_{\alpha/2} \times S^2}{E^2} \qquad (2)$$

Sanders *et al.* (1983) provided a very comprehensive outline on methods for determining sampling frequencies for single station–single variable, single station–multiple variables, and multiple station–multiple variables.

The following is an example of the use of Equation 1.

* * *

EXAMPLE:

The samples to be measured for a given parameter are expected to have a mean annual concentration of approximately 0.1 ppm, with standard deviation of 0.05 ppm, and the tolerable error in the stated value of the mean at the 95% confidence level (Z = 1.96) is not to exceed 20% of the mean (0.02 ppm). If it can be assumed that the measurement error is small in comparison with the measured values, find the number of samples required per year to estimate future annual means for the given parameter.

With the above values, the number of samples required will be:

$$N = \frac{(1.96)^2 \times (0.05)^2}{(0.02)^2}$$

= 24 samples per year, or 2 samples per month

* * *

The appropriate size of samples to be collected for any given parameter will depend on the estimated natural concentration level of that parameter within the aquatic source. The smaller the existing analyte concentration, the larger the volume necessary for obtaining a reasonable measurement (a measurement equal to or greater than the limit of quantitation, subsequent to volume reduction).

For organics, samples of 100 L or more are collected for solvent extraction. New field in situ solvent extraction techniques are now providing alternatives to handling large liquid volumes. However, for a number of diverse parameters, there is still a need for estimating the sample volumes required to measure levels of interest. A number of empirical and semi-empirical methods have been developed for determining sample sizes so as not to exceed a predetermined level of sample uncertainty. One approach (Ingamells and Switzer 1973) is based on the knowledge that the between-sample standard deviation decreases as the sample size increases.

A modified form of the Ingamells' Equation (see Equation 4) can be adopted to determine the sample volume that is required for an estimated standard deviation and a predicted sampling uncertainty. The equation is of the form:

$$VR^2=M_s \qquad (3)$$

where: V is the volume of sample analyzed;
R is the relative standard deviation (in percent) of sample analyte; and
M_s is the sampling constant, corresponding to the volume of sample required to limit the sampling uncertainty to a chosen percentage value (e.g., 1%) with a proposed level of confidence (e.g., 68%).

The magnitude of M_s can be determined by estimating the between-sample standard deviation S_s from a series of measurements of samples of volume V.

Once M_s is evaluated for a given sample, the minimum volume V required for a maximum relative standard deviation of R% can be calculated. An example of the use of M_s follows.

* * *

EXAMPLE:

Estimate the volume of sample to be collected for the analytical measurement of a given parameter, when the values of M_s and R are given as 100 mL and 20%, respectively.

Inserting the values of M_s and R in Equation 3, V is calculated as:

$$V = \frac{100\,mL}{(0.2)^2} = 2.5\,L$$

* * *

When a reduction of the original volume (via solvent extraction, distillation, evaporation, etc.) is required before analysis, back-calculations must be carried out to determine the original volume that must be collected to give the desired measurable level of the parameter. For example, if V_1 is the volume calculated from the use of M_s and R, then the required original volume V_0 is obtained from the product of V_1 and the volume reduction factor. For a volume reduction of 100, V_0 would be $100V_1$.

The Ingamells' Equation:

$$WR^2 = K_s \qquad (4)$$

can be applied to measurements involving bottom sediment samples. W represents the weight of the sample analyzed, R has the same meaning as in Equation 3, and K_s replaces M_s in the same equation.

Whether volumes or weights are being considered, every effort should be made to assure homogeneity in the sample matrix before any subsampling is carried out.

2.6 Field and Laboratory Analytical Practices

The sampling plan must consider the role of good field and laboratory analytical practices. The basic QA/QC field testing and laboratory analytical protocols that will be adopted must be spelled out in the sampling plan. For the laboratory, the scope of both the intra- and interlaboratory QA/QC programs should be referenced and should be consistent with the size of available resources.

The sensitivity of each analytical method is an important characteristic and must be known. The precision and accuracy of measurements must reflect the level of confidence placed on the measurements.

The accuracy required from the measuring system should be determined by the objectives of the sampling program and by the use of the data. The measurement system should avoid superfluous and expensive accuracy and precision determinations if there is a failure in the monitoring program to meet the stipulated objectives. The analytical measurements must also be rugged and complete.

2.7 Data Quality Objectives

Data generated from the sampling program must possess a number of quality factors if they are to be used objectively in judging the water quality and if correct and unbiased management decisions are to be made.

All data used in examining and diagnosing water quality should be accompanied by the following characteristics:

- accuracy;
- precision;
- completeness;
- method detection limit (MDL);

- representativeness; and

- traceability

and should be suitable to meet network objectives. Accuracy, precision, completeness, and the MDL are established through accepted statistical principles and are distinguishable from other quality characteristics by their quantifiability (see Part V for a discussion of data handling principles and guidelines).

Another aspect of data quality relates to data comparability and compatibility among similar data sets from agencies involved in water quality monitoring. To achieve data comparability and compatibility, a number of steps must be taken, including:

- employing techniques and practices that can be duplicated at different locations and times and by different agencies;

- providing results that can be understood and tested on a comparative basis;

- reducing systematic errors and increasing comparability between measurements by additional interlaboratory sample testing exercises; and

- using certified reference materials (CRMs) or carefully selected natural areas as common reference sites, especially when there is a need to compare the results from different existing measurements.

The achievement of data quality objectives can often be difficult. Limitations of resources, methodologies, equipment, and technical expertise may reduce the quality and amount of data collected. However, any consideration of the limitations associated with the data collection process should lead to the establishment of objectives that adequately satisfy the goals of the project.

2.8 The Use of Statistical Analysis

One of the basic aims of the WQB monitoring program is to provide representative water quality data that can be understood and tested on a statistical or comparative basis. The most appropriate statistical techniques must be employed to produce comprehensive analyses and interpretations (see Part V).

The choice of appropriate statistical analyses should flow logically from the objective(s) of the sampling or monitoring study, the null hypothesis, and the sampling plan. The hypothesis model should determine the statistical model.

An efficient statistical analysis method should be as conservative, powerful, and robust as possible (Green 1979). If it is conservative, it will have a low probability α of making a type I error.

If the statistical method is powerful, it will have a low probability β of making a type II error. If the method is robust, the stipulated error levels will not be seriously affected by the kinds of data commonly encountered in environmental studies.

Overcomplicated statistical approaches should be avoided as much as possible, and results of the statistical analysis should always be reported in a form that is understandable by the data users.

2.9 Cost Considerations

Quality costs for environmental monitoring systems have been treated by Rhodes and Hochheiser (1971) and by Strong *et al.* (1980). Table 1 is an example of how quality costs can be itemized by category, group, and activity.

Program managers involved with water quality monitoring programs must be concerned with program cost-effectiveness. The operation of a quality cost system would help to minimize the cost of operational activities that are directed towards controlling data quality. The basic concept of the quality cost system is to minimize total quality costs through proper allocation of planned expenditures for the QA/QC efforts, while maintaining an acceptable level of data quality.

The quality cost system structure provides a means of identifying quality-related activities and of organizing these activities into prevention, appraisal, and failure cost categories, which are defined as follows:

(1) Prevention costs are associated with planned activities whose purpose is to ensure the collection of data of acceptable quality and to prevent the generation of data of unacceptable quality.

(2) Appraisal costs are associated with measurement and evaluation of data quality. These include the measurement and evaluation of materials, equipment, and processes used to obtain quality data.

(3) Failure costs are incurred directly by the monitoring agency or organization producing the failure (unacceptable data).

There is no preset formula for determining the optimum mode of operation. Rather, the cost-effectiveness of quality costs is optimized through an iterative process requiring a continuing analysis and evaluation effort. Maximum benefits are realized when the system is applied to a specific measurement method in a stable, long-term monitoring program. For example, a monitoring program with a fixed number of monitoring sites, scheduled to operate for more than one year, would be a desirable candidate for a quality cost system.

The size and diversity of a water quality monitoring operation are important determining factors in decisions related to formalization of the QA program. Operations concerned with defined programs

Table 1. Total Quality Cost Summary

Cost group	1st quarter ($)	2nd quarter ($)	3rd quarter ($)	4th quarter ($)
Prevention				
Planning and documentation				
Procurement				
Training				
Preventive maintenance				
System calibration and operation				
Total prevention costs				
Appraisal				
QC measures				
Audits				
Data validation				
QA assessment and reporting				
Total appraisal costs				
Failure				
Problem investigation				
Corrective action				
Lost data (unacquired data)				
Total failure costs				
Total quality costs				

of work may be more easily adaptable to formal QA activities, and the program costs can be easily justified. For operations involving numerous repetitive measurements, as in the case of WQB monitoring activities, a formal QA program is a necessity, and the time and effort devoted to its development should be compensated by cost-effective, reliable analytical data. The overall cost of the activities should not exceed 10 to 15% of the total operational effort, and "repeats" and lost data should be minimized.

2.10 References

Désilets, L. 1986. Water Quality Branch Strategy for Assessments of Aquatic Environmental Quality, and Criteria for Basin Selection and Sampling Station Macrolocation. Water Quality Branch, Inland Waters Directorate, Environment Canada, Ottawa, Ont.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley-Interscience Publication, John Wiley & Sons, Toronto, Ont.

Ingamells, C.O., and P. Switzer. 1973. Talanta 20: 547.

Kister, L.R., and W.B. Garrett. 1983. Field Guidelines for Collection, Treatment and Analysis of Water Samples. Arizona District, Ariz.

Kittrell, F.W. 1969. A Practical Guide to Water Quality Studies of Streams. Federal Water Pollution Control Administration, CWR-5, U.S. Department of the Interior, Washington, D.C.

Rhodes, R.C., and S. Hochheiser. 1971. Quality costs for environmental monitoring systems. Am. Soc. Qual. Control Tech. Conf. 77: 151.

Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian, and V. Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, Colo.

Strong, R.B., J.H. White, and F. Smith. 1980. Guidelines for the Development and Implementation of a Quality Cost System for Air Pollution Measurement Programs. EPA Contract No. 68-02-2722, Research Triangle Institute, Research Triangle Park, N.C.

3.0 TRAINING OF FIELD PERSONNEL

All personnel involved in sample collection, field analysis, data collection, and QC procedures must be adequately trained. They must be made aware of the importance of obtaining data of known quality, must become familiar with the QA program, and must be knowledgeable of the various ways of achieving and maintaining good-quality data.

Training personnel for field operations will involve a broad range of disciplines, because of the diversified requirements (e.g., mechanical, physical, chemical, biological, hydrological).

Personnel training must start before operations begin and must continue during operations, with periodic reviews and retraining when indicated.

In order to maintain and foster consistency in the quality of collected data, all field personnel (particularly new employees) should become familiar with project objectives, sampling protocols, station location, equipment operation, and safety rules prior to the commencement of fieldwork.

Lay sampling personnel (i.e., persons engaged in water quality sampling for the WQB that are not WQB employees) should be strictly supervised until it is fully demonstrated that they can conduct the required operations in the desired manner. In cases where provincial staff are involved in water quality sampling activities (e.g., Federal–Provincial Water Quality Monitoring Agreement programs), proper training should be given by WQB personnel. Training may take the form of a course on water quality sampling with hands-on participation. Station microlocations and initial sampling runs should be conducted as a collaborative effort between personnel of the two agencies, in order to ensure compatibility between provincial and federal procedures.

Suggestions for general on-the-job training for sampling and field testing personnel include:

- studying operation manuals and the sampling program plan;
- observing a trained and experienced operator perform the various tasks in the data collection process;
- performing field operations under the direct supervision of a trained and experienced person and then independently performing operations related to instruments, sample collection, sample analysis, and data handling;
- attending and participating in in-house seminars that are directed towards problem solving and confidence building;
- gaining experience and acquiring new skills and technical information through interchange programs between the various regions, in cases where the interchange becomes a beneficial exercise; and

- announced and unannounced periodic proficiency testing of operational personnel to determine the need for additional training.

4.0 FIELD TRIP PREPARATION

Before starting out on a water quality data collection trip, field personnel should determine the types of samples that are to be collected and the field measurements required in order to assemble the proper sampling and testing equipment. Field personnel should also determine that the equipment being transported to the field is in good working order and that there are spares available for those pieces of equipment that are hard to replace once out in the field. The present WQB field manual, Sampling for Water Quality (Environment Canada 1983), gives a detailed outline of necessary preparation for field trips, including:

- obtaining specific instructions on sampling procedures;
- preparing an itinerary according to the sampling schedule;
- preparing lists of required equipment, supplies, and materials;
- ensuring that all sample bottles and bottle covers have been cleaned in accordance with standard procedures and are safely packed;
- ensuring that the laboratory has prepared the chemical reagents and standards needed for the trip; and
- preparing a checklist (including road maps, station location descriptions, field sampling sheets, labels, equipment manuals, tool box).

An additional presampling preparation that is necessary for some regions is the consultation of the project flowchart or project description manual (program outline) in order to determine which sites are to be sampled in sequential triplicate (as part of the field QC program) on a particular sampling trip.

4.1 Selection of Sampling Equipment

The collection of water samples requires the use of a variety of sampling equipment depending on the station, the medium to be sampled, and the parameter list. Generally, the WQB has used sampling irons, discrete samplers such as the Kemmerer sampler and the Van Dorn bottle, multiple samplers, dissolved oxygen samplers, and sediment grab samplers. The choice of sampler type must be closely related to the parameter list in order to avoid sample contamination. Past experience with the PVC Kemmerer-Plus has indicated contamination of heavy metal samples (Arseneault and Howell 1987). In addition to being parameter- and station-specific, samplers must be such that they provide suitable sample volumes and are suitable for use in a wide variety of environmental conditions.

4.2 Sample Bottles

Sample bottles are generally provided by the analytical laboratory. The selected bottles (including the liners and caps) must be free of contamination and suitable for the water quality parameters to be analyzed. Table 2 gives a description of the recommended containers for specific parameters and outlines the washing and preparation procedures used for ensuring the cleanliness of the sample bottles. The volume of sample required depends on the type and number of parameters to be analyzed, the analytical method, and the expected concentrations of the parameters in the water.

Recommended sample containers and washing and preparation procedures associated with the collection of suspended sediment and bottom sediment samples are outlined in the manual Sampling for Water Quality (Environment Canada 1983).

Prior to sampling, every bottle (or sample container) should be rinsed twice with the water to be sampled. During the rinsing process, bottle caps should be put on loosely to ensure adequate rinsing of the cap, liner, and bottle threads. If replicate samples are taken, every set of bottles should be rinsed twice and then filled, one set after the other (i.e., sequentially).

Proper labelling of the sample bottle is necessary to avoid possible mistakes and confusion when the sample arrives at the laboratory. All sample and subsample containers must be affixed with a strip of vinyl labelling tape when the containers are warm and dry. The NAQUADAT station number, sampling date and time, parameter(s) to be analyzed, and the program submitter identification (Sub. I.D.) must be clearly printed on each container with a waterproof fine-tipped nylon marker. Blank bottles should be similarly identified using a special NAQUADAT number that has been reserved for the identification of field blank samples.

4.3 References

Arseneault, R., and G. Howell. 1987. The Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program. Version 2. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Environment Canada. 1983. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

Table 2. Recommended Containers and Washing Procedures for Selected Water Quality Parameters

Parameter(s) to be analyzed	Recommended container	Washing procedure
Physical parameters Major ions Nitrogen: NO_2/NO_3 Nitrogen: NH_3 Nitrogen, total Dissolved organic carbon Dissolved inorganic carbon	1-L polyethylene bottle	Rinse 3 times with tap water, once with chromic acid, 3 times with tap water, once with 1:1 nitric acid, and then 3 times with distilled water, in the given order.
Suspended solids	2-L polyethylene bottle	Same as for physical parameters.
Phosphorus, total	0.5-L glass (sovirel) bottle	Rinse 3 times with tap water, once with chromic acid, 3 times with tap water, once with 1:1 nitric acid, and then 3 times with ultrapure distilled water, in the given order.
Metals, extractable	0.5-L polyethylene bottle	Rinse 3 times with tap water, once with chromic acid, 3 times with tap water, once with 1:1 nitric acid, and then 3 times with ultrapure distilled water, in the given order.
Chromium, extractable	0.25-L polyethylene bottle	Same as for metals, but chromic acid is used.
Mercury	0.125-L glass (sovirel) bottle	Rinse 3 times with tap water, once with chromic acid, 3 times with tap water, once with 1:1 nitric acid, and then 3 times with ultrapure distilled water, in the given order.
Organic toxic chemicals	1-L or 4-L glass (amber) bottles with Teflon-lined cap	Rinse 3 times with tap water, once with chromic acid, 3 times with organic-free water, twice with washing acetone, and twice with pesticide-grade hexane, then dry the uncapped bottle in a hot-air oven at 360°C for at least 1 hour.
Urea	1-L glass (amber) bottle	Same as for organic toxic chemicals.

5.0 ACCESSIBILITY TO SAMPLING SITES

Ready accessibility to some sampling stations that extend across the width of a water body can sometimes be difficult, and it is therefore customary to find a number of WQB stations located on bridges. The main station is generally located at the midpoint, and additional stations are established when spatial discontinuities are expected. All stations are located by permanent identification marks on the bridge structures.

Although sampling from bridges has some obvious advantages, there are some possible contamination problems. Most of these structures are made of metals, concrete, or creosoted timber, and caution must therefore be exercised to avoid heavy metal, major ion, and organic contamination, respectively. In addition, many of these structures are subject to heavy vehicular traffic, and thus there is a possibility of sample contamination by organics, heavy metals (e.g., leaded fuels), and road salts.

In order to avoid sample contamination while sampling from a bridge, all sampling should be conducted from the upstream side of the structure. When sampling from concrete structures, care must be taken to ensure that the movement of the sample rope does not result in concrete dust formation by the abrasive action of raising and lowering the sampler.

Some WQB stations require that samples be collected from the shore, which also results in QA/QC problems. Before establishing these stations, it may be necessary to perform some cross-sectional sampling to ensure that the littoral samples are representative of overall quality conditions. If samples are collected by wading, water should be taken upstream from the technician's position in order to avoid contamination by resuspended sediments.

Some samples are collected from stations that require the use of aluminum boats, rubber rafts, and, occasionally, helicopters. Use of these means of transport must be project-specific, with particular emphasis placed on the parameter list. Thus, if heavy metals are the major concern, the rubber boat should be used, whereas the aluminum boat is more suitable for the toxic organics program.

Regardless of the type of craft used, samples should never be taken off the stern of the boat, where floating oil and gasoline from the outboard motor might contaminate samples.

For lakes that have poor accessibility, it is often necessary to use a helicopter, and this increases the risk of contamination of samples for toxic organic chemicals with fuel and kerosene fumes. Thus, the data obtained from helicopter sampling of toxic organic chemicals must be viewed with extreme caution (Arseneault and Howell 1987).

5.1 References

Arseneault, R., and G. Howell. 1985. Field Quality Control/Quality Assurance Program Results of 1984/85 Sequential Triplicate Sampling. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

6.0 WATER QUALITY SAMPLING

6.1 Sample Representativeness

For new sampling programs, pilot projects should be conducted to resolve logistic problems and to implement procedures that would support the objectives of the program. This should include an evaluation of each collection site in terms of temporal and spatial homogeneity. Also, the collection frequency for priority variables, sample size, preservation effectiveness, shipping efficiency, and minimization of contamination risks could be determined or checked at this stage. In addition, proper sampling techniques must be used to ensure that a sample is representative of the flow in the sampled cross section.

The number of verticals to be sampled in a cross section should relate primarily to the collection of a representative sample and secondarily to the volume of the sample required (Kister and Garrett 1983). If field measurements of specific conductance, temperature, pH, and dissolved oxygen show the water to be well mixed, then a sample obtained at a single vertical near the centroid of the flow may be assumed to be representative of the total flow.

In situ measurements of pH, temperature, specific conductance, and dissolved oxygen should be taken before actual water samples are collected, in order to minimize water column disturbance caused by the lowering and raising of the sampling apparatus. This is particularly important in the case of residence time water bodies and during periods of ice cover when the vertical distribution of many parameters may be highly variable. Once the in situ measurements are conducted, water samples may be collected, starting at the surface and proceeding down through the water column if applicable. Collection of sediment samples should be performed after all water column samples have been taken.

Samples that are to be collected for suspended sediment or samples to be analyzed for total constituents should be collected by using approved and properly treated water–sediment depth-integrating samplers. Suggested instances where use of these samplers is not required are as follows (Kister and Garrett 1983):

- if there is extreme low flow, so that use of the sediment sampler is impractical; samples may be collected by immersing the bottle by hand;
- if there are flood conditions, when the velocity of flow is so great that the sampler cannot be lowered;
- when collecting samples of dissolved chemical constituents that are well mixed within the sampled cross section; and
- when aseptic collection techniques are required for bacteriological samples; these samples can be collected at a single vertical near the centroid of flow or by immersing the sample bottle by hand (when the water source is shallow).

In sampling for suspended sediment, cross-section samples should be obtained at every opportunity, using either the equal discharge increment or equal width increment method to obtain cross-section coefficients to apply to the point samples obtained by the sampler (Guy and Norman 1970).

6.2 Automatic Sampling

In many instances, particularly in accessible remote areas, automatic samplers are used very effectively. Automatic sampling can be more cost-effective at remote sites than manual sampling. The cost-effectiveness is closely linked to the instrument's reliability, improved capabilities, and greater sampling frequency ability.

For the efficient and effective use of an automatic water quality sampler, flow data should be available at every sampling station where the sampler is used. Flow data can provide a sound basis for interpretation of water quality results.

Unfortunately, automatic samplers are limited in their application. The sample handling procedures required in quantitative measurements for a wide range of water quality parameters (e.g., nutrients, major ions, total metals, toxic organics) make it virtually impossible to use automatic sampling for detecting and measuring parameters other than dissolved oxygen, pH, specific conductance, temperature, colour, and turbidity.

Samples collected at remote sites by automatic samplers should be retrieved at the earliest possible time. Samples collected in this manner should be analyzed only for constituents that do not require on-site preparation and could be assumed to be representative of the associated flow event.

High concentrations of suspended solids and floating materials can limit the performance of automatic samplers and reduce the accuracy of the measurements made. Suspended sediment samples collected by automatic samplers should be analyzed individually, and the specific conductance should be measured for each sample. This helps to determine the representativeness of the individual samples.

Periodic manual sampling should be conducted to verify the accuracy and representativeness of results obtained from automatic samplers. In selecting these instruments for limited water quality monitoring, attention must be paid to such factors as:

- the range of intended use;
- the level of accuracy desired;
- the skill level required for installing and servicing the automatic sampler;
- mechanical reliability of the sampler;

- adaptability of the sampler to varying weather patterns and flow regimes; and
- cost of the sampler.

6.3 Sampling in Winter Conditions

The WQB conducts its monitoring program throughout the entire year and must deal with a number of QA/QC problems during winter sampling, where ice conditions and low temperatures have considerable influence on sampling protocols (Arseneault and Howell 1987). Heavy ice conditions at many sites require the use of power ice augers, and thus contamination of toxic organic chemical samples by gasolines and oils must be considered.

During thaw periods, there is often a layer of meltwater immediately under the ice, and this water is not representative of the water chemistry of the system. Thus, care must be taken to ensure that samples are collected from a stratum that is below the ice–water interface.

The in situ measurement of pH, specific conductance, temperature, and dissolved oxygen during winter conditions must be carefully scrutinized, because some of the measurement meters (e.g., pH meters) do not function well in cold temperatures. Conductivity meters such as the Hydro TC-2 may give erroneous results (usually biased low) if slush or ice is allowed to build up around the thermistor or in the conductivity cell (Arseneault and Howell 1987). Similarly, dissolved oxygen probes can also suffer from buildup of ice and slush around the membrane. Therefore, in these winter conditions it may be necessary to determine dissolved oxygen concentrations by the Winkler method, even though biochemical oxygen demand bottle breakage and dissolved oxygen sampler freeze-up can also cause difficulties.

Another problem associated with winter sampling involves sample handling. It is essential that water samples are not allowed to freeze prior to analysis. This is particularly important for samples with high concentrations of organic matter, as freezing and subsequent thawing can result in flocculation of dissolved and colloidal organic compounds. Thus, it is necessary to work from a heated vehicle, such as a mobile laboratory, during the winter months.

6.4 References

Arseneault, R., and G. Howell. 1987. The Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program. Version 2. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Guy, H.P., and V.W. Norman. 1970. Field Methods for Measurement of Fluvial Sediments. Techniques of Water Resources Investigations of the United States Geological Survey, Book 3. U.S. Government Printing Office, Washington, D.C.

Kister, L.R., and W.B. Garrett. 1983. Field Guidelines for Collection, Treatment and Analysis of Water Samples. Arizona District, Ariz.

7.0 SAMPLING FOR SPECIFIC PARAMETERS IN SPECIFIC MATRICES

Many pesticides and heavy metals are toxic to aquatic life, even when these parameters are in low concentrations. The toxicity of a particular substance is dependent to a large extent on other water quality characteristics, such as temperature, pH, alkalinity, and dissolved oxygen. The complex nature of toxic chemicals (particularly trace organic toxic chemicals) demands the application of a number of QA/QC efforts that are additional to those generally used for less complex substances.

In the toxic chemical monitoring programs, water, bottom sediments, fish (forage and predator), and other biota samples are collected and analyzed.

7.1 Toxic Chemicals in Water

This section focuses on trace organic toxic chemical sampling, preservation, and analysis. However, the protocols and procedures that are outlined can be applied (with less complexity) to trace inorganic toxic chemicals (metals).

Sampling for toxic organic chemicals in water is conducted in a manner similar to any other type of water sample collection. Customarily, 4-L amber glass bottles with Teflon-lined caps are used, and rinsing of these bottles in the field is not required. Sample preservation is carried out using the procedures listed in Table 3.

Recently, a number of improved field QA/QC procedures have been instituted by the WQB (Atlantic Region) with regards to toxic chemical survey projects (Arseneault and Howell 1987). These procedures include the following:

- preparation of field blanks;
- preparation of spiked field blanks;
- preparation of spiked field samples (often in replicates);
- sequential replicate sampling; and
- preservation of spiked blanks and samples a few hours after spiking, to permit prolonged interaction between sample matrix and spiking solution.

In the preparation of a field blank, distilled deionized water is first added to a sample bottle prior to the field trip. The appropriate preservative is then added to the water blank in the field at the same time the collected samples are being preserved. The resulting blank is called a field blank. The field blank is used to check on:

- the cleanliness of sample containers;

Table 3. Procedures Used for the Preservation* of Water Samples Collected for Toxic Chemical Analyses

Parameter	Preservation
Pesticides • Organochlorine • Organophosphorous Polyaromatic hydrocarbons Polychlorinated biphenyls	Add 200 mL of hexane per 4-L sample (in a 4.5-L amber glass bottle), shake well, and deliver to the laboratory as soon as possible. Place a piece of aluminum foil over the mouth of the bottle before screwing on the cover (except when there are Teflon liners in the cap).
Chlorophenols	Adjust to pH 4 with 25% H_2SO_4.
Carbamates	Adjust to pH 3 with 25% H_2SO_4.
Chlorobenzenes	Add 50 mL of hexane per 1-L sample and shake.
Triazines	Adjust to pH 8 with 0.1 N NaOH solution, add 200 mL of methylene chloride per 4.5-L sample, and shake.

*All samples are kept as cool as possible in the field.

- random field contamination; and
- the purity of added preservatives.

Solvent blanks are also run during the analytical determinations in the laboratory.

Prior to its use in the blanks' preparation, the distilled deionized water should be checked for its purity in the laboratory. It is possible that the water may require some treatment (e.g., solvent extraction, filtering through XAD-type resin columns) before it is suitable for the required use.

Preparation of spiked field blanks is also required to check on the recovery of a spiked analyte of interest in a "clean" matrix. Spiking solutions are prepared by laboratory personnel prior to field trips. Preparation of spiked field blanks also serves to monitor the stability of the spiking solution in a "clean" matrix and to check the accuracy of its preparation. The analytical results of the spiked field blanks and the spiked field samples (which are prepared at the same time in the field) are compared to see if the analytical measurement process is indicating analyte recovery at the same level in a natural sample (matrix effects) as in a blank (no matrix effects).

Spikes are prepared by using an appropriate microsyringe (which has been previously rinsed with solvent) to add a precise volume (generally 100 to 200 μL) of a spiking solution (of known

concentration) into blanks or regular samples. All field personnel involved in the preparation of spikes should be trained in the correct use of a microsyringe. Larger volumes of less concentrated spiking solutions may be appropriate for the preparation of spikes. This could minimize possible manipulation errors when a spiking solution is added to a sample.

Spiked samples are generally prepared in duplicate or triplicate in the field to verify the reproducibility of spike recoveries and to gain a better understanding of the spiking process (e.g., matrix effects, systematic errors). Replication of spikes on all occasions is recommended. Based on resource availability, duplication of spikes could be an acceptable minimum, although triplication would possibly provide more information.

The addition of preservative to spiked blanks and spiked samples should be carried out a few hours after spiking; this time delay permits physical interaction between the sample matrix and the spiking solution.

In reporting the analytical results on the spikes, both the absolute concentrations and the percent recoveries should be reported by the laboratory.

The frequency with which this set of field QA/QC samples–namely, blanks, spiked blanks, spiked samples (in replicate)–should be prepared would be dictated by the water quality study or project and available resources.

Random sequential replicate sampling (see Chapter 8) should also be conducted regularly in addition to the above activities. Conducting sequential replicate sampling for 10 to 15% of all samples collected may be an adequate level of operation for the toxic chemicals QA/QC program. This sampling replication provides a means of monitoring the short-term variability of the parameters of interest. Additionally, the replication program in concert with information obtained from other field QA/QC samples may permit the discernment of contamination problems or contamination sources.

7.1.1 Liquid–Liquid Extraction Techniques

The use of liquid–liquid extraction techniques by the National Water Research Institute (NWRI) in collaboration with the WQB has permitted the detection of very low concentrations (sub–parts per trillion) of organic contaminants (Fox 1985; Environment Canada 1987). In the extraction process, large quantities of solvent are needed for 40-L samples, and the purity of the solvent must be assured. When the 40-L sample is filtered by the extractor (e.g., CRAB APLE Extractor), the suspended solids retained on the filter can also be kept for analysis.

The CRAB APLE Extractor is relatively inexpensive and readily available. Water collection, filtration, and extraction of both phases can be accomplished in less than an hour. Fox (1985) presented some data on recoveries of organochlorines in spiked organic-free water and natural Lake Ontario water. In general, recoveries were very good. However, its use was limited to organochlorine

contaminants, and thorough testing of this extractor is being conducted for other contaminants of interest.

Future studies involving the CRAB APLE and other extractors should address a number of QA/QC aspects, such as the preparation of field blanks and field spikes, to verify the recoveries attained by the extraction techniques and to investigate any possible contamination originating from the use of the extractor.

Another positive aspect associated with the CRAB APLE extraction system is its capacity for filtering the water and retaining the suspended solids on a filter for analysis. It could be used routinely for this purpose only.

7.1.2 Column Extraction

In the past, use has been made of Amberlite XAD-4 resin columns in the field for sample extraction of fenitrothion and aminocarb in rainwater (WQB [Atlantic Region]). The rainwater was collected and extracted with the aid of a special sampling (collection) device equipped with Amberlite XAD-4 columns. The sample columns were replaced each week by fresh ones, and the spent ones were brought to the laboratory and eluted with an appropriate solvent for the parameters of interest. The parameters were then identified and quantified analytically. After elution, the columns were regenerated with specific solvents, wrapped in clean aluminum foil, and returned to the field.

It is possible that these types of extraction columns could have some field application in the extraction of large volumes of surface water for the determination of specific organic contaminants (Leenheer and Noyes 1984; Woodrow *et al.* 1986). However, one drawback associated with resin columns is the requirement of exhaustive cleaning with various solvents to remove all trace contaminants. Preparation of blanks, spiked blanks, and spiked samples (in replicate) as well as sequential replicate sampling should be included as part of the specific QA/QC procedures that are needed, if these Amberlite XAD-4 resin columns (or others) are used more extensively in the future.

The inclusion of QA/QC procedures would assist in routinely checking the purity of the resin column(s), the cleanliness of the apparatus, the recoveries attained by the extraction and elution processes, the reproducibility of the extraction and elution efficiencies, and the representativeness of the water samples.

7.2 Toxic Chemicals in Bottom Sediment

The general procedures employed by the WQB for obtaining bottom sediment samples involve:

- the use of a 15 x 15 cm or 30 x 30 cm Ekman grab sampler as the collection device;
- transferring the sample to a Pyrex tray;

- removing the top 2.0 cm of sediments for toxic organic chemical analysis, using an aluminum scoop; and

- obtaining a sample for heavy metal determinations, using a plastic scoop; the subsample is taken from the centre of the grab sample to avoid sediments that have been in direct contact with the dredge.

In addition to the first grab, three or more grabs are taken, and the top 2.0 cm of sediments of each grab are put in a large plastic or stainless steel bowl (depending on the type of analyses required). The three or more portions are mixed together, using an appropriate scoop, to obtain a well-homogenized sample. Subsamples (often triplicate splits) are then placed in individual aluminum containers or, in the case of samples to be analyzed for metals, in whirl-pak polyethylene bags and frozen until they are used for analysis (Arseneault and Howell 1987).

All apparatus and containers involved in the collection and storage of sediment samples for organics are prewashed and rinsed with acetone and hexane.

In addition to the determination of the concentrations of the selected parameters, each bottom sediment sample is characterized by measuring water content and percent organic carbon (CHN analysis). This information is essential for the interpretation of results from the toxic chemicals analysis. In some studies, samples are also characterized through particle size analyses. When sampling bottom sediments, it is preferable to collect samples with high clay and organic matter content instead of rocks and sand, because it is known that pollutants are likely to be observed in the former type of bottom sediment matrices. This approach obviously places a bias on the sampling site selection, but adherence to the criterion (for the specific sediment constituents) must be encouraged (Arseneault and Howell 1987).

Generally, bottom sediment samples are taken from an enlargement of a river, which permits deposition of suspended sediments on the river bottom. In a lake, the situation is usually less critical, and samples are generally collected from the deepest point of the lake, especially when toxic chemical screening is the study objective. However, to obtain a good estimate of the spatial variability of parameters of interest within the bottom sediments, sampling should be performed at as many sites as possible within the given lake or river that is being surveyed.

As is the case with the water matrix, spiking of bottom sediment subsamples with organic contaminant standard solutions is carried out regularly in the field. About 20.0 g of bottom sediments are weighed in a prewashed aluminum sample container. Then 100 or 200 μL of spiking solution are added and thoroughly mixed with the sample. Following this, the container is closed and stored in a freezer (Arseneault and Howell 1987).

It is felt that the present containers used for holding the 20 g of spiked samples should be replaced with smaller ones, in an effort to prevent or reduce the risk of volatilization of some of the spiked contaminants. The WQB (Atlantic Region) has made a switch to 50-mL prewashed sovirel

bottles for the preparation of spiked bottom sediment samples. The sovirel containers appear to be better suited than the aluminum containers for housing the bottom sediment samples when toxic organic chemicals are the parameters of interest (Arseneault and Howell 1987).

At present, there is no bona fide body of reference samples of bottom sediments that can be used to prepare spiked field blanks, and it is proposed that a bank of reference material be prepared for this purpose. As is the case for water, blanks and spiked blanks are needed to check on contamination caused by containers and on accuracy, reproducibility, stability, and recovery efficiency of the spiking process.

The reference bottom sediment bank could consist of a large bottom sediment sample that would be previously approved by NWRI and the National Water Quality Laboratory (NWQL) through subsample analyses and spiked subsample recoveries. The banked sample would have to be sterilized and kept this way in order to eliminate any biological activity that could have an effect on the integrity of the added spiking solutions, when the spiked field blanks are prepared. For storing and handling purposes, it may be better to dry the bottom sediment sample and rehumidify portions of it as needed. With the use of the prepared standard reference material (SRM), comparisons can be made between spiked sample and spiked blank recoveries. This approach could permit a better interpretation of the analytical results.

Some QA/QC testing of the storage procedure currently used for bottom sediments is recommended. As previously noted, bottom sediment samples are frozen after collection, then brought to the laboratory where they are kept frozen until analysis. Samples are then thawed, so that an aliquot can be obtained for the analysis of specific parameters. After aliquot extraction(s), the samples are refrozen. Before taking an aliquot, samples are homogenized as best as possible in the container. As additional aliquots are taken within a few days (or weeks) for analyses of the same or other parameters, the same bottom sediment sample will be rethawed and refrozen successively. This procedure will be repeated until all analyses and rechecks are completed.

The effect that this multiple thawing and freezing has on the contaminants of interest in a sample is unknown (Arseneault and Howell 1987). Also, when the bottom sediment sample undergoes thawing, a water layer is very often observed in the container, the size of which seems to depend on the type of bottom sediments collected. It is uncertain how the water layer interacts with the solid sediment component during the thawing period. Perhaps some of the chemicals of interest may be released into the water by the sediment particles. This could be verified by conducting the required analyses on the water component.

Another cause for concern is the ineffective mixing of the water with the rest of the bottom sediment sample during the rehomogenizing process, should there be some of the chemicals of interest in the water as a result of the thawing action. Some laboratory analysts have indicated that it is very difficult to properly homogenize thawed bottom sediment samples in the aluminum containers that are currently in use, and that a move to more appropriate containers may be justified.

7.3 Toxic Chemicals in Biota

The collection and interpretation of biological data are more difficult than the collection and interpretation of water and bottom sediment data. Many of the organisms collected are mobile, and thus it is difficult to class them as representative of the site from which they are taken. Some water quality investigations involving biological parameters require monitoring or surveillance of an aquatic area over a long period of time. Such systematic re-sampling usually employs either a "transverse" or a "longitudinal transect" system, or a "grid" or "quadrant" system.

In a stream, transect sampling consists of collecting samples either along a section of its length or along a line across the stream. In a lake or a reservoir, transect sampling consists of collecting samples along a line that may be delineated by buoys. Samples may be collected at uniform intervals along the transect line or at a random location selected with the help of a table of random numbers.

A sampling grid or quadrant consists of an imaginary or physical rectangular arrangement of lines, covering all or part of a given habitat. For example, assume that a riffle measuring 10 x 30 m^2 is to be sampled for benthic organisms. If a 1-m^2 sampling device is used, there are potentially 300 sampling units, each of which could be assigned a number. If 15 sampling units are to be sampled, they may be selected by numbers taken from a table of random numbers (Snedecor 1956; Snedecor and Cochran 1967) or in some other randomized fashion. A grid or quadrant sampling scheme should, as does the transect scheme, give equal consideration to the various habitat types.

Regardless of the type or purpose of the biological sampling study, sample collection should be designed carefully. A large number of samples collected at the wrong time or place have less value than a few samples carefully selected as to time and place of collection. The frequency of sample collection will depend on the variability of environmental and biological factors and on the study objectives. The greater the habitat variability, the more intensive the sampling program must be. Life history events of the organisms must also be considered in the design of a sampling program. Some organisms have two or more generations per year; others have but one. Moreover, many aquatic insects spend only the juvenile period of their lives in water, emerging just prior to the adult transformation. In lakes and reservoirs, phytoplankton and zooplankton succeed one another in a somewhat rhythmic fashion; one may be abundant while the other is scarce. Vertical and horizontal movements of phytoplankton and zooplankton are also common in lakes and reservoirs. These facts, plus many others, increase the variability of the results of biological investigations.

In spite of the temporal and spatial variability of aquatic populations, statistical techniques are available for the design of sampling programs and for the evaluation of biological data. The subject of statistical sampling and analysis is discussed by Snedecor (1956), Steel and Torrie (1960), Stanley (1963), and Snedecor and Cochran (1967). Specific work on sampling statistics in aquatic biology has been reported by Needham and Usinger (1956), Ricker (1958), Elliot (1971), and Chutter (1972).

Many biological sampling activities of the WQB are centred around the collection of fish, with occasional sampling of freshwater clams, macrophytes, phytoplankton, and zooplankton. Fish are

collected by gill and seine netting, depending on the species required. For toxic organic and inorganic chemical analyses, small-size fish species (forage fish) are pooled in order to provide adequate sample volumes, whereas large fish are dissected and the egg mass, livers, and muscle analyzed separately. Care must be taken during dissection to avoid cross-contamination of samples. Samples are stored in aluminum containers or aluminum foil (for toxic organic chemical analyses) or plastic bags (for metal analyses) and frozen until analysis.

In the QA program for biota sampling, attempts must be made to maintain statistically significant sample sizes for each site. In addition, where appropriate, factors such as population size, size of habitat, feeding habits, weight, sex, lipid content, and geographical disposition of habitat (relative to industrial sites, agricultural lands and developments, waste dumps, and sewage inputs) should be determined to help in data interpretation.

Selection of discrete samples of the aquatic biota should be random, and the mean of five or more measurements on individual samples should be determined in order to give representative values of the parameters of interest during the laboratory analysis stage.

All dissecting equipment and sample containers must be prewashed appropriately, and dissection of fish by WQB personnel should be carried out preferably at the collection site in order to reduce the risk of contamination from avoidable extraneous sources.

Freshwater clams may be collected by hand in the shallow littoral zone and by a benthic drag in deeper regions. The clams may be shucked and pooled to provide adequate sample volumes and then stored in a manner similar to the storage of fish tissue.

Information on species' identification, sex, shell length, total weight, and body weight should be gathered. As in the case of fish species, this information could prove useful for data interpretation during consideration of toxic organic or inorganic chemical concentrations.

The infrequency in the sampling of macrophytes, phytoplankton, and zooplankton has resulted in a lack of established standardized QA procedures for treating the collection and preservation of these entities and has made it necessary at the present time to treat these classes of aquatic biota in a similar fashion to fish and freshwater clams.

Field spiking of biota should be attempted in those cases where a sample processor (grinder, homogenizer) of some type is available to prepare and mix samples with spiking solutions. Precaution against field contamination would be required during the extra manipulations of cutting, grinding (homogenizing), and extraction of the homogenate. Wherever possible, the laboratory should regularly prepare spiked samples in order to verify the efficiency of the employed analytical process (Arseneault and Howell 1987).

7.4 Toxic Chemicals in Packed Snow

Sample collection in packed snow can be accomplished by excavating a pit to the full depth of the annual snowpack, such that a clean cross section is available for sampling. The pit should be dug with a clean, acetone-rinsed metallic shovel or Teflon-coated aluminum shovel; the shovel should be subsequently cleaned in nearby snow (Sandberg *et al.* 1986). The bags in which the snow is placed either should be sealed airtight at the snow pit site or should be folded closed and sealed airtight at the first possible opportunity.

Several centimetres (in depth) of snow should be left undisturbed at the bottom of the pit so as to minimize incorporation of terrestrial material other than that which is incorporated within the profile through wind action. Additional samples could be taken in wide-mouth 500-mL polyethylene containers for the purpose of determining snow density. The polyethylene containers should be filled in such a way as to represent the entire section (as density varies with layers), thereby providing a composite density measurement. Whenever possible, replicate samples should be taken.

Once the sample (consisting of a number of boxes of snow per site) is melted, the water can be extracted using a continuous-flow liquid–liquid extractor (Gregor 1987). The container of dichloromethane, the solvent employed for the extraction, should be kept sealed until its use is required at the extraction site. The dichloromethane should be pretested for impurities prior to use, and dichloromethane blanks must be submitted to the laboratory with the samples.

The bottles used to hold the samples and the cap liners must have been washed, rinsed with pesticide-grade acetone and hexane, then baked for 8 hours prior to being sealed and shipped to the sampling or operational site. The bottles with their contents are sealed and stored frozen until shipped to the laboratory.

Samples for metals and trace elements should be placed in 250-mL Teflon bottles that have been washed in accordance with WQB bottle-washing procedures (Environment Canada 1983) and double-rinsed with sample water. Preservation should be carried out using 1 mL of 50% HNO_3 with subsequent refrigeration until delivery to the laboratory.

Major ion samples should be stored in 1-L polyethylene bottles that have been prewashed and subsequently double-rinsed with sample water. These samples are then refrigerated until delivery to the laboratory.

The in-field liquid-phase continuous-flow extraction technique is useful for the extraction of organic compounds and provides detection of the ultratrace concentrations expected in the snow. In addition, the extraction technique helps to minimize the possibility of sample contamination and simplifies the logistics and expense of shipping large quantities of samples from the snowpack region (e.g., the Arctic) to a laboratory that is thousands of miles away.

Refined QA/QC protocols for snowpack sampling will evolve as the frequency of this type of sampling increases.

An abridged description of precipitation sampling including QA/QC general procedures for site selection, sample collection, and handling is outlined in Sampling for Water Quality (Environment Canada 1983). A more detailed account of this sampling activity can be found in the Canadian Network for Sampling Precipitation (CANSAP) Manual (Environment Canada 1980).

Effective QA/QC procedures in the area of toxic organic chemical investigations related to aquatic sources are steadily evolving, and greater statistical confidence is being attached to both field and laboratory measurements. However, there is still a long way to go before the point is reached where specific QA/QC practices and protocols for toxic organic chemicals in various matrices (e.g., water, bottom sediment, biota) are comparable in effectiveness to those employed in the area of inorganic parameters in water.

7.5 Bacterial Sample Collection

Samples for bacteriological examination must be collected in bottles that have been carefully cleaned and autoclaved for approximately 20 minutes at 121°C at 15 psi. Also, glassware, except when in metal containers, may be sterilized in a hot-air oven for not less than 1 hour at a temperature of 170°C. Glassware in metal containers may be sterilized at 170°C for not less than 2 hours (U.S. Geological Survey 1982).

Samples from shallow streams should be collected from a single vertical near the centroid of flow. When collecting the sample, field personnel should face upstream, hold the collection bottle near its base, and immerse the bottle with its mouth directed towards the current.

Samples from deep streams can be collected from a bridge or cableway with a weighted bottle by immersing the sampler at a single vertical near the centroid of flow. The empty sampler should be immersed in the stream before the sample bottle is inserted in order to avoid the possibility of contamination of the sample. When the sample is collected, air space should be left in the bottle to facilitate mixing of the sample.

Bacteriological determinations should be commenced as soon as possible after sample collection, preferably within 1 hour and not more than 6 hours after sample collection. Samples must be chilled in ice during the time between collection and filtration. However, samples must not be subjected to freezing.

Residual chlorine in a water sample will destroy the biological population and may prevent an accurate determination of bacteria in the sample, unless the chlorine is destroyed at the time of sample collection. Therefore, if a sample from a halogenated water supply is taken, 1.0 mL of a 10% solution of sodium thiosulphate should be added to the 1-L sample bottle before sterilization in the field service

unit. This reagent should neutralize about 15 mg of residual chlorine per litre of sample and should show no effect upon variability or growth (Kister and Garrett 1983).

7.6 Sampling for Radioactive Parameters

Limited QA/QC directives for the sampling of radioactive parameters are outlined in the manual on Sampling for Water Quality (Environment Canada 1983). Additional principles and guidelines will evolve in this area as more specific sampling operations are conducted in aquatic systems that receive their radioactive pollutants from the major sources of radioactivity (e.g., uranium mining, nuclear power industries, nuclear weapons testing, the peaceful applications of nuclear materials and devices).

7.7 References

Arseneault, R., and G. Howell. 1987. The Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program. Version 2. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Chutter, F.M. 1972. A reappraisal of Needham and Usinger's data on the variability of a stream fauna when sampled with a Surber Sampler. Limnol. Oceanogr. 17(1): 139–141.

Elliot, J.M. 1971. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biol. Assoc. Sci. Publ. 25: 1–144.

Environment Canada. 1980. The Canadian Network for Sampling Precipitation (CANSAP) Manual. Atmospheric Research Directorate, Atmospheric Environment Service, Toronto, Ont.

Environment Canada. 1983. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

Environment Canada. 1987. Proceedings of a Workshop on Preconcentration of Trace Organic Compounds from Large Volumes of Water. Water Quality Branch and National Water Research Institute, Canada Centre for Inland Waters, Burlington, Ont.

Fox, M.E. 1985. A Practical Sampling and Extraction System for the Quantitative Analysis of Sub-ng/L of Organochlorine Contaminants in Filtered Water and Suspended Solids. National Water Research Institute, Canada Centre for Inland Waters, Environment Canada, Burlington, Ont.

Gregor, D.J. 1987. Evidence of Atmospheric Transport and Deposition of Toxic Substances in Canadian Arctic Snow. Water Quality Branch (Western & Northern Region), Inland Waters Directorate, Environment Canada, Regina, Sask.

Kister, L.R., and W.B. Garrett. 1983. Field Guidelines for Collection, Treatment and Analysis of Water Samples. Arizona District, Ariz.

Leenheer, J.A., and T.I. Noyes. 1984. A Filtration and Column Adsorption System for On-site Concentration and Fractionation of Organic Substances from Large Volumes of Water. U.S. Geological Survey Water Supply Paper 2230, Washington, D.C.

Needham, P.R., and R.L. Usinger. 1956. Variability in the macrofauna of a single riffle in Prosser Creek, California, as indicated by the Surber Sampler. Hilgardia 24(14): 383–409.

Ricker, W.E. 1958. Handbook of computations for biological statistics of fish populations. Fish. Res. Board Can. Bull. 119: 1–300.

Sandberg, D.K., N.R. McQuaker, R.J. Vet, and H.A. Weibe. 1986. Snowpack: Recommended Methods for Sampling and Site Selection. Federal/Provincial Research and Monitoring Coordinating Committee, Quality Assurance Subgroup, Downsview, Ont.

Snedecor, G.W. 1956. Statistical Methods Applied to Experiments in Agriculture. 5th ed. Iowa State Univ. Press, Ames, Iowa, 534 pp.

Snedecor, G.W., and W.G. Cochran. 1967. Statistical Methods. 6th ed. Iowa State Univ. Press, Ames, Iowa, 593 pp.

Stanley, J. 1963. The Essence of Biometry. McGill Univ. Press, Montreal, Que., 147 pp.

Steel, R.G.D., and J.H. Torrie. 1960. Principles and Procedures of Statistics. McGraw-Hill Book Co., New York, N.Y., 481 pp.

U.S. Geological Survey. 1982. Methods for Collection and Analysis of Aquatic Biological and Microbiological Samples. Book 5, Laboratory Analysis, Chapter A4. Techniques of Water-Resources Investigations (TWRI).

Woodrow, J.E., M.S. Majewski, and J.N. Seiber. 1986. Environ. Sci. Health B21(2): 143.

8.0 SEQUENTIAL TRIPLICATE SAMPLING

The sequential triplicate sampling program is designed to assess the efficiency and effectiveness of the overall water quality monitoring program. The program serves to evaluate whether the water sampling performed is representative, aids in discerning contamination or analytical problems, and reveals problems associated with water quality data management. Overall, the sequential triplicate sampling program covers the entire QA picture by providing some information on the combination of field, laboratory, and data management components. When a problem is revealed through the use of sequential triplicate sampling, other QC measures (field or laboratory) can be used to elucidate and correct that problem (Arseneault and Howell 1987).

Sequential triplicate sampling may not define the exact source(s) of a problem, but it can be used as an aid in determining the cause(s) of the problem.

Triplicate sampling is more effective than duplicate sampling; the use of three replicates delivers more statistical information than the use of two replicates. Higher replication (e.g., quadruplicates, quintuplicates) would be more statistically significant, but the increased cost could outweigh the benefit of the potential additional information. In cases where there are very specific problems (regarding, for example, sampling, preservation, contamination, storage, or analysis), increased repeat sampling may be deemed necessary in order to provide conclusive experimental and statistical evidence relating to the source(s) of the problems.

In order to conduct a comprehensive evaluation of the triplicate results, it is essential to define the acceptable degree of reproducibility between triplicate samples. One accepted rule of thumb is to set the maximum variation between triplicate results at a level of 10% (Arseneault and Howell 1985). The arbitrary limit depends to some extent on the complexity of any or all of the stages within the monitoring program for any given parameter or suite of parameters. In the case of toxic organic chemicals, the maximum limit could possibly be extended from 10 to 20% for results at sub-parts per million and sub-parts per trillion levels or 50% or greater near the detection limit. However, upper-parameter-based limits of variation should not be chosen in an ad hoc fashion but should be consistent not only with valid and documented triplication data, but also with parameter-associated operational complexities.

The frequency of sequential triplicate sampling in a number of WQB regions is currently running at a level of 15 to 20% of all collected samples. This commitment should remain the same until rigorous data assessments are made of the results collected in the past 5 to 10 years. For those projects that have a small number of stations, one triplicate sample may be collected for each sampling trip, and thus, in practice, replication may occur in the range of 20 to 25%.

In order to ensure equal weighting within designated projects, stations to be replicated should be selected from a predetermined rotational schedule. This approach overcomes the possibility of field technician bias towards easily accessible stations.

Currently, only the median value originating from the three replicate results for each parameter per site is recorded in the NAQUADAT system. However, it is possible that more information from the sequential triplicate sampling program will be included in the system in the future.

Table 4 presents a partial record of the results from a sequential triplicate sampling program conducted by the WQB (Atlantic Region) in the period 1984 to 1985. Over 80% of the variations from the median values were less than 10%, indicating the possibility for achieving substantial success with the program.

The usefulness and ultimate success of a sequential triplicate sampling program will rest heavily on the efforts that are directed towards the evaluation and interpretation of the collected data, as well as on the use of the results from these two data processes.

8.1 References

Arseneault, R., and G. Howell. 1985. Field Quality Control/Quality Assurance Program Results of 1984/85 Sequential Triplicate Sampling. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Arseneault, R., and G. Howell. 1987. The Water Quality Branch Atlantic Region Field Quality Assurance/Quality Control Program. Version 2. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Table 4. Partial Results of Sequential Triplicate Sampling of Selected Stations Included in the Nova Scotia LRTAP Lakes Monitoring Program (1984 to 1985).

NAQUADAT difference No.	Sample date	Aluminum (Al) extractable		Manganese (Mn) extractable	
		Actual values (mg/L)	% Difference from median	Actual values (mg/L)	% Difference from median
01NS01DA0004	02/05/84	0.110	0	0.02	0
01NS01DA0004	02/05/84	0.110*	-	0.02*	-
01NS01DA0004	02/05/84	0.110	0	0.02	0
01NS01DA0006	02/05/84	0.087*	-	0.02	50.00
01NS01DA0006	02/05/84	0.100	14.94	0.01*	-
01NS01DA0006	02/05/84	0.085	2.30	0.01	0
01NS01DA0007	02/05/84	0.210*	-	0.02	0
01NS01DA0007	02/05/84	0.200	4.76	0.01	50.00
01NS01DA0007	02/05/84	0.220	4.76	0.02*	-
01NS01EA0010	02/05/84	0.100	0	0.02	0
01NS01EA0010	02/05/84	0.100*	-	0.02*	-
01NS01EA0010	02/05/84	0.110	10.00	0.02	0
01NS01ED0013	02/05/84	0.170*	-	0.01	0
01NS01ED0013	02/05/84	0.160	5.88	0.01*	-
01NS01ED0013	02/05/84	0.170	0	0.01	0
01NS01ED0021	02/05/84	0.210	5.00	0.01	0
01NS01ED0021	02/05/84	0.200*	-	0.01*	-
01NS01ED0021	02/05/84	0.200	0	0.01	0
01NS01DA0003	16/10/84	0.014	0	0.01	0
01NS01DA0003	16/10/84	0.014*	-	0.01*	-
01NS01DA0003	16/10/84	0.018	28.57	0.01	0
01NS01EA0002	16/10/84	0.022	0	0.02	0
01NS01EA0002	16/10/84	0.022*	-	0.02*	-
01NS01EA0002	16/10/84	0.022	0	0.02	0
01NS01ED0006	17/10/84	0.043	4.44	0.05	0
01NS01ED0006	17/10/84	0.045*	-	0.05*	-
01NS01ED0006	17/10/84	0.045	0	0.05	0
01NS01ED0008	17/10/84	0.210	8.70	0.01	0
01NS01ED0008	17/10/84	0.240	4.35	0.01*	-
01NS01ED0008	17/10/84	0.230*	-	0.01	0
01NS01ED0010	17/10/84	0.280	9.68	0.01	0
01NS01ED0010	17/10/84	0.320	3.23	0.01*	-
01NS01ED0010	17/10/84	0.310*	-	0.01	0
01NS01ED0013	17/10/84	0.090	3.44	0.01	0
01NS01ED0013	17/10/84	0.065	25.29	0.01*	-
01NS01ED0013	17/10/84	0.087*	-	0.01	0
01NS01ED0017	17/10/84	0.100	0	0.01	0
01NS01ED0017	17/10/84	0.100*	-	0.02	50.00
01NS01ED0017	17/10/84	0.100	0	0.01*	-

* Median value.

9.0 SAMPLE HANDLING, PRESERVATION, STORAGE, AND TRANSPORTATION

Efforts must be made to minimize errors that can be introduced as a result of collecting and handling the sample. The objective is to provide the laboratory with a set of samples that closely represent the aquatic environment from which they were taken.

To ensure consistency and efficiency, sample handling (e.g., filtration, decantation, centrifugation, sample splitting), preservation, storage, and transportation procedures must be properly and accurately documented and adhered to by field personnel.

Preservatives should be prepared from Ultrex-grade or similar grade chemicals, and care must be taken to ensure that the water sample is not contaminated by impurities residing in the added preservative.

In adding preservatives to field blanks, the same level of caution should be exercised as with actual samples. The practice of adding ultrapure distilled water to the field blank bottles in the laboratory prior to the field trip should be encouraged. The preservation of blanks can then be carried out in the field. Previously, in some regions, large carboys of ultrapure distilled water were taken on each trip to be used as rinse water and for blank preparation. However, as this water was kept in the containers for varying lengths of time, it was felt that contamination by leaching of materials (e.g., metals) from the walls of the vessels was a possibility.

It is necessary to pass the water sample through a filter of specified porosity, type, and quality when the concentration of dissolved inorganic constituents has to be determined. Some filters and filtration apparatus may require laboratory pretreatment and must also be rinsed with a portion of the collected sample (subsequent to rinsing with ultrapure water) before the filtrate is collected. The filtrate should be preserved (if required) by adding the proper preservative. The glassware utilized should be specific to the parameter that is to be analyzed. The filtering apparatus should be placed in transportation cases that permit proper storage and provide protection against shock and dirt. The transportation cases can also be used as stable work platforms for filtration.

Filtration should be performed at the end of the day, to ensure the same conditions for all samples. A filtration blank for each parameter or suite of parameters should be prepared from ultrapure distilled water using the same procedure as that used for the sample. At the end of a set of filtration runs, the glassware must be rinsed with an appropriate solution (e.g., diluted acid solution for phosphorous and metal parameters) and afterwards with ultrapure distilled water. All glassware should be rinsed with ultrapure distilled water prior to filtration and between filtrations of individual samples. The glassware and filter should be rinsed by filtering the excess water, and the sample bottle used for collecting the filtrate should be rinsed twice with filtrate.

In order to prevent chemical transformation due to photooxidative processes or biologically mediated reactions, bottles containing samples should be stored in coolers during transportation. Refrigeration at 4°C should be employed for the maintenance of the quality of both preserved and

unpreserved water samples. When cooler space is limited, priority should be given to the housing of unpreserved samples in the mobile laboratory's refrigerator. In the absence of a refrigeration unit, "koolatrons" should be used to keep the samples cooled during transportation. If a WQB mobile laboratory is not equipped with a refrigerator, field activities and transportation schedules must be planned in such a manner to ensure expedient return of samples to the laboratory.

Water samples must be well stoppered and packed to prevent spillage or breakage. Labels bearing the sample's identification, destination, and the word "FRAGILE" must be attached to each container. The top of the carton must be clearly identified as "THIS END UP," and the containers in a shipment must be numbered.

A check must be made to ensure that all sample bottles recorded on the field sampling sheets have been placed in a given carton before shipping. The shipping date and mode of transport must be indicated on the field sampling sheet.

Samples from any one location should be kept together, except in cases where all bottles of one size must be shipped together because of container size. When samples from one station must be separated and placed in more than one carton, a copy of the field sampling sheet pertaining to the bottles must be enclosed in each box.

Additional QA/QC information on sample handling, preservation, storage, and transportation is given in the WQB Sampling for Water Quality manual (Environment Canada 1983).

9.1 References

Environment Canada. 1983. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

10.0 MAINTENANCE AND CALIBRATION OF INSTRUMENTATION

10.1 General Maintenance and Calibration

The performance of an instrument must be checked out and appraised by a qualified person before it is transported to the field. This involves a visual inspection and verification of its operation, including the zero and full-scale calibration.

Documents on standardization, calibration, maintenance, equipment safety, and spare parts should accompany each instrument.

Specially designed briefcases should be used to pack and protect each instrument and piece of equipment during transportation. Conductivity meters, pH meters, dissolved oxygen meters, turbidity meters, and thermometers should be encased in robust and shock-resistant housings, and the electrically operated meters should be open to as little electrostatic and electromagnetic interference as possible. In addition, caution notices for the protection of the equipment during transportation should be visible, and packing instructions indicating environmental tolerance limits (e.g., in relation to temperature and humidity) should be included in order to prevent the exposure of equipment to conditions that could affect performance.

The standardization of instruments should be performed using reference standards when these are available, or against certified standard instruments when they are not. This should be done before the instrument is transported to the field; in some cases, the reference instrument could be transported to the field and employed under actual field conditions. Laboratory-standardized instruments should be rechecked upon arrival in the field.

Calibrations must be conducted under the same instrumental and chemical conditions as those that will exist during the measurement process. The frequency of calibration will depend on the accuracy requirements of the investigation and the stability of the instruments. Daily calibration checks are recommended when the instrument is in daily use; calibration checks should be performed immediately prior to a series of measurements at other times. For unstable instruments, the calibration should be checked prior to each series of measurements, in between measurements, and after the last measurement.

The calibration process is vital to all measurement programs and should be governed by a calibration plan. Such a plan should provide for:

- calibration procedures and record forms;
- stated calibrated frequencies;
- appropriate sources for obtaining certified and high-quality standards, or the best means of producing accurate in-house standards;

- a list of all calibration standards (including nomenclature and assigned identification numbers);
- specifications of environmental conditions; and
- intended range of validity.

Calibration procedures should include information on the following:

- the specific equipment or groups of equipment to which the procedure is applicable;
- a brief description of the scope, principle, or theory of the calibration method (an example and a reference may also be included);
- calibration specifications, such as the number of calibration points, environmental requirements, and precision and accuracy requirements;
- a list of the calibration standards and accessory equipment needed to perform an effective calibration, manufacturer's name, and instrument model number;
- a complete, clear, concise, step-by-step written calibration procedure;
- specifications for calibration facilities, equipment, temperature, and humidity, and physical protection for calibration standards; and
- specific instructions for obtaining and recording the test data (data forms should also be furnished).

Pre- and postshipping checks may verify the performance of an instrument during its operation and could reveal the occurrence of measurement drift.

10.2 Maintenance of Specific Instrumentation

The precision thermistor temperature sensor of the Hydrolab Model TC-2 should be checked at least every six months against an ASTM Certified Mercury Thermometer in the laboratory at controlled temperatures that span the range of temperatures encountered in the field.

The four-electrode configured specific conductance cell of the Hydrolab Model TC-2 should be checked before the commencement of each field trip. The check should be performed against the specific conductance of two freshly prepared standards at the 74 and 177 μS/cm levels. Calibration can be routinely carried out at each sampling site by switching to the provided internal standards.

The combination glass electrode should be calibrated daily in the field at pH values of 7.0 and 4.0 using commercially prepared buffers that are kept at a temperature difference of $\pm 1.0^{\circ}$C from the

sampling temperature. If the temperature of the pH buffers is not ±1.0°C from that of the sample(s), it is adjusted by placing the buffer bottles in a cold or hot water bath or in cold air, whichever is applicable.

Calibration curves should be prepared for each range of the turbidity meter by using appropriate standards that have a range of 0 to 1000 nephelometric turbidity units (NTUs). The standards are generally supplied by the manufacturer but can be prepared from stock standard turbidity suspensions. Tests should be conducted on at least one standard in each range to be used.

The YSI (Yellow Springs Instrument) dissolved oxygen probe should be visually examined prior to a field trip to check the condition of the membrane. Defective membranes must be replaced. The instrument should be tested each day at the first site by using the "Azide Modification of the Standard Iodometric Method" (with freshly prepared reagents) on two samples. The resulting values should be averaged and the meter then set (with the calibration knob) to read the average value of the samples. On the same day, only air calibrations need to be performed at the other sites. Calibration of the meter can also be accomplished using a calibration chamber in which the dissolved oxygen probe is placed and immersing the chamber in the stream or lake to carry out the measurement. The latter approach provides an adequate environment for proper calibration.

All equipment must be kept clean and in good working condition, using the techniques described by the manufacturer or those techniques set out in the Field Methods Manual (Environment Canada 1983).

Each instrument should have a logbook detailing its operational history, and all data originating from checks, calibrations, and standardizations should be recorded in the logbook as support data.

10.3 References

Environment Canada. 1983. Field Methods Manual. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

11.0 FIELD TESTING

11.1 Quantitative Analysis

11.1.1 Application and Scope

This practice is applicable to field analyses involving the use of automated measurements for the determination of pH, specific conductance, dissolved oxygen, and turbidity.

11.1.2 Preparation of Standards, Blanks, Reagents, and Sample Solutions

Stock standard solutions should be prepared in the laboratory before the sampling trip is undertaken. The stock solutions should be well preserved according to the directions in the analytical methods outline (for the particular parameter to be determined).

The intermediate standards must be freshly prepared in the field by the method of serial dilution, just before the commencement of calibration and measurement exercises. A minimum of five to six working level standards spaced evenly over the analytical range should be acceptable.

The blanks must of necessity be composed of the same matrix (deionized water or ultrapure distilled water) with which the samples, standards, and reagent solutions are constituted.

Reagents should preferably be fresh or kept preserved as set out in the Analytical Methods Manual (Environment Canada 1979). Note that some reagents will deteriorate with age or on continuous exposure to such agents as ultraviolet light and heat.

Whenever possible, modification of the original sample solutions should be avoided, except in cases where operations such as decantation, filtration, or dilution are deemed necessary.

Accurate volumetric measurement of sample aliquots cannot be overstressed.

11.1.3 Calibration and Measurements

The apparatus should be set up as specified in the appropriate analytical procedure, and the proper analytical working conditions should be selected.

The analytical procedure for calibrating the instrument should be followed, and a complete set of standards (minimum of five or six) should be run. By using blanks at the beginning and end of a calibration run, changes in instrument behaviour may be detected.

When the calibration curve has been satisfactorily constructed, a number of appropriate sample solutions should be chosen and the values of the parameter of interest determined using the calibration curve.

11.1.4 Standard Addition Technique

The method of standard addition (U.S. Geological Survey 1982) is used to compensate for known matrix effects or analytical interferences that can cause inaccuracies in analyte determinations. This technique is useful only if the measured analytical response is linear with respect to the added quantity of the parameter being determined, and if the observable interference is independent of the same quantity.

Because both samples and standards are affected equally, it is not necessary to prepare matrix water that is comparable with the unknown sample in order to correct the analyses.

In practice, a preliminary analysis is made for the water quality parameter of interest, and a blank and three standards containing different amounts (generally multiples of each other) of the parameter to be analyzed are prepared. Equal volumes of the sample are then added to the blank and three standards. The analytical responses are plotted on the vertical axis, and the horizontal axis is used for plotting the known quantities of the parameter prior to the addition of the sample. The horizontal axis is extended to the left of the vertical axis and is scaled backwards from the zero (blank) quantity of the analyte of interest.

A line is drawn through the plotted points (Figure 2) and then extrapolated to the zero analytical response. The value at the intercept is recorded as the quantity of the parameter in the sample.

11.2 Precision and Accuracy Estimates

For replicate analyses involving precision and accuracy measurements, a value that conveniently lies at a position of either one-half or three-quarters of the calibration range is chosen for the parameter of interest. For the total precision estimate (involving collection, handling, and analytical processes), measurements must be made on samples collected in replicates; for the precision estimate related to the collection and handling processes, measurements must be made on sample splits (replicates from a single water sample).

Ideally, for good estimates of both types of precision, no less than 10 repeat measurements should be performed. Five to six replicated measurements may be adequate in situations of time constraints or shortfalls in sample volumes or sample collections.

From the determinations of both the total precision (CV_T) and the analytical precision (CV_L), the field sampling and sample handling precision (CV_F) can be calculated with the use of the approximate Equation 5:

$$\textbf{Field precision } (CV_F) = \textbf{Total precision } (CV_T) - \textbf{Analytical precision } (CV_L) \qquad (5)$$

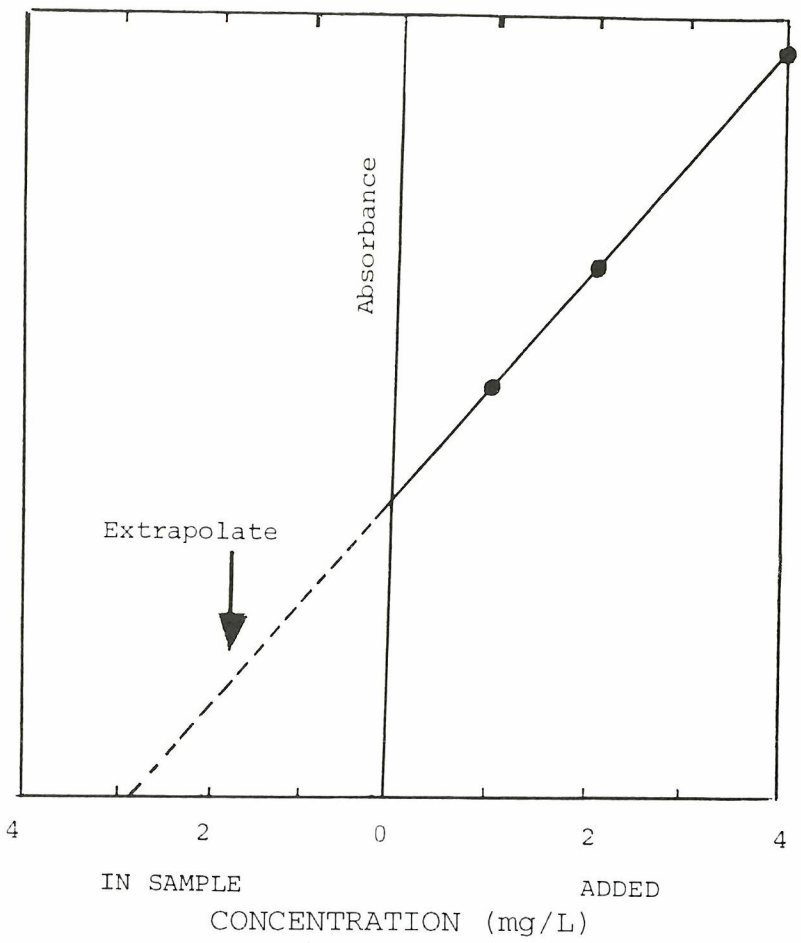


Figure 2. Example of standard addition method (from Skougstad *et al.* 1979).

The coefficient of variation (CV), otherwise called the relative standard deviation (RSD), is generally used to represent the precision of a set of replicate measurements and is defined as follows:

$$CV\ (\%) = \left(\frac{S}{\bar{X}}\right) \times 100 \qquad (6)$$

where, for N replicate measurements,

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad (7)$$

and

$$S = \sqrt{\frac{\Sigma (X_i - \overline{X})^2}{N - 1}} \quad (8)$$

If standard reference materials (SRMs) are available, the accuracy of the field analytical measurements can be determined. The accuracy of the determination can be expressed as a percent error according to Equation 9:

$$\% \text{ error} = \left(\frac{\overline{X} - X_{SRM}}{X_{SRM}} \right) \times 100 \quad (9)$$

where: $\overline{X}$ is the mean value of replicate determinations of an SRM; and
X_{SRM} is the certified value of an SRM.

In the absence of SRMs, it is possible to estimate the accuracy by determining the "spike recovery" of a given parameter (see Chapter 5 of Part III). Recovery is calculated on a percentage basis as follows:

$$\% \text{ recovery} = \left(\frac{C_F - \overline{C}_B}{C_A} \right) \times 100 \quad (10)$$

where: C_F is the measured concentration in the sample;
$\overline{C}_B$ is the average concentration of the blank; and
C_A is the known concentration added to the sample.

The following example outlines the determination of precision and accuracy statistics in field analytical measurements for samples taken from a given site.

* * *

EXAMPLE:

Field and laboratory analytical measurements (Tables 5 and 6) were analyzed for precision and accuracy values, and a comparison was made between field and laboratory results.

Table 5. Specific Conductance (μS/cm) Results Obtained in the Field

Data for individual samples (replicates)		Data for sample splits		Spiked and unspiked blanks
(1)	306.1	(1)	305.0	(1) 0.3 (blank)
(2)	305.2	(2)	305.0	(2) 706.2 (spiked)
(3)	306.2	(3)	304.4	(3) 706.0 (added)
(4)	307.1	(4)	305.0	
(5)	305.3	(5)	305.2	
(6)	304.5	(6)	305.0	
(7)	305.4	(7)	305.0	
(8)	306.2	(8)	305.0	
(9)	307.0	(9)	305.0	
(10)	308.0	(10)	304.4	

Table 6. Specific Conductance (μS/cm) Results Obtained in the Laboratory

Data for individual samples (replicates)		Data for sample splits		Spiked and unspiked blanks
(1)	307.0	(1)	305.0	(1) 0.2 (blank)
(2)	306.1	(2)	305.0	(2) 706.4 (spiked)
(3)	307.0	(3)	305.2	(3) 706.0 (added)
(4)	306.3	(4)	305.1	
(5)	304.8	(5)	305.0	
(6)	305.2	(6)	305.1	
(7)	305.4	(7)	305.2	
(8)	308.0	(8)	305.1	
(9)	306.2	(9)	304.3	
(10)	306.0	(10)	305.0	

The use of Equation 5 permits the calculation of precision values associated with field sampling and sample handling. These precision values (Table 7) are larger than the corresponding values obtained for the analytical operation. The higher the numerical value of the precision statistic, the lower the measurement precision.

Precise data for analytical operations can be acquired from measurements that are under statistical control and should therefore show less variability (greater precision) than the corresponding field sampling and sample handling activities.

Table 7. Precision and Accuracy Statistics Associated with Field and Laboratory Measurements of Specific Conductance

Statistic	Field value (%)	Regional laboratory value (%)
(1) Total precision = CV_T (individual samples)	$\frac{0.940}{306.1} \times 100$	$\frac{0.953}{306.2} \times 100$
	= 0.307	= 0.311
(2) Analytical precision = CV_L (subsamples)	$\frac{0.260}{304.9} \times 100$	$\frac{0.258}{305.0} \times 100$
	= 0.085	= 0.085
(3) Precision for field sampling and sample handling = CV_F ($CV_F = CV_T - CV_L$)	0.222	0.226
(4) Accuracy	99.99	100.03

Table 7 also reveals that the precision and accuracy values that were determined in the laboratory show only small differences from the corresponding values obtained in the field measurements. This suggests that the samples did not undergo any significant change in constitution during the preservation, storage, and transportation phases.

* * *

The comparison of field and laboratory statistics pertaining to the same sample(s) can generally serve as an indicator of sample contamination or sample change during the sample handling process; if the difference of the means (M_d) for the corresponding field and laboratory analytical results is compared with the average standard deviation (S_a) for the same two sets, and

$$\frac{M_d}{S_a} > 4 \qquad (11)$$

then there exists the possibility that contamination, deterioration, or some change has occurred within the sample. In the above example, the application of the criterion gives favourable results. For example:

$$\frac{306.2 - 306.1}{0.309} = 0.32 \qquad < 4$$

and

$$\frac{305.0 - 304.9}{0.085} = 1.18 \qquad < 4$$

The criterion in Equation 11 is arbitrary and serves only as an "Empirical Test Error" (Taylor 1984; Gaskin 1986).

11.3 Measurement Comparisons

Intercomparison programs could be used appropriately to evaluate the performance and variability (bias, precision, and accuracy) of any part or all of the measurement system. The intercomparison exercises could consist of on-site comparison of different instruments measuring the same parameter. Standard solutions should be used in the measurements.

Unsatisfactory results from intercomparison tests should lead to the reassessment of equipment calibration procedures, frequencies of instrumental checks and calibrations, and operating procedures. Errors in operator performance may also be pinpointed during assessment of the test results.

11.4 References

Environment Canada. 1979. Analytical Methods Manual. Water Quality Branch, Inland Waters Directorate, August.

Gaskin, J.E. 1986. Procedures for the Acquisition of Precision and Accuracy Data in Field Sampling and Field Testing. Water Quality Branch (Western & Northern Region), Inland Waters Directorate, Environment Canada, Regina, Sask.

Skougstad, M.W., M.J. Fishman, L.C. Friedman, D.E. Erdman, and S.S. Duncan (eds.). 1979. Methods for the Analysis of Inorganic Substances in Water and Fluvial Sediment. U.S. Geological Survey Techniques of Water-Resources Investigations (TWRI), Book 5, Chapter A1, p. 32.

Taylor, J.K. 1984. Quality Assurance of Chemical Measurements. Center for Analytical Chemistry, National Bureau of Standards, Gaithersburg, Md.

U.S. Geological Survey. 1982. Methods for Collection and Analysis of Aquatic Biological and Microbiological Samples. Book 5, Laboratory Analysis, Chapter A4. Techniques of Water-Resources Investigations (TWRI).

12.0 DATA COLLECTION

12.1 Recording Field Data

Accurate and reliable records must be kept of all field data. Detailed accounts of the station and sampling locations must include distances related to specific reference points. Descriptions of the water body should include the banks on either side, the bed material, irregularities in morphology that affect flow or water quality, seasonal conditions, and natural or man-made conditions that may have a bearing on water quality. Wherever and whenever possible, a detailed sketch should be made of the location of the station. A combination of a map that locates the sampling station on a larger scale with respect to roads, highways, and towns and the sketch of station location should provide complete location information.

The station data form must be completed so that the recorded information can be subsequently stored in NAQUADAT. The information required to complete this form is outlined in the NAQUADAT (NAtional water QUAlity DATa base) Guide to Interactive Retrieval (Whitlow and Lamb 1983) and includes:

- the station number;
- location, as latitude and longitude or UTM coordinates;
- narrative description;
- reference station; and
- station parameters.

A number of recordings must be made on the field sheets that are issued to the sampling personnel. Records must be made of such items as:

- experimental observations;
- sampling date;
- time;
- location;
- weather patterns;
- excessive algae growth;

- dead fish or debris;
- grease or oil slicks on the water surface;
- unusual colour or odour of the water;
- a preponderance of leaves and detritus; and
- other phenomena.

An example of a systematic format for recording field analyses and observations is provided in Figure 3.

Particular attention must be paid to QC requirements for the sampling operation and the field measurements. In addition, records must be kept of the sampling techniques employed (e.g., depth-integrated, grab, sequential replication, spatial or temporal sampling), and the data obtained from field analytical determinations must be accurately recorded.

Details regarding either the erratic behaviour or the less-than-expected performance of any piece of equipment or apparatus must be recorded. These details may later help to explain any anomalies in the experimental data.

12.2 Support Data

Support data are data collected during the measurement program, and their use could play a valuable role during the interpretation process. Support data must be easily available via EDP capability or field and laboratory books. A variety of information may be considered as support data:

- data charts and printouts;
- equipment performance records;
- calibration records;
- operation logs;
- environmental conditions (e.g., weather patterns, tides, precipitation) prior to and during sampling;
- measurement comparison records;
- quality control and system audit records; and
- records of corrective actions.

STATION NO. _____
DESCRIPTION: _____
DATE OF SAMPLING DY _____ MO _____ YR _____
TIME OF SAMPLING HR _____ MI _____ TIME ZONE _____
SAMPLED BY _____

FIELD MEASURED PARAMETERS
Water Temp. °C _____ Air Temp. °C _____
pH _____ Specific Cond. _____ Diss. Oxygen _____ Turb. _____
Depth of Water _____ Depth at which Sample Taken _____
Ice Thickness _____
Other _____
Remarks _____
_____

INSTRUMENT CALIBRATION
Diss. Oxygen Meter Model _____ Winkler Calibration _____ mg/L
Meter Reading before Adjustment _____
Conductivity Meter Model _____
pH Meter Model _____ Calibration Buffers Used _____
Remarks _____
_____

WATER QUANTITY MEASUREMENT DATA
Location Description _____
_____
Description of Gauge _____
Stage Height _____
Time _____

SAMPLING APPARATUS USED AND PROCEDURES
_____
_____
_____
_____
_____

SAMPLE SPECIFICS

	Container Material	Volume Collected	Preparation	Quality Control
Major Ions				
Metals				
Organics				
Pesticides & Herbicides				
Mercury				
Phenols				
Nutrients				
BOD & COD				
Others				

Figure 3. General format for a field sampling sheet.

QUALITY CONTROL REMARKS _____

_____

_____

GENERAL REMARKS _____

_____

_____

MODE OF TRANSPORT _____

SHIPPING DATE _____

Figure 3. Continued

Support data are important in determining the validity of the measurement program data and should be used, for example, in deciding whether or not an outlier is valid or is an artifact.

Ideally, unusual conditions should be recorded on the field data report sheets. Failure to report such circumstances can lead to rejection or misinterpretation of valuable data.

12.3 References

Whitlow, S., and M. Lamb. 1983. NAQUADAT Guide to Interactive Retrieval. Water Quality Branch, Inland Waters Directorate, Environment Canada, Ottawa, Ont.

13.0 CHAIN OF CUSTODY

Complete records must be kept of every transfer of data or samples to an individual, laboratory, or storage facility. Such records will permit an investigator to determine who had custody of the material and where it was at any given time. Custody documentation is part of the support data and should be available for review by an independent auditor.

Transfer documents must be updated and maintained at a central location, such as the office of the program or project manager. These records are of great importance in establishing the validity of any questionable data.

Chain of custody procedures must assure that:

- only authorized personnel handle the sample;
- only the field sampling techniques specified for the measurement program are used;
- a record tag is attached to the sample immediately after collection and includes the following information:
  - program or project identification;
  - sample field number;
  - location;
  - depth;
  - collection date;
  - time; and
  - collector.
- blank samples with and without added chemicals (preservatives) are interspersed with the actual samples;
- all record forms are completed;
- the transfer of samples is documented; and
- transfer procedures provide for proper protection and preservation (e.g., if samples are mailed, they should be sent by certified mail with a request for an acknowledgment of receipt).

14.0 FIELD SAFETY

Safety procedures and guidelines affecting every phase of the sampling operation not only must be documented, but must be enforced if accidents and hazardous conditions are to be avoided.

All field personnel must have the Field Safety Manual (Environment Canada 1983a), which gives the procedures and precautions to be observed during the collection of water samples, and personnel must be provided with the appropriate safety equipment (e.g., protection hat, boots, float-jackets, security belt). A knowledge of the provisions of the Canada Labour Code, Part IV, as it applies to safety during field operations is very important.

Brochures, documents, and books relating to field safety, laboratory safety, occupational health and safety, hazardous chemicals, and transportation of dangerous goods should be made readily accessible and available to all field personnel. A copy of each document should be kept in the sampling vehicle or the mobile laboratory.

The following are some of the precautionary measures that must be taken during sampling:

(1) The use of a circular flasher is advised when sampling is performed from bridges where traffic interference presents a safety hazard.

(2) Areas where power lines are strung close to bridges should be avoided.

(3) Ice thickness should be tested before the commencement of sampling in order to avoid falls through the ice.

(4) When wading in streams in which unexpected large holes and swift currents may be encountered and footing may become unsafe, the investigator should use a wading rod and be attached to a rope that is safely anchored.

(5) All field personnel should be able to swim as a safeguard against drowning. Life jackets are necessary equipment.

(6) Boats should be given a thorough servicing prior to the sampling trip. When sampling for an extended period of time in remote areas, personnel should take along an extra motor and emergency rations.

(7) Light airplanes used for transportation to remote areas and helicopters from which sampling is conducted should also be subject to thorough servicing and evaluations for safety before the commencement of a sampling trip.

(8) The handling of chemicals requires care. Inhalation of chemical vapours or direct contact of chemicals with the skin, eyes, and clothing must always be avoided. Acid and base

preservatives or organic extractants must always be carefully stored when not in use. Any spillage of acids or bases should be cleaned up immediately by dilution with large quantities of water, neutralization, or careful mopping up of the chemical solution followed by disposal of the contaminated material. Spilled organic solvents (e.g., n-hexane and dichloromethane) should be sucked up with a mechanical or automatic suction pump and the waste stored for disposal.

(9) Rubber gloves, safety eyeglasses (goggles), and face masks should be worn by personnel involved in cleanup operations related to chemical spills. Contact of chemicals with skin or eyes should be remedied immediately by washing affected areas with plenty of water, followed by an application of a neutralizing solution (in the case of an accident with a base or acid). Eye or skin injuries should be subsequently treated by a physician.

(10) Oral pipetting of solutions is forbidden.

(11) Routine care must be taken to prevent glassware breakage during sampling, packing, unpacking, and storage in subzero weather.

(12) Clothing should be appropriate to the season and weather conditions. Survival kits and rations should be taken along on a sampling trip, especially on extended trips and trips in remote areas. The Sampling for Water Quality manual (Environment Canada 1983b) lists some of the items and articles that would comprise a survival kit and rations.

(13) A first-aid kit should always be included among the supplies and equipment that are taken on the sampling trip. The Canada Labour Code, Part IV, lists items that the first-aid kit should contain for various types of fieldwork.

(14) At least two persons should travel together to collect samples during periods involving long sampling trips, sampling in wilderness and remote areas, or sampling in hazardous winter and flood conditions.

In addition to the health and safety regulations that are issued by the Treasury Board Secretariat (1982), Labour Canada (1986), Transport Canada (1980), and Consumer and Corporate Affairs Canada (1981), a number of informational texts on health and safety practices in the workplace (field and laboratory) can be obtained from provincial and territorial agencies.

14.1 References

Consumer and Corporate Affairs Canada. 1981. Hazardous Products Act and Regulations. Ottawa, Ont.

Environment Canada. 1983a. Field Safety Manual. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

Environment Canada. 1983b. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ontario.

Labour Canada. 1986. Canada Occupational Safety and Health Regulations, SOR/86/304. Canada Gazette Part II, p. 1105.

Transport Canada. 1980. Transportation of Dangerous Goods Act and Regulations. Ottawa, Ont.

Treasury Board Secretariat. 1982. Personnel Management Manual. Personnel Policy Branch, BT-32-19-1/1982, Vol. 12.

15.0 FIELD AUDIT PROGRAM

Performance and systems audits must be instituted and should become stable and continuing features in the field QA/QC program.

The systems audit should consist of a review of the total data production process, which includes on-site reviews of the field's operational systems, the physical facilities for sampling, sample handling, storage, and transportation, and the measurement protocols.

The audit system should be used as one of the channels for detecting, flagging, and correcting errors or defects at any point in the field QA/QC process.

Guidelines in the audit program should specify who will conduct the audit, what protocols and procedures will be used, and to whom the audit reports will go. Garfield (1984) and Freeberg (1980) proposed a number of procedures for conducting an audit program for laboratory operations. These procedures and guidelines are easily adaptable to the field sampling and testing operations, as there is a similarity in the mechanics of both types of program (e.g., constructing the audit team, planning prior to the audit, developing audit methods, subjective and objective measurement procedures, evaluating and assessing the audit fundings, using the audit report to effect change).

15.1 References

Freeberg, F.E. 1980. Optimizing chemical laboratory performance through the application of quality assurance principles. *In* Meaningful Quality Assurance Program for the Chemical Laboratory. Edited by F.M. Garfield *et al.* Association of Official Analytical Chemists, Arlington, Va., pp. 13–23.

Garfield, F.M. 1984. Quality Assurance Principles for Analytical Laboratories. Association of Official Analytical Chemists, Arlington, Va.

16.0 QA/QC TECHNICAL AND ADVISORY COMMITTEE

The QA/QC Technical and Advisory Committee should be established for water quality monitoring programs. Its technical and advisory support functions should fall into four broad categories:

- assisting the monitoring program with the most appropriate, up-to-date, and effective technical advice, to aid in the production of reliable data;
- developing and implementing a system for continually evaluating the performance of field sampling and testing, laboratory analysis, and data management operations and evaluating the personnel who are involved in these operations;
- regularly evaluating the adequacy of operational methods, instruments, and equipment being used; and
- enhancing the overall capability and performance of sampling and laboratory analytical activities through corrective actions and training.

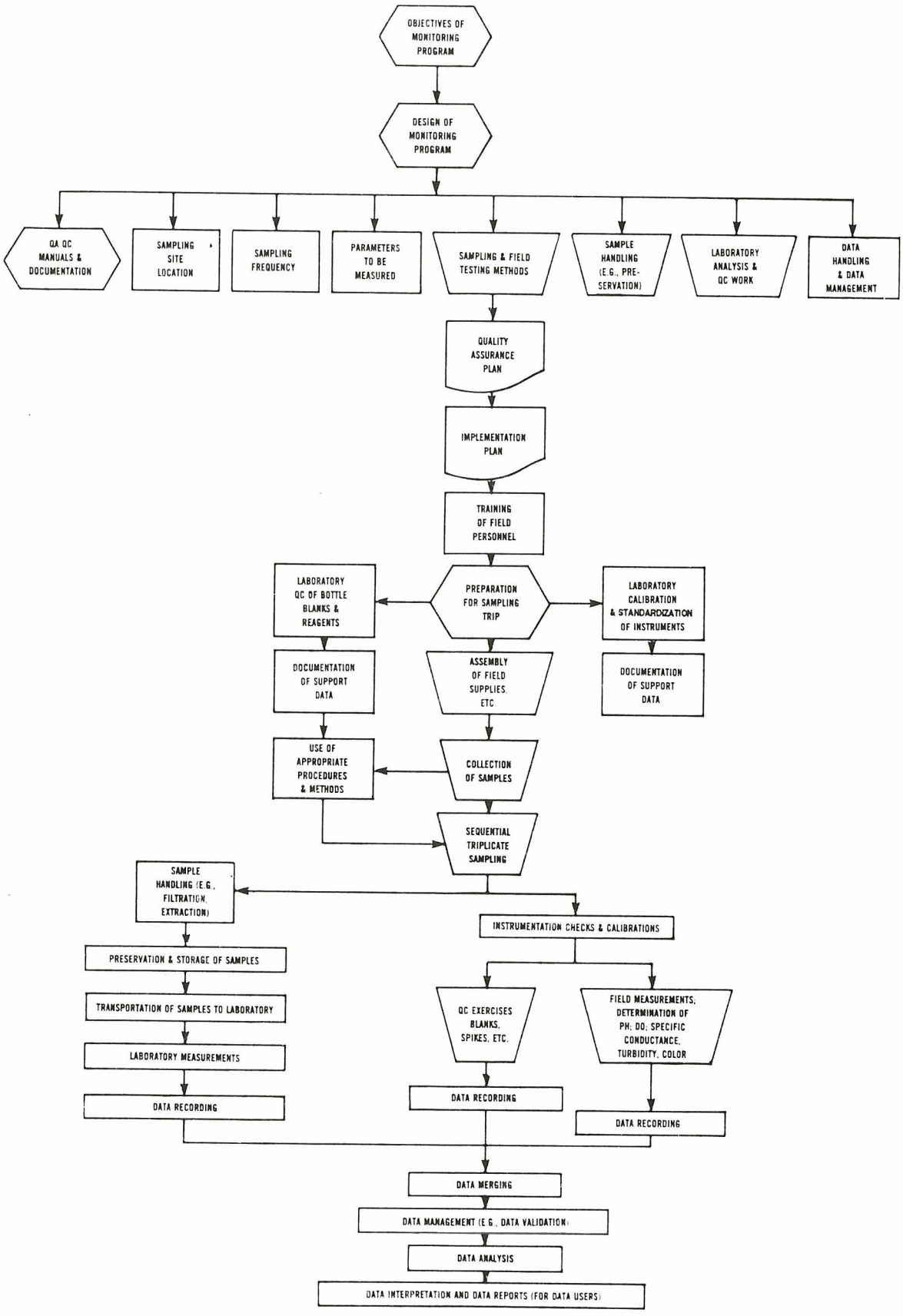
On a more definitive basis, the activities of the QA/QC Technical and Advisory Committee may be associated with the following terms of reference:

- to advise on, and recommend measures to ensure the fulfillment of, the QA/QC objectives of water quality monitoring programs;
- to assist in improving regional QA/QC programs by identifying quality defects in procedures and methodologies through the audit program and by suggesting or recommending appropriate corrective actions;
- to ensure that personnel within the WQB are properly trained and able to carry out their QA/QC responsibilities;
- to ensure that documentation of QA/QC guidelines, principles, and protocols is effected and implemented by all participating groups;
- to provide information, advice, and recommendations to operational managers within the WQB concerning all aspects of its QA/QC programs;
- to provide technical assistance and coordination in communicating the WQB QA/QC protocols to collaborating provincial and U.S. agencies, and to assist them in implementing identical or compatible methodologies so that the groundwork for the generation of comparable and compatible data can be established;

- to assist in the implementation and monitoring of the performance of QA/QC programs, so that collected data meet established standards of precision, accuracy, and completeness; and

- to organize periodic QA/QC workshops and advanced seminars for WQB personnel, and to identify suitable training courses in interdisciplinary areas of water quality monitoring and water quality technology.

Finally, the QA/QC Technical and Advisory Committee should be responsible for ensuring that the guidance and services provided are used effectively in all measurement activities.



Appendix 1. Flowchart of QA/QC activities for water quality sampling (the arrows serve only to outline the general sequential steps).

Part III

QUALITY ASSURANCE IN THE NATIONAL WATER QUALITY LABORATORY

Haig Agemian

Contents

Page

FOREWORD viii

PREFACE ix

ACKNOWLEDGMENTS x

1.0 INTRODUCTION III-1

2.0 INTRALABORATORY QUALITY ASSURANCE PLAN III-3

3.0 DEFINITIONS OF ANALYTICAL TERMINOLOGY III-7
- 3.1 Quality assurance, quality control, and quality assessment III-7
- 3.2 Precision III-7
- 3.3 Accuracy III-8
- 3.4 Bias III-9
- 3.5 Sensitivity III-9
- 3.6 Confidence interval III-10
- 3.7 Limit of detection III-10
  - 3.7.1 Instrument detection limit III-11
  - 3.7.2 Method detection limit III-11
  - 3.7.3 Practical detection limit III-12
- 3.8 Limit of quantitation III-12
- 3.9 Reporting of analytical results III-13
- 3.10 References III-13

4.0 PROTOCOLS III-15
- 4.1 Protocol for rejection or reporting of analytical data below the level of quantitation III-15
- 4.2 Protocol for validation of analytical methods III-15
- 4.3 Protocol for approval of new methodology III-16
- 4.4 Protocol for assignment of new parameter code to a method III-16
  - 4.4.1 Substrate or matrix III-17
  - 4.4.2 Concentration range III-17
  - 4.4.3 Instrumental technique III-17
  - 4.4.4 Confidence level/interval III-17
  - 4.4.5 Cleanup III-17
  - 4.4.6 Derivatization III-17
- 4.5 Data validation protocol III-17

Contents (Continued)

Page

5.0 QUALITY CONTROL PROCEDURES III-19
5.1 Intralaboratory quality control procedures III-19
5.2 Interlaboratory quality control III-20
5.3 Standard additions method III-21
5.4 Quality control charts III-22
5.5 Calibration curves by least squares method III-27
5.6 References III-29

6.0 QUALITY CONTROL PRACTICES IN THE NATIONAL WATER QUALITY LABORATORY III-30
6.1 Sample Receiving Facility III-30
6.2 Bottle Preparation Laboratory III-31
6.3 Major Ions, Nutrients, and Physical Parameters Laboratory III-32
6.3.1 Operational considerations III-32
6.3.1.1 Deionized distilled water blanks III-32
6.3.1.2 Analytical standard preparation III-32
6.3.1.3 Calibration procedures III-33
6.3.1.4 Precautions to check quality of glassware III-33
6.3.1.5 Use of control samples III-33
6.3.1.6 Reproducibility and performance checks III-33
6.3.1.7 Calculation and data management III-33
6.3.2 Precision, accuracy, and bias considerations III-34
6.3.2.1 Calculation of standard deviation III-34
6.3.2.2 Recheck analysis III-34
6.3.2.3 Analysis of reference sample and spike recoveries III-34
6.3.3 Controls to detect the occurrence of unexpected observations . . . III-34
6.3.3.1 Calibration curves III-34
6.3.3.2 Reproducibility within and between runs III-34
6.3.3.3 Accepted limit for replicate III-34
6.3.3.4 Accepted limit for the spike recovery III-34
6.3.3.5 Analysis of control samples III-35
6.4 Atomic Spectroscopy Laboratory III-35
6.4.1 Operational considerations III-35
6.4.1.1 Deionized distilled water blanks III-35
6.4.1.2 Reagent blanks III-35
6.4.1.3 Analytical standard preparation III-35
6.4.1.4 Calibration procedures III-35
6.4.1.5 Precautions to check quality of glassware III-36
6.4.1.6 Use of reference samples III-36

Contents (Continued)

Page

6.4.1.7 Reproducibility and performance checks III-36
6.4.1.8 Calculation and data management III-36
6.4.2 Precision, accuracy, and bias considerations III-36
6.4.2.1 Calculation of standard deviation III-36
6.4.2.2 Within-run and between-run standard deviation III-36
6.4.2.3 Calibration rerun III-37
6.4.2.4 Replicate analysis III-37
6.4.2.5 Analysis of reference sample and spike recoveries III-37
6.4.3 Controls to detect the occurrence of unexpected observations . . . III-37
6.4.3.1 Calibration curves III-37
6.4.3.2 Reproducibility within and between runs III-37
6.4.3.3 Accepted limits for replicates III-37
6.4.3.4 Accepted limits for the spike recovery III-37
6.4.3.5 Analysis of control samples III-37
6.4.3.6 Certified reference materials III-38
6.5 Organic Analysis Laboratory III-38
6.5.1 Operational considerations III-38
6.5.1.1 Reagent purity III-38
6.5.1.2 Cleaning of glassware III-38
6.5.1.3 Analytical reference standards III-38
6.5.1.4 Calibration procedures III-38
6.5.2 Precision, accuracy, and bias considerations III-39
6.5.2.1 Replicates III-39
6.5.2.2 Recovery studies (spiked samples) III-39
6.5.2.3 Calculation of standard deviation III-39
6.5.2.4 Within-run and between-run precision III-39
6.5.2.5 Quality control check standards III-40
6.5.2.6 Confirmation techniques III-40
6.5.3 Record keeping in the Organic Analysis Laboratory III-40
6.5.4 Reporting III-40
6.6 Gas Chromatography/Mass Spectrometry Laboratory III-41
6.6.1 System applications III-41
6.6.2 System operational characteristics III-41
6.6.3 System performance III-41
6.6.4 Applications criteria III-42
6.6.4.1 Qualitative confirmation III-42
6.6.4.2 Quantitative confirmation III-42
6.6.4.3 Qualitative identification of unknown materials III-43

Contents (Continued)

Page

6.6.4.4 Quantitative analysis of target organic contaminants . . . III-43
6.6.4.5 Screening and characterization for unknown contaminants III-43
6.6.5 Target compound analysis performance records III-44
6.6.6 System setup procedure III-44
6.6.6.1 INCOS data system III-44
6.6.6.2 Shrader data system III-45
6.7 Automated Water Quality Laboratory Data Acquisition and Management System (AWQUALABS) III-45
6.7.1 Present activities III-46
6.7.1.1 Sample initialization III-46
6.7.1.2 Sample analysis III-46
6.7.1.3 Data entry III-47
6.7.1.4 Data verification III-47
6.7.2 Future development III-48
6.7.2.1 General limits screening III-48
6.7.2.2 Site-specific limits screening III-48
6.7.2.3 Chromatographic decision matrix III-48
6.7.2.4 Laboratory performance III-49
6.8 References III-49

7.0 GUIDELINES FOR GOOD LABORATORY PRACTICE III-50
7.1 Laboratory facilities and safety III-50
7.1.1 Facilities and safety equipment III-50
7.1.2 Safety practices III-51
7.2 Staff relations and work environment III-52
7.3 Chemicals, reagents, and standards III-53
7.4 Apparatus III-54
7.5 Methodology III-55
7.6 Data recording, reporting, and documenting III-56
7.7 References III-56

8.0 GUIDELINES FOR INSTRUMENT PERFORMANCE CONTROL CHECKS III-58

9.0 QUALITY ASSESSMENT FORMS III-70

Tables

		Page
1.	Atomic spectroscopy performance control checks	III-59
2.	Molecular spectroscopy performance control checks	III-63
3.	Chromatography performance control checks	III-64
4.	Mass spectrometry performance control checks	III-67
5.	Miscellaneous instrument performance control checks	III-69

Figures

		Page
1.	Intralaboratory quality assurance	III-6
2.	Standard additions curve	III-23
3.	Accuracy control chart ($\bar{x}$-chart)	III-24
4.	Precision control chart (R-chart)	III-25

Foreword

The need for quality assurance (QA) has long been recognized by the Water Quality Branch (WQB). Activities undertaken by the WQB have increased the demand for measurement data and hence augmented the need for a comprehensive QA program. Quality Assurance in the National Water Quality Laboratory describes the steps that must be taken to ensure that all measurement and data processing activities that affect the quality of the data produced in a modern laboratory satisfy pre-established standards. This document is intended as a reference manual for laboratory staff and data users.

Preface

During the last decade, there has been a growing awareness of the impact of toxic chemicals, at trace and ultratrace levels, on the environment. Social pressures have created a demand for elucidation of the effects of chemicals on the ecosystem and on humans. Increased emphasis is also being placed on how chemicals migrate in the ecosystem. Analytical data are continually being released by many laboratories, and increased method sensitivity and selectivity are being sought to determine toxic chemicals at ultratrace levels.

Recently, many agencies have shown a real concern about the validity of analytical data generated at ultratrace levels. The main reason for such concern is believed to be methodology differences and lack of documentation of procedures and quality control (QC) practices used in various laboratories.

In the field of environmental analysis, QC and quality assurance (QA) are being given greater attention by laboratory managers and users of data. Although there are a number of excellent monographs, reviews, and manuals on QC in environmental monitoring and analysis, it is nevertheless essential for each laboratory engaged in the generation of data to document the internal QC practices and protocols that comprise its QA program. This would assist analysts in the generation of high-quality data as well as substantiate data quality to laboratory clients and data users. In addition, it is recognized that quality assessment activities, whereby the effectiveness of QC practices is determined, are an integral part of a successful QA program. This is achieved by the proper documentation of QC data on Quality Assessment Forms and their systematic evaluation.

The National Water Quality Laboratory (NWQL) of the Inland Waters Directorate of Environment Canada has developed a QA document for analysts within the laboratory. It is also aimed at providing pertinent information to clients to indicate what precautions are taken to ensure proper calibration and standardization of instruments and analytical systems. It also describes the routine practices that are being implemented by analysts to ensure that data are generated with well-defined confidence levels and intervals.

Application of these QC practices within NWQL is a step towards a dynamic system of process deviation control. It is anticipated that by implementing these practices, systematic errors will be minimized. The topics covered in this document include definitions of analytical terminology, protocols, QC procedures and practices, guidelines for good laboratory practice, guidelines for instrument performance control checks, and Quality Assessment Forms. By following these standard operating QC procedures, it will be possible to identify problems, improve data quality, and design anticipatory actions and options to improve proficiency of analysts and the overall performance of the laboratory.

Acknowledgments

Thanks are due to B.K. Afghan for his valuable suggestions, J. Gaskin for his extensive editorial comments, A. Demayo and E. Watt for contributions towards Chapter 2, and the following for their contributions to Chapters 5, 7, and 8: J. Carron, O. El Kei, B. Francoeur, M.A. Forbes, G.H. Jamro, R.C.J. Sampson, Y. Sheik, and D. Sturtevant.

Quality Assurance in the National Water Quality Laboratory

Haig Agemian

1.0 INTRODUCTION

The National Water Quality Laboratory (NWQL) generates analytical data for Environment Canada and other government agencies to establish the trend, fate, and effects of environmental contaminants in the aquatic environment. In some cases, these measurements are used, or misused, to establish the degree of pollution loadings and may form a basis for regulatory policy decisions.

Heightened concerns about water quality and new concerns about the possible effects of toxic chemicals at trace and ultratrace levels have contributed greatly to an increased demand for impartial, objective, and independent data that can be used to assess the impact of these contaminants in the aquatic environment.

Agencies that have been involved in environmental measurements for some time have observed the tremendous increase in analytical sensitivity over the past 10 to 15 years. For some environmental contaminants, detection limits as low as a few femtograms have been reported. At the same time, greater disparities are found between the reported data at very low levels and the ability of techniques or methods to detect and measure contaminants at ultratrace levels. In addition, it is very rare to find reported sensitivities, detection limits, or data associated with degrees of uncertainty (confidence levels and intervals). In many cases, the numbers reported by scientists or laboratories are used as absolute values by other scientists and government agencies to predict loadings and effects, and to make regulatory and control decisions. Often these numbers find their way into the media and lend to increased public concerns. Sometimes people are unnecessarily alarmed by the apparently increasing deterioration of environmental quality without any idea that the assessments and statements are based on data that may have a large degree of uncertainty.

In view of the above, it is essential that scientists who are skilled in making these measurements, at trace and ultratrace levels for particularly toxic chemicals, provide valid analytical data with well-defined confidence levels and intervals to facilitate the objective interpretation for regulatory policy decisions.

What are valid, reliable, and quality data, and how can they be assured? There is no single definition that will encompass everyone's understanding of quality. The dictionary defines quality as "excellence, the degree of excellence, relative goodness or grade." In testing laboratories such as the NWQL, quality can be interpreted as the ability of the laboratory to generate accurate analytical data with well-defined total error. Total error in the laboratory can be defined as the sum of random and

systematic errors that can occur during the acquisition, processing, handling, analyzing, and reporting of analytical data.

In order to ensure data quality, it is very important that the persons involved in generating these data are knowledgeable and fully committed to the production of high-quality data. This is achieved by carefully following well-established operating procedures and practices and minimizing errors during analysis, data processing and verification, and data reporting.

The purpose of this document is to develop standard operating procedures for those activities of the laboratory that affect data quality. The main objectives are to:

- provide laboratory personnel with quality control (QC) procedures and practices that will minimize errors from the time of receipt of sample to the reporting of results;

- assess the quality and overall performance of the NWQL. By following these QC practices, it will be possible to ascertain that the personnel carrying out various activities within the laboratory are taking appropriate precautions during analysis to ensure that the equipment is properly standardized, calibrated, and functioning in accordance with specific and well-defined limits, that methods are performing within the accepted tolerance levels, and that data verification is undertaken prior to reporting;

- quantitatively determine single operator/multiple operator and overall precision for various determinands and provide confidence intervals and range of variability of analytical results produced within NWQL;

- institute a quality assessment program by the periodic evaluation of the effectiveness of QC practices; and

- provide an information document to clients so they can ascertain the degree of quality in all measurements being carried out within NWQL. It is anticipated that this will result in increased client confidence.

2.0 INTRALABORATORY QUALITY ASSURANCE PLAN

Quality management involves the activities of senior laboratory managers pertaining to the development and implementation of policies and plans to ensure effective and successful quality assurance (QA) programs. Quality management is implemented by the coordination and harmonization of various components of QA to ensure optimized and fine-tuned operations and to provide high-quality data within available human and financial resources.

QA is the overall verification program that provides producers and users of data the assurance that predefined standards of quality at predetermined levels of confidence are met. The objectives of a comprehensive QA program are to:

- establish policies and protocols on laboratory QC;
- document QA methodology;
- standardize data QC;
- provide guidelines for good laboratory practices;
- establish a quantitative approach to determine single operator/multiple operator and overall precision and confidence intervals of analytical results;
- make available data quality information documents for clients and data users;
- implement a mechanism for auditing laboratory operations;
- establish a framework for high-calibre analytical practices; and
- provide QC statements to support analytical practices.

These objectives can normally be fulfilled by the implementation of a comprehensive QA program that involves the adoption of four major activities, in the following order:

- Step I — operational and scientific protocols;
- Step II — guidelines on good laboratory practices;
- Step III — QC procedures and practices; and
- Step IV — quality assessment activities.

The first step includes the implementation of QC protocols of an operational and scientific nature. The purpose of these protocols is to establish standardization and conformity of activities throughout the analytical process from receipt of samples at the laboratory to final data reporting to clients. The implementation of these rules ensures that objective performance is maintained in all processes.

The second step in the in-house QA program is the adoption and ongoing application of a set of guidelines called Good Laboratory Practices. These are aimed at providing up-to-standard laboratory facilities with respect to safety and working environment, staff relations, and analytical proficiency, quality of chemicals, reagents, and standards, performance of analytical instrumentation and apparatus, standardization, adaptation, validation, documentation of analytical methodology, data recording, reporting, and documentation.

In the optimization and control of an analytical activity, one must address every variable that may affect data quality. All analytical processes have a large series of factors that must be constantly fine-tuned and controlled. These are normally classified into four main groups, as follows:

(1) *Materials*—standards, chemicals, reagents, control samples, standard reference materials (SRMs), consumables, etc.

(2) *Machines*—instrument meters, recorders, detectors, laboratory equipment, etc.

(3) *Methods*—preservation techniques, sample preparation procedures, analytical methodologies, data management, and interpretation protocols.

(4) *Men and women*—management, supervisory, scientific, and technical staff.

Each one of these elements should be optimized and controlled on an ongoing basis to ensure effective operations. The "quality" activities that are engaged to enhance the quality of the above groups of factors are the following:

(1) *Materials*—purification, standardization, intercomparison.

(2) *Machines*—maintenance, troubleshooting, optimization, and calibration.

(3) *Methods*—adaptation, validation, standardization, specification testing, and documentation.

(4) *Men and women*—training, good laboratory practices, performance assessment, safety procedures, adherence to protocols and guidelines.

When both QA steps I and II are implemented and the four categories of variables are fine-tuned, then an analytical laboratory is said to have the fundamentals to monitor and evaluate the

quality of its product "data." Steps I and II ensure that all variables that affect data quality are optimized.

Step III, the implementation of QC procedures and practices, is designed to provide a quantitative measurement of the quality of the data by the analysis of a set of QC samples together with every analytical run and the interpretation of these data to ensure that the process in question is in control. Normally, up to five types of QC sample tests may be run for a given batch of samples based on the analytical requirements. They are as follows:

- blanks;
- within-run or between-run duplicates;
- surrogate or internal standards;
- spikes; and
- reference samples (in-house controls or SRMs).

These samples are called QC samples and are analyzed for each set of unknown samples. The QC data for each batch of samples are normally evaluated to provide information on freedom from contamination (blanks), precision (between-run and within-run duplicates), recovery (spikes), and accuracy (control samples and SRMs). These data can be evaluated using statistical and graphical control charting techniques (statistical process control) to provide long-term, single operator/multiple operator performance data as well as confidence limits for analytical methods. The data may be used by laboratory analysts, supervisors, and managers as well as by data users and clients to provide QC substantiation of analytical data.

Step IV of the QA program, quality assessment, consists of the activities predominantly concerned with ensuring that the QC step is carried out effectively. This final step involves evaluating and auditing QC data and taking corrective action on an ongoing basis to fine-tune QA activities and ensure a successful program.

Figure 1 shows a schematic of the interrelationship of these four major QA steps. A comprehensive intralaboratory QA program is said to be in place only when all of these areas are well coordinated with the required checks and balances to harmonize and fine-tune the QA activity. The result of a comprehensive QA program is a quality product, namely quality data, and a satisfied client.

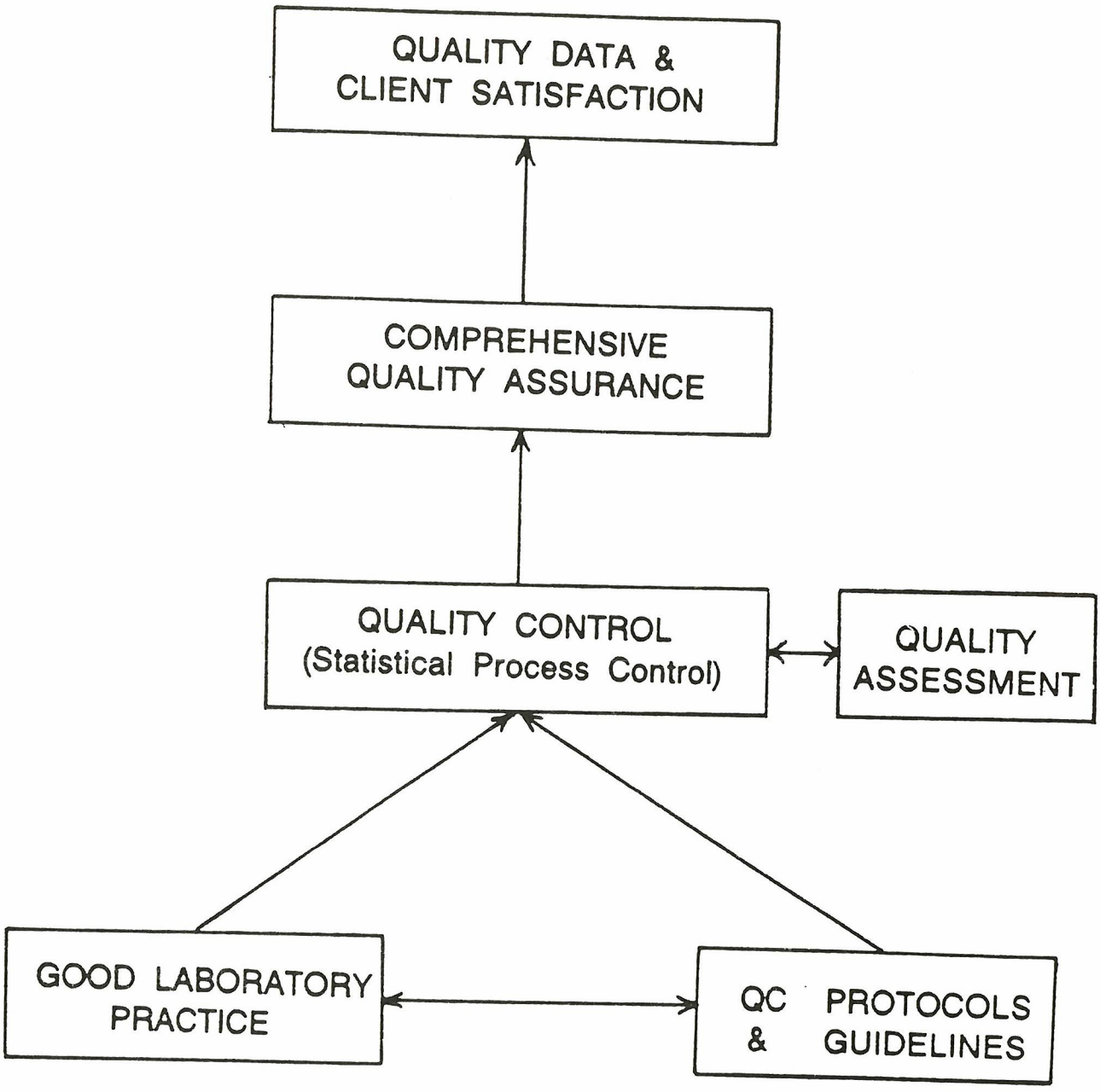


Figure 1. Interlaboratory quality assurance.

3.0 DEFINITIONS OF ANALYTICAL TERMINOLOGY

The following definitions have been adopted from currently available publications, which are listed in Section 3.10.

3.1 Quality Assurance, Quality Control, and Quality Assessment

Quality assurance (QA) is the overall verification program that provides producers and users of data the assurance that predefined standards of quality at predetermined levels of confidence are met.

Quality control (QC) is the overall system of guidelines, procedures, and practices that are designed to regulate and control the quality of products or services with regards to previously established performance criteria and standards.

Quality assessment is the overall system of activities that ensure that QC is being performed effectively. This is carried out immediately following QC and involves evaluation and auditing of QC data to ensure the success of the QC program (see Chapter 9).

3.2 Precision

Precision or *reproducibility* describes the degree of closeness or agreement between the data generated from replicate or repetitive measurements by applying the same experimental procedure several times under prescribed conditions. Statistically, the concept is referred to as dispersion, and it measures the variability of the analytical method resulting from random errors. Precision is generally reported as the standard deviation (SD) or relative standard deviation (RSD). The precision of an analytical method has two components that should be recorded for each method — within-run and between-run precision.

Within-run precision (SD_W) measures the random error during processing of a single batch of samples analyzed at the same time.

Between-run precision (SD_B) measures the variability between various batches of samples analyzed.

The precision of a method is ordinarily measured by the standard deviation (SD):

$$SD = \sqrt{\sum_i \frac{(X_i - \bar{X})^2}{n-1}} \qquad (1)$$

where: n is the number of replicate results of the same sample;
$\bar{X}$ is the mean of n determinations; and
X_i is the value of the ith determination.

Where SD is independent of concentration (i.e., SD is constant for a given range of concentration), results of duplicate analyses of different samples may also be used to calculate the precision. In such a case:

$$SD = \sqrt{\sum_i \frac{(x_i - y_i)^2}{2n}} \qquad (2)$$

where x_i and y_i are n paired sets of duplicate analysis results of different samples.

Often precision is expressed relative to the concentration level at which it was determined. This term is called *% relative standard deviation* (% RSD) or *coefficient of variation*. It is calculated as follows:

$$\% \text{ RSD} = \frac{SD \times 100\%}{\bar{x}} \qquad (3)$$

3.3 Accuracy

Accuracy refers to the correctness of the data and defines the degree of agreement of the measurements with the true value of the magnitude of the quantity concerned.

The accuracy of a determination is expressed as *% error* as follows:

$$\% \text{ error} = \frac{\bar{X} - X_{SRM}}{X_{SRM}} \times 100\% \qquad (4)$$

where: $\bar{X}$ is the mean value of replicate determinations of a standard reference material (SRM); and
X_{SRM} is the certified value of the SRM.

In the absence of SRMs, it is possible to estimate the accuracy by determining the *spike recovery* of a given parameter. During method validation, spike recovery is established by spiking the analyte into an appropriate matrix at a minimum of three levels that span the range of interest. Because this procedure introduces the analyte in the soluble form, the recovery does not account for extraction efficiency from solid matrices such as sediment and biological tissue or from suspended and colloidal matter in liquid samples. Recovery is calculated on a percentage basis as follows:

$$\% \text{ recovery} = \frac{C_{found} - \overline{C}_{blank}}{C_{added}} \times 100 \quad (5)$$

Where: C_{found} is the measured concentration in the sample;
$\overline{C}_{blank}$ is the average concentration of the blank; and
C_{added} is the known concentration added to the sample.

For each method, the recovery should be reported as the mean, range, and standard deviation. For example:

	Mean recovery (%)	Recovery range (%)	SD (%)
Pyrene–GC	63	26–129	40

Recovery is monitored by spiking at least 10% of the actual samples and calculating by the equation:

$$\% \text{ recovery} = \frac{C_{(sample + spike)} - \overline{C}_{sample}}{C_{added}} \times 100 \quad (6)$$

where: $C_{(sample + spike)}$ is the measured concentration in the spiked sample;
$\overline{C}_{sample}$ is the average concentration of the unspiked sample; and
C_{added} is the known concentration added to the sample.

Inaccuracy results from imprecision (random error) and bias (systematic error) in the measurement process.

3.4 Bias

Bias is a systematic deviation of the average of one set of data from another.

3.5 Sensitivity

Sensitivity is a measure of instrumental response factor as a function of concentration. It is commonly measured as the slope of the calibration curve.

Given a linear calibration curve of the form $R = SC + I$ and an intercept I on the ordinate R, the sensitivity (S, slope) may be expressed as follows:

$$S = \frac{R}{C} \quad (7)$$

where: R is the signal response; and
C is the concentration.

3.6 Confidence Interval

A *confidence interval* is an expression of the range of values defined by upper and lower limits at a statistically defined confidence level.

The confidence interval for a sample mean is given by:

$$\bar{X} \pm \frac{t_{(n-1)}SD}{\sqrt{n}} \qquad (8)$$

where: $\bar{X}$ is the sample mean;
SD is the standard deviation;
n is the sample size; and
$t_{(n-1)}$ is the two-sided Student's t value with (n−1) degrees of freedom for the specified level of confidence.

3.7 Limit of Detection

The *limit of detection* (LOD) is one of the most important terms used for comparison between various analytical procedures, techniques, or instruments. It is defined as being the lowest concentration of analyte that can be distinguished with reasonable confidence from blank or background. In Water Quality Branch (WQB) laboratories, the confidence level of 95% is adopted as a standard for LOD for all analytes.

Different statistical approaches are reported in the literature to calculate LODs. The LOD for a determinand can easily vary an order of magnitude through the use of different statistical approaches. This problem is well documented in the literature (Long and Winefordner 1983).

Recently, detection limits for selected parameters analyzed within WQB laboratories were calculated using different statistical approaches. The study covered a range of analytes, including major ions, nutrients, trace metals, and trace organics. Analytical techniques employed included titrimetry, colorimetry, atomic spectroscopy, and gas chromatography. The statistical approaches for calculation of LODs included IUPAC model, graphic approach, and propagation of error approach. The results of the above study confirmed that significant variation in detection limits can occur depending upon the approach employed. IUPAC model consistently gave the lowest values, whereas the largest values were obtained using propagation of error approach.

A simplified version of IUPAC, described below, is selected as a standard for calculating detection limits for all analytes determined within the WQB. It is fully recognized that this approach may give the lowest values; however, the proposed approach is simple to apply and will ensure meaningful comparison and objective assessment of the sensitivity of the techniques, procedures, and methods used within the WQB.

The LODs are represented in three different forms depending upon the requirements — instrument detection limit (IDL), method detection limit (MDL), and practical detection limit (PDL).

3.7.1 Instrument Detection Limit

The *instrument detection limit* (IDL) is the lowest concentration of analyte that an analytical instrument can detect and which is statistically different from the response obtained from the background instrumental noise.

The IDL is established by adding the analyte in reagent (blank) water or appropriate organic solvent to give a final concentration within five times the estimated IDL and calculating standard deviation (SD) by introducing the solution directly into the instrumental system to obtain seven or more replicate measurements. The IDL is then calculated using the 95% confidence level, as follows:

$$IDL = t_{(n-1)} \times SD \text{ for } n = 7 \text{ or more} \qquad (9)$$

where $t_{(n-1)}$ is the value for a one-sided Student's t distribution for n−1 degrees of freedom.

The IDL should be used to indicate absolute sensitivity of the analytical technique or instrument.

3.7.2 Method Detection Limit

The *method detection limit* (MDL) is the lowest concentration of analyte in distilled water that a method can detect reliably and which is statistically different from the response obtained from a blank carried through the complete method including chemical extraction or pretreatment of the sample. When the MDL is experimentally evaluated for each matrix by the analysis of samples or spiked samples, then the practical detection limit (PDL) is determined (see Section 3.7.3).

When repeated analyses of blanks show a positive response for the analyte, the MDL is defined as:

$$MDL = \overline{S_b} + t_{(n-1)}SD \qquad (10)$$

where: $\bar{S}_b$ is the average signal (or level) for the blanks;
SD is the standard deviation of the replicate determinations; and
$t_{(n-1)}$ is the one-sided Student's t distribution for n–1 degrees of freedom at a confidence level of 95%.

If the repeated analysis of method blanks does not show a positive response for the analyte, the MDL can be calculated by spiking reagent water or the samples to give a final concentration within five times the estimated IDL and by using standard deviation (n = 7 or more) and 95% confidence level as shown in Equation 10.

For methods requiring concentration and pretreatment, the MDL is determined by selecting a standard sample size (e.g., 1 L in the case of water), volume of final aliquot (e.g., 1 mL prior to analysis by high-pressure liquid chromatography [HPLC], gas chromatography [GC], gas chromatography/mass spectrometry [GC/MS]), and volume required for analysis (e.g., 2 μL for GC).

3.7.3 Practical Detection Limit

The *practical detection limit* (PDL) is the lowest concentration of analyte in a real sample matrix that a method can detect reliably and which is statistically different from the response obtained from a blank carried through the complete method. It is calculated in the same manner as the MDL in Equation 10.

For a specified method and analyte, the PDL will vary with different sample matrices, as these may affect reproducibility, blanks, and interference levels.

The PDL is always equal to or greater than the MDL, but it will never be less than the MDL.

3.8 Limit of Quantitation

The *limit of quantitation* (LOQ) is defined as follows:

$$LOQ = \overline{S_b} + 10\ SD \qquad (11)$$

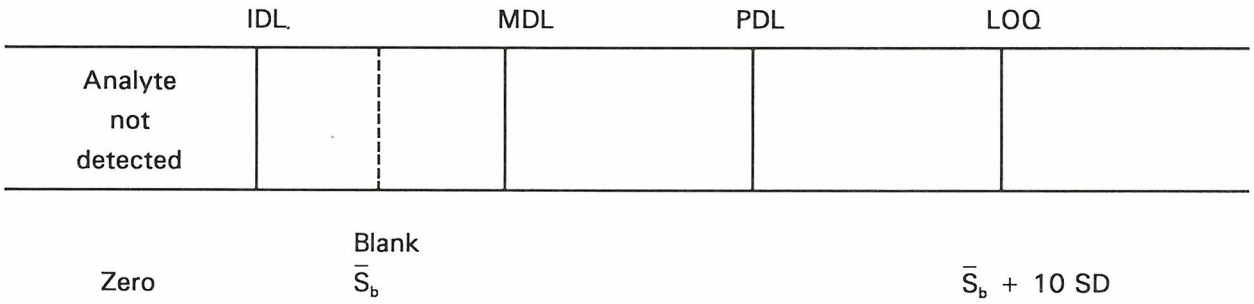
where: $\bar{S}_b$ is the average signal (or level) for the blank; and
SD is the standard deviation of the replicate determinations.

This defines the level above which quantitation is reliable and also a region between MDL and LOQ, where detection is reliable but quantitation is not. The LOQ is the level above which quantitative results may be obtained with a specified degree of confidence.

3.9 Reporting of Analytical Results

In no case should the term 0 (zero) be used to report an analytical result. In all cases for which no analyte was detected, the expressions ND (none detected) and L (less than) should be used to report the result. The value following L is the MDL. If the result falls between the MDL and the PDL, the value may be reported in parentheses, with an explanation that the result is less than the PDL and the recovery and precision have not been evaluated at that level.

Analyte concentration (mg/L)



In some cases, data and analyte recovery or concentration may be reported down to the IDL provided the necessary explanation is accompanied with the results, indicating changes in sample size, final aliquots prior to analysis, or method of quantitation (use of internal standards or surrogates).

3.10 References

American Chemical Society. 1983. Principles of environmental analysis. Anal. Chem. 55: 2210.

Bauer, E.L. 1971. A Statistical Manual for Chemists. 2nd ed. Academic Press, New York, N.Y.

Caulcutt, R., and R. Boddy. 1983. Statistics for Analytical Chemists. Chapman and Hall Ltd., London, U.K.

Crummett, W.B., and J.K. Taylor. 1980. Guidelines for data acquisition and data quality evaluation in environmental chemistry. Anal. Chem. 52: 2242.

Kaley, R.G., II. 1984. Standard Operating Procedure for Establishing and Reporting Performance Characteristics of an Analytical Procedure. Presented at the American Chemical Society's 107th Annual Meeting, April, St. Louis, Mo.

Kirchmer, C.J. 1983. Quality control in water analyses. Environ. Sci. Technol. 17(4): 174A.

Long, G.L., and J.D. Winefordner. 1983. Limit of detection, a closer look at the IUPAC definition. Anal. Chem. 55: 712A.

Taylor, J.K. 1985. The quest for quality assurance. Am. Lab. 1985 (Oct.): 67.

4.0 PROTOCOLS

4.1 Protocol for Rejection or Reporting of Analytical Data Below the Level of Quantitation

By definition, data above the LOQ are statistically well substantiated and may be reported with a high degree of confidence. Data at levels of the PDL, MDL, and IDL may be reported with decreasing levels of confidence, as defined in Chapter 3.

There are two schools of thought on the level at which data should be rejected or reported. The first maintains that because data below the LOQ are not quantifiable, they should be totally rejected. The second argues that concentrations below the LOQ should nevertheless be reported, because they are defined with appropriate degrees of confidence and thus have usefulness. The NWQL has taken the latter position. It is strongly believed that a laboratory is responsible to clients in reporting all detectable analyte concentrations, provided they are well defined with appropriate levels of statistical confidence. It is felt that a laboratory that takes the initiative to censor and eliminate a certain amount of detected data is indeed doing an injustice to data users. These data, irrespective of their level of statistical confidence, may contain valuable environmental information. It is not the mandate of the NWQL to interpret data or to make decisions on their usefulness. The NWQL must endeavor to make every effort to generate high-quality data that are substantiated by appropriate limits of confidence. In this way, the data user would have a valuable product that could be used or not used as dictated by project needs.

4.2 Protocol for Validation of Analytical Methods

The following elements constitute specifications or performance criteria that characterize the capabilities and scope of an analytical method. Method validation is carried out by systematically checking these elements and documenting the results in the validation statement shown below:

(1) Analyte(s)

(2) Substrate(s)

(3) Method summary

(4) Times for analyses (h)
- batch samples
- single samples

(5) Interferences

(6) Range (concentration)

(7) Detection criteria
- instrument detection limit (IDL)
- method detection limit (MDL)
- practical detection limit (PDL)

(8) Sensitivity (slope of calibration curve)

(9) Accuracy (SRMs, spike recovery, or other method compatibility)

(10) Precision (at 20, 50, and 80% of concentration range).

4.3 Protocol for Approval of New Methodology

The NWQL has adopted specified practices to be followed in the event of modification or alteration of existing analytical methodology or the adoption of new methodology. They are as follows:

(1) The new method or change in existing method is submitted to the chemist-in-charge for consideration. The chemist-in-charge reviews the method and ensures that the necessary validation data are available.

(2) The method is then adopted as a "tentative method" to generate precision and accuracy data, using standard and control samples.

(3) The method is then tested on environmental samples for the period of three to six months to ensure that the method covers specified scopes in terms of detection limits and the range and type of samples. The method is then documented in standard format and submitted to the laboratory manager for review.

(4) The laboratory supervisor reviews the method and accompanying data and submits them for adoption as a routine method.

(5) The section head, after appropriate review and peer group evaluation, recommends the method for acceptance as a routine or on-line method.

(6) The final approval is given by the Chief, NWQL.

4.4 Protocol for Assignment of New Parameter Code to a Method

When modifications to existing methods are made, it is necessary to decide whether or not the change constitutes a new method. Methodology differences in one or more of the following elements are considered to constitute a new method and thus require new parameter codes.

4.4.1 Substrate or Matrix

Similar methods that apply to a different substrate or matrix, such as water, sediment, fish, or plants, should be categorized separately and assigned different codes based on their classification.

4.4.2 Concentration Range

Each method should cover a well-defined concentration range, such as 1 to 10 or 10 to 100 units. Therefore, a change in detection limit must constitute a significant change of at least an order of magnitude.

4.4.3 Instrumental Technique

Methods with similar sample preparation steps requiring different instruments for quantitative analysis should be assigned separate codes. Different models of instruments do not require different code numbers unless their operation is based on completely different scientific principles or requires substantial changes to the method.

4.4.4 Confidence Level/Interval

Methods with statistically different confidence levels/intervals should be assigned separate codes.

4.4.5 Cleanup

In many cases, different cleanup procedures are deployed to remove certain interferences. The methods should be assigned a separate code when the cleanup procedures are different.

4.4.6 Derivatization

In some cases, it is essential to derivatize or complex the analyte prior to quantitation. Methods employing different complexing agents or derivatization should be assigned separate codes.

4.5 Data Validation Protocol

The following are the steps taken at the NWQL in validating data prior to release to clients and archiving:

(1) The technician reviews analysis data to ensure blanks, spikes, replicates, and SRMs are within the expected ranges. If a problem is detected, unknowns and QC samples are reanalyzed.

(2) The laboratory unit head (CH-02) reviews all analytical reports, verifies data on all samples, and makes a recommendation to the Laboratory Head (CH-03) for acceptance and reporting.

(3) The Laboratory Head (CH-03) reviews the data, accepts each sample, and prepares the preliminary AWQUALABS sample data report.

(4) Data on NAQUADAT samples are compared against historical ranges to ensure they are reasonable values. Outliers are reanalyzed to confirm analytical results. The Laboratory Head (CH-03) then prepares a final AWQUALABS sample data report and recommends it for reporting to the Section Head (CH-04).

(5) The Section Head (CH-04) checks approximately 5% of the data against historical levels, verifies the QC data, and reports the data to the client.

(6) A record of sample data reports is kept at each level in the hierarchy from technician to Section Head, and a copy of the client data report is kept in the Central Registry office at the Canada Centre for Inland Waters.

5.0 QUALITY CONTROL PROCEDURES

QC procedures are the means by which a QA program is implemented. A QA program is usually divided into two classifications: the intralaboratory (in-house) and interlaboratory (between-laboratory) QC programs. The former program is a continuing and systematic regime carried out by each laboratory individually. Unless several appropriate SRMs are available for the specific parameter, substrate, sample matrix, and concentration levels under study, an in-house QA program can reflect the performance of each laboratory in isolation only. In order to assess or ensure the reliability and compatibility of sets of data generated by different laboratories, an interlaboratory QA program is necessary in addition to the intralaboratory QA program.

QC procedures are intended to make available to analysts a standardized approach for minimizing analytical errors and assuring the generation of good data with the best possible precision and accuracy. This is achieved by implementing into normal laboratory practices steps that would ensure freedom from such sources of error as contamination, matrix effects, human and instrumental bias and random errors, fluctuating instrumental sensitivity, and discrepancies in analytical standards.

Before analysts carry out any specific analytical QC procedures, they should consult the Good Laboratory Practice guidelines (Chapter 7). Good laboratory practice and QC procedures complement one another, and together they lead to a successful QA program.

The following sections provide a brief description of accepted QC procedures and techniques that are commonly used. It is recommended that the NWQL adhere to these practices as much as possible, so that the data generated are of the best possible quality.

5.1 Intralaboratory Quality Control Procedures

For every batch of analyses, the following steps should be followed:

(1) Before deciding to use an analytical standard, cross-check with standards of other manufacturers in order to determine purity and compatibility.

(2) Run an appropriate reagent blank to check for contamination.

(3) Spike reagent blanks at various concentrations to check the performance of the analytical instrument.

(4) Spike about every 10th sample with a concentration of analyte similar to that found in the natural sample. This step checks for matrix interference effects and determines recovery.

(5) If matrix interference is encountered, use the method of standard additions to calculate the concentrations of the analyte (see Section 5.3).

(6) In order to check the performance of the overall procedure, prepare synthetic samples at various concentration levels. Matching the sample matrix gives added information to the analyst.

(7) "Close bracketing" of standards should always be exercised in order to check on changes in sensitivity. This involves running a standard about every 10th sample, as well as running the complete calibration curve before and after the samples.

(8) Run a replicate of about every 10th sample to check the precision of the system. Use Equation 2 in Section 3.2 to calculate the precision of paired determinations.

(9) To determine the precision of n replicates of one sample, analyze one sample 10 or more times and use Equation 1 given in Section 3.

(10) If available, run one or more certified reference materials (CRMs) in order to check the accuracy of the overall method.

(11) Run secondary in-house reference materials (control samples) in order to further check on recovery. Because of the limited supply and high cost of CRMs, in-house reference standards (control samples) should be used. They are also useful in the preparation of QC charts.

(12) For parameters determined routinely, QC charts provide a powerful tool to establish control of the analytical system. Use this technique whenever feasible (see Section 5.4).

(13) If difficulty is encountered in drawing the analytical calibration curve by the freehand technique, use the method of least squares (see Section 5.5).

5.2 Interlaboratory Quality Control

In addition to in-house QC practices, the NWQL participates in the Water Quality Inter-Regional Quality Control program as well as other interlaboratory programs. These studies are designed to determine the degree of compatibility of the data generated by participating laboratories. Furthermore, such studies help determine interlaboratory precision and accuracy and standardize analytical methodology. They also provide valuable data for the certification of SRMs.

A well-designed multisample interlaboratory evaluation (QA) program provides the following:

- the necessary documented information to data users on the overall "competency" of a laboratory *through time* in providing data to costly environmental programs;
- confidence to correlate and present future data sets from a single or different laboratory, because the interlaboratory levels of confidence, through time, can be established;

- a neutral evaluation of the effectiveness of in-house QC procedures and early warning to the operational laboratories;

- a vital mechanism to spot bias in the laboratory measurement process that is undetected in intralaboratory QC;

- quantitative measurement of systematic bias (by a series of well-designed SRMs);

- a valuable data base to compare different analytical systems for the same parameters; and

- realistic criteria for acceptability of data.

5.3 Standard-Addition Method

This technique is a simple way of calculating the concentration of an unknown. Instead of the usual determination of the unknown relative to external standard solutions, several aliquots of the unknown sample are spiked at different levels of concentration of the analyte (usually one blank and two or more spikes). By this process, the standard is subjected to the same chemical matrix as the unknown. The technique then becomes useful whenever matrix interference prevents the complete recovery of the analyte. Therefore, the standard-addition method removes inaccuracies caused by matrix effects.

Although this technique is useful, it is not applicable to all types of interference. The effect of the interfering component on the analyte in the sample solution must be the same as that on the spiking solution. Therefore, the chemical form in which the analyte is added should be as close to, if not the same as, that in the unknown.

For more information on the uses and limitations of this technique, the analyst should consult Klein and Hach (1977). Additional criteria for the proper use of this procedure are that the response should be linear in the range of the additions and that the magnitude of additions should be in the same range as that of the unknown.

The procedure to be used is as follows:

(1) Take three or more aliquots of the sample solution to be analyzed.

(2) Determine the concentration level at which the solutions should be spiked. Spike the solutions with concentrations of 0, 1x, 2x, 4x,..., where x is the predetermined unit of concentration. The first solution is a zero spike or unspiked sample. Note that it is advantageous to make the volume of the spike negligibly small in order not to introduce a dilution factor. However, if this is not possible, the dilution factor introduced to the solutions should be constant. The zero spike should be diluted accordingly.

(3) After analyzing the solutions, plot the scale reading or instrumental response against the concentrations of two or more of the additions (Fig. 2).

(4) Trace a straight line through the points and extend it until it intersects the abscissa. The point of intersection on the abscissa gives the concentration of the unknown sample solution (Fig. 2).

5.4 Quality Control Charts

The use of a QC chart (American Society for Testing and Materials 1976; U.S. Environmental Protection Agency 1984) is simply an extension of the use of a certified or in-house SRM for the purpose of QA of analytical data. This technique allows already-generated data on SRMs or data that are to be generated on them to be used to improve the precision and accuracy of a method and to give the analyst more control of the quality of data he or she generates. Basically, the technique involves the preparation of charts with specified control limits (Figs. 3 and 4). The charts give the analyst a visual graphic presentation of the precision and accuracy of the system as a function of time and provide immediate detection of erroneous and unsuitable data. Furthermore, long-term variations in these data are monitored, and trends in method or operator performance can be determined. Experience has shown that the use of QC charts generally improves the performance of analysts and makes laboratory supervisors better informed about personnel, methods, and equipment.

The essential requirement for the preparation of charts is a large enough supply of an SRM, either a certified SRM (or CRM) purchased commercially or an in-house bulk SRM prepared by the analyst (control samples). The latter is more useful, because it is much cheaper to prepare and the sample matrix can be chosen to be similar to the real samples analyzed.

The SRM must be homogeneous, and the concentration of the analyte in it must be in the same range as that of the real samples. Furthermore, the SRM should be large enough in quantity to last for an appreciable amount of time, during which the concentration of the analyte must be constant. The initial step of preparing the chart is to obtain about 20 pairs of data, each performed on a different day. This allows the calculation of the control limits. Once the chart is prepared, the analyst runs a pair of analyses on the SRM together with every batch of real samples analyzed. Simply plotting the average and difference of the SRM results on the accuracy (Fig. 3) and precision (Fig. 4) charts, respectively, determines whether or not the data are within the control limits. If they are outside the limits, a problem is identified, and proper steps are taken to eliminate it. If the data are within the limits, control of the system is assured, and credibility is improved.

The procedure to be used is as follows:

(1) Prepare a bulk reference sample of the same background matrix as that of the majority of sample types analyzed in the laboratory. Ensure the sample is homogeneous, stable for the lifetime of the standard, and has a concentration of analyte in the same range as the samples.

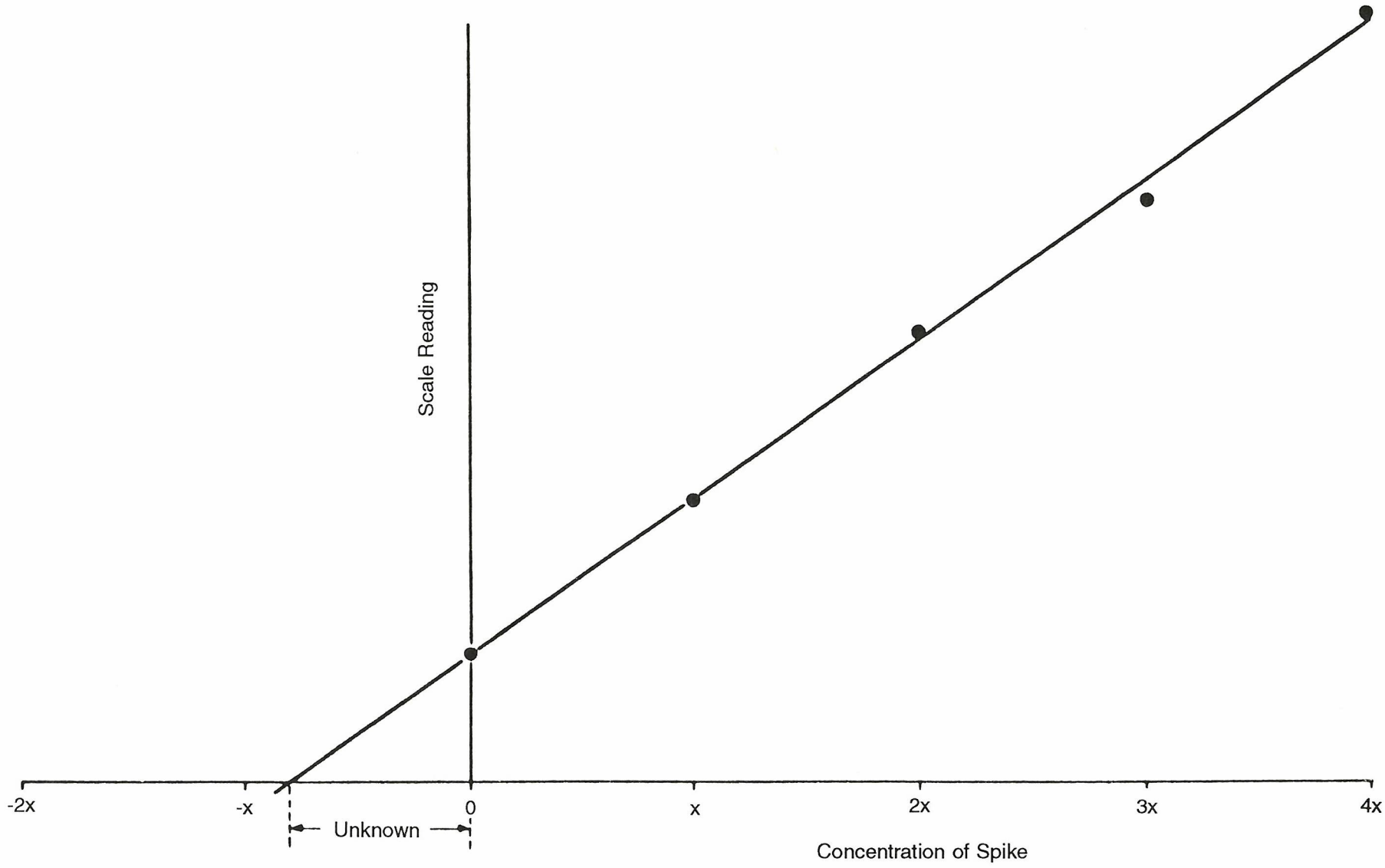


Figure 2. Standard additions curve.

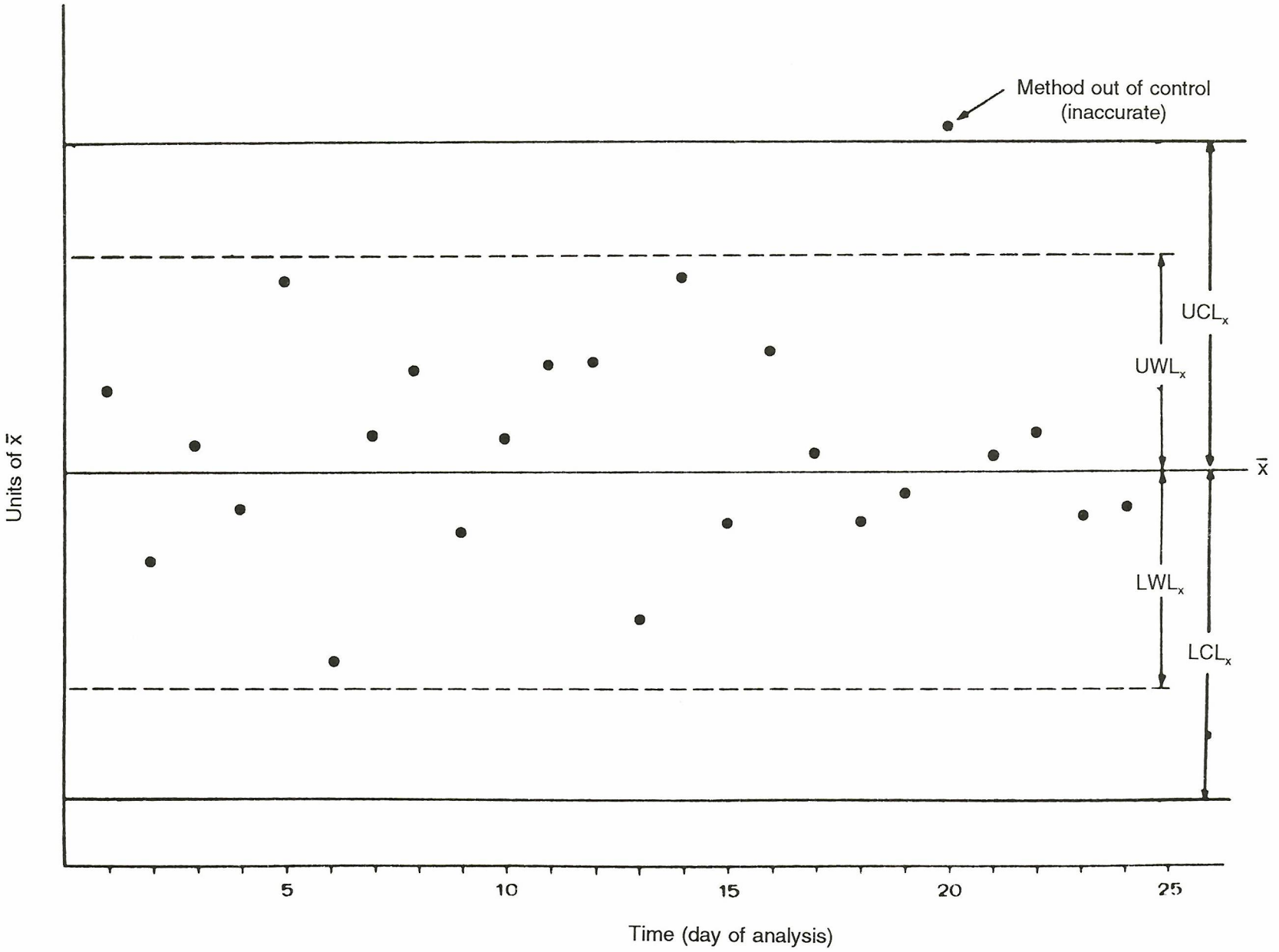


Figure 3. Accuracy control chart ($\bar{x}$-chart).

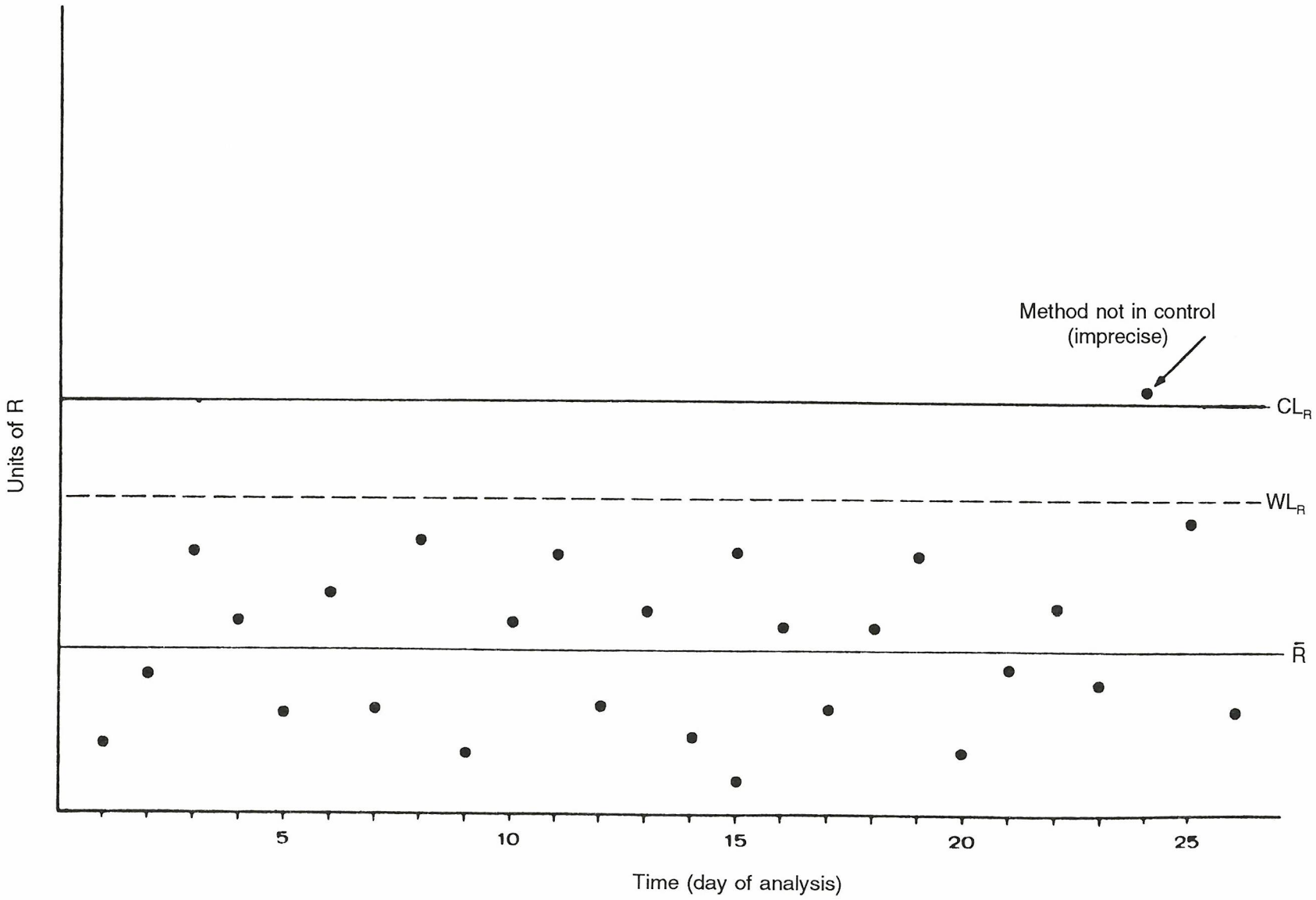


Figure 4. Precision control chart (R-chart).

(2) With the analytical method in control, collect about 20 pairs of results from different days of analysis. The daily data set need not be restricted to pairs of analyses. Statistical treatment of data can be used for any number of determinations of control samples by modifying the warning and control limits (American Society for Testing and Materials 1976). However, for all practical purposes, pairs of analyses are all that are needed. Also, because a laboratory must run control samples at different levels of the calibration curve (ideally at low, medium, and high levels), minimization of the number of QC samples becomes important.

(3) For every pair of results (a_i, b_i), calculate the mean ($\bar{x}_i$) and range (R_i) as follows:

$$\bar{x}_i = \frac{a_i + b_i}{2} \quad (12)$$

$$R_i = |a_i - b_i| \quad (13)$$

(4) For n means and ranges, calculate the grand mean $\bar{\bar{x}}$ and the average range $\bar{R}$ as follows:

$$\bar{\bar{x}} = \frac{(\bar{x}_1 + \bar{x}_2 + \ldots \bar{x}_n)}{n} \quad (14)$$

$$\bar{R} = \frac{(R_1 + R_2 + \ldots R_n)}{n} \quad (15)$$

where n is the number of pairs of data.

(5) Calculate control limits CL_x and CL_R and warning limits WL_x and WL_R for x and R, respectively, as follows:

$$CL_x = 1.880\ \bar{R} \quad (16)$$

$$WL_x = 1.254\ \bar{R} \quad (17)$$

$$CL_R = 3.267\ \bar{R} \quad (18)$$

$$WL_R = 2.178\ \bar{R} \quad (19)$$

(6) Prepare an accuracy control chart ($\bar{x}$-chart), as shown in Figure 2, by drawing upper and lower control limit lines CL_x units above and below the grand mean ($\bar{\bar{x}}$) value, respectively:

$$UCL_x = \bar{\bar{x}} + 1.880\ \bar{R} \quad (20)$$

$$LCL_x = \bar{\bar{x}} - 1.880\ \bar{R} \quad (21)$$

and upper and lower warning limit lines WL_x units above and below the grand mean ($\bar{\bar{x}}$) value, respectively:

$$UWL_x = \bar{\bar{x}} + 1.254\ \bar{R} \quad (22)$$

$$LWL_x = \bar{\bar{x}} - 1.254\ \bar{R} \quad (23)$$

The chart is now ready for use.

(7) Prepare a precision control chart (R-chart), as shown in Figure 3, by drawing upper control limit and upper warning limit lines, as in Equations 18 and 19. The chart is now ready for use.

(8) For a batch of real samples, run a pair of analyses on the designated in-house reference sample. Calculate $\bar{x}_i$ and R_i and plot them on the $\bar{x}$- and R-charts, respectively. If these points lie outside the control limits, a problem has been identified. At this point, resolve the problem and, if necessary, reanalyze the samples.

(9) As more $\bar{x}_i$ and R_i points become available, recalculate the control limits by taking the larger set of points available.

5.5 Calibration Curves by Least Squares Method

The analytical calibration curve is the means by which concentrations of unknowns can be related to standards. Because these curves provide a graphical representation of analytical standards, it is imperative that they be of good quality. Any error involved in drawing the calibration curve will affect the result. The process of drawing a curve involves plotting a set of points on a rectangular coordinate system to give a scatter diagram of the relationship of analytical signal to the actual concentration of the analyte in the standard. The next step requires the fitting of these points to the best straight line that represents the above relationship.

The common method of curve fitting is the freehand technique, where individual judgment of the analyst is used in obtaining the best line. The technique is, however, susceptible to human error and, depending on a given case, may possibly contribute to large errors in the standardization of the analytical system.

The least squares method of curve fitting removes the possibility of human error by eliminating the need for individual judgment by the analyst. The technique simply makes possible the calculation of the intercept and slope of the best line, which is drawn such that the square of its distance from every point has the smallest possible magnitude. The analyst simply uses the calculated value of the slope and intercept of the straight line to obtain the mathematical expression describing the line and then uses this expression to calculate the new line.

Given a set of n data points (x_i, y_i) on a rectangular coordinate system that behave in a linear fashion, the equation of the straight line that describes this relationship is as follows:

$$y = a + bx$$

where: a is equal to the y– intercept of the straight line; and
b is the slope of the line.

Using the least squares approach, a and b are calculated as follows:

$$a = \bar{y} - b\bar{x} \tag{24}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{25}$$

These calculations are best performed using a simple computer program. Automated curve-fitting programs are available in the AWQUALABS data management system. TYFIT is the general curve-fitting program for first- to fifth-degree equations. Other curve-fitting programs, such as THILO, TYSUL, and TYCHL, are also available for special cases. Curves may also be automatically plotted using the program TYPLT. The AWQUALABS documentation should be consulted for more information. The use of automated curve fitting eliminates the inter-operator curve-plotting errors and subsequent uncertainties present in a determination.

5.6 References

American Society for Testing and Materials. 1976. ASTM Manual on Presentation of Data and Control Chart Analysis. STP-15D, October.

Klein, R., Jr., and C. Hach. 1977. Standard additions, uses and limitations in spectrophotometric analysis. Am. Lab. (July): 21.

U.S. Environmental Protection Agency. 1984. Handbook for Analytical Quality Control in Water and Wastewater Laboratories. EPA-600/4-79-019, March.

6.0 QUALITY CONTROL PRACTICES IN THE NATIONAL WATER QUALITY LABORATORY

At the NWQL, a systematic intralaboratory QC program is recognized to be an essential element of the activities of an analytical laboratory. This practice is applied to every step of the analytical process, from sample receiving to final data reporting and recycling of field sampling containers.

The intralaboratory QC program is divided into the following areas and is customized to the various disciplines of analytical specialty:

- Sample Receiving Facility;
- Bottle Preparation Laboratory;
- Major Ions, Nutrients, and Physical Parameters Laboratory;
- Atomic Spectroscopy Laboratory;
- Organic Analysis Laboratory;
- Gas Chromatography/Mass Spectrometry Laboratory; and
- Automated Water Quality Laboratory Data Acquisition and Management System (AWQUALABS).

Each unit within the NWQL performs intralaboratory QC activities, which consist of calibration checks, analysis of blanks, spikes, replicates, and reference and control samples on a regular basis, as well as logistic and data validation and verification checks. The following are QC activities that are conducted on a regular basis within the various laboratory units.

6.1 Sample Receiving Facility

(1) Laboratory clients are required to accompany every batch of samples being submitted for analysis with an environmental sample submission form (Environment Canada form 067-2152 [02/84]).

(2) The sample submission form contains all pertinent information required to identify the samples and to describe the nature of requests, including the following:

- Project Number
- Project Leader

- Sample Submitter
- Date of Shipment
- Method of Shipment
- NAQUADAT Submitter I.D.
- NAQUADAT Station I.D.
- Sample I.D.
- Date and Time of Sampling
- Sample Type
- Analytical Request.

(3) Information on samples received at the NWQL is checked immediately to verify data on submission sheets against bottle labels. Any discrepancies or irregularities such as missing samples, mislabelled bottles, or damaged containers are recorded on submission sheets, and the information is relayed to the project leader.

(4) Each sample is assigned an AWQUALABS laboratory number and appropriate analytical schemata to identify requested parameters. Samples are labelled with this information and then initialized in the AWQUALABS computer system.

(5) A parallel set of AWQUALABS laboratory numbers is kept in a hard-copy backup logbook for later verification.

(6) Accuracy of sample identification information is verified the following day against a list generated by the AWQUALABS system.

(7) After verification, copies of submission forms (with AWQUALABS laboratory numbers and remarks) are returned to project leaders as acknowledgment of receipt of samples and for future reference.

(8) Copies of submission forms are also sent to various laboratory unit heads in the NWQL.

(9) Samples are stored according to conditions specified in the Water Quality Analytical Methods Manual (Environment Canada 1979) until ready for analysis.

6.2 Bottle Preparation Laboratory

(1) The NWQL provides an ongoing bottle washing and preparation service to WQB regional offices. After data are reported on analyzed samples, used bottles are rewashed under strict analytical and QC protocols. When verification is complete, the clean bottles are returned to the original client.

(2) Bottles are treated according to techniques specified in the Water Quality Analytical Methods Manual (Environment Canada 1979).

(3) After cleaning, approximately 2 to 5% of every batch of clean bottles are randomly selected for QC analysis by the appropriate laboratory.

(4) The QC chemist inspects the data, and any suspect batch of bottles is returned to the laboratory for rewashing.

(5) A copy of the bottle QC report is returned with every batch of washed bottles for client information.

(6) Bottles belonging to each WQB region are washed as a batch separate from other WQB regional sample containers, and QC analyses are carried out specifically for that batch.

(7) Because bottles are color-coded for each WQB region, clients are assured of obtaining bottles with a sample history pertaining only to their region.

(8) Several times a year, as requested by the WQB regions, analytical preservatives are tested at the NWQL and forwarded to the regional central offices.

6.3 Major Ions, Nutrients, and Physical Parameters Laboratory

6.3.1 Operational Considerations

6.3.1.1 Deionized Distilled Water Blanks

Deionized distilled water blanks are monitored routinely in the laboratory. A set of two blanks is run immediately after the calibration standards and at the end of the run, following either another full set of calibration standards or a given standard. The blanks are checked against the baseline water, which usually stays unchanged from day to day. All containers used to store distilled deionized water for field activities such as cruises on the Great Lakes are checked for blanks. If any of the containers give a positive response, they are flushed with distilled deionized water and checked again. This procedure is also performed on the sample cups. The deionizing cartridges (ultrapure D0809) are changed before the meter on the deionizing column shows a conductivity reading of 0.2 MΩ.

6.3.1.2 Analytical Standard Preparation

Stock standards are prepared every six to eight months or when nearly depleted. New stock is checked against the old stock standard.

Intermediate and working standards are prepared covering different analytical ranges suitable for the various groups of parameters, as specified in the analytical method.

Working standards for nutrients and alkalinity are prepared daily, and stock solutions are kept at 4°C. Standards for major ions maintain their integrity for long preservation periods, and are therefore used until they are consumed.

6.3.1.3 Calibration Procedures

Instruments with linear output (e.g., autoanalyzers) are calibrated with a minimum of two standards (mixed standard if more than one parameter are analyzed simultaneously). In the case of nonlinear output, the instrument is calibrated with four to five standards covering low to high ranges of concentration (Ca, Mg, Na, K). Linearity test and reproducibility of calibration curves are frequently performed. Instrument response is normally monitored by keeping track of the standard calibration curve of the system. For the chloride/nitrate/sulphate ion chromatographic technique, the instrument is calibrated with three standards covering low, medium, and high levels of the working range. For the chloride/sulphate/silica autoanalyzer technique, the system is calibrated with six levels of standards.

Instrument response is monitored by keeping records of standard calibration curves on each instrument.

6.3.1.4 Precautions to Check Quality of Glassware

The glassware used in the preparation of the reagents is washed with tap water and rinsed several times with deionized water before and after use. The graduated cylinders used for dispensing different types of acid and reagents are labelled appropriately to avoid cross-contamination.

6.3.1.5 Use of Control Samples

Three control samples (high, medium, and low) of known concentration are run per batch of 40 to 60 samples. If the analytical result falls outside the normal range of variation, then the calibration standards, reagents, and blanks are checked. Analysis is resumed only when the problem is rectified.

6.3.1.6 Reproducibility and Performance Checks

The within-run and between-run reproducibility is checked by means of reference samples, and replicate samples are usually analyzed at the beginning and at the end of a batch. Depending on the size of the batch, two or more replicates may be analyzed. Often three blanks and two replicates are sufficient for a batch of 40 to 60 samples.

6.3.1.7 Calculation and Data Management

Analytical data are fed into the computer either manually or by direct data capture from the analytical instrument, depending on the automated setup. All calculations, curve fitting, etc. are done by computer. A hard copy of the data is obtained by each technician, and cross-checked with the printout from the instrument or recorder charts before the final transfer into the data base.

6.3.2 Precision, Accuracy, and Bias Considerations

6.3.2.1 Calculation of Standard Deviation

The standard deviation is calculated on a regular basis using replicates of control samples and duplicates of real samples.

6.3.2.2 Recheck Analysis

Analysis of replicates is performed on a regular basis at a rate of one per 15 to 20 samples.

6.3.2.3 Analysis of Reference Sample and Spike Recoveries

Analysis of a spiked sample is performed and recoveries are calculated once a day; recoveries are plotted on a control chart on a regular basis.

6.3.3 Controls to Detect the Occurrence of Unexpected Observations

6.3.3.1 Calibration Curves

Repeatability of calibration standards should be within the accepted limits based on the 95% confidence interval. Calibration curves are constantly monitored for linearity.

6.3.3.2 Reproducibility Within and Between Runs

Differences in standards between and within runs should fall within the normal range of variation characterized by the 95% confidence limits of the system.

6.3.3.3 Accepted Limit for Replicate

Control charts with the mean difference of duplicate samples are used for all methods. The upper and lower control limits are set as described in Section 5.4. This is done with the within-run and between-run duplicates. The data used for the original calculation are accumulated over a period of approximately two to three months. Any values that fall outside those limits are reexamined and, if necessary, reanalyzed.

6.3.3.4 Accepted Limit for the Spike Recovery

Control charts are also set up for spike recovery. The upper and lower limits for each parameter are based on data accumulated over a certain period of time (i.e., approximately one month). The spiked sample can be either a composite sample (i.e., a mixture of many samples) or the sample by itself. Composite samples are usually made from a large batch of samples with the same matrix (i.e., lake samples). Here, again, the upper and lower limits are set as described in Section 5.4.

6.3.3.5 Analysis of Control Samples

Control samples are analyzed on a regular basis. If the analytical result falls outside the normal control limits, then the calibration, standards, reagents, and blanks are checked. Analysis is resumed only when the problem is rectified.

6.4 Atomic Spectroscopy Laboratory

6.4.1 Operational Considerations

6.4.1.1 Deionized Distilled Water Blanks

To ensure a good supply of deionized distilled water, the cartridges in the deionizing column are monitored regularly to ensure that a reading of less than 0.2 MΩ is maintained in the column.

6.4.1.2 Reagent Blanks

Two reagent blanks are run with every batch of 50 or less samples. In the case of the solvent extraction technique, this will show the composite blank level of the deionized distilled water, chelating agent, buffer, acids, and glassware. The blanks are studied regularly to identify possible contaminant sources.

6.4.1.3 Analytical Standard Preparation

High-quality certified stock standards (1000 ppm) for each metal are obtained from commercial sources. New stock solutions are checked against old standards. Mixed intermediate standard solutions containing 12 metals (Al, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb, V, Zn) are prepared at levels of 50, 1, and 0.1 ppm in 0.2% nitric acid. These intermediates are used for consecutive batches until exhausted (normally two months). When new intermediates are prepared, they are cross-checked against old standards to ensure continuity.

Working standards (mixed 12) are made at levels of 1, 2, 5, 10, 20, 50, and 100 ppb for each analytical run.

6.4.1.4 Calibration Procedures

(1) Atomic absorption and emission spectrophotometers are optimized using instrumental specifications.

(2) Instrument performance is tested using a standard test solution (e.g., Cu^{2+}) for atomic absorption spectrophotometers.

(3) Batches of an average 50 samples are run, preceded and followed by the seven levels of calibration standards.

6.4.1.5 Precautions to Check Quality of Glassware

The glassware used is washed with deionized distilled water and then soaked overnight in 5% HNO_3. This glassware includes volumetric flasks, graduated cylinders, and pipettes.

All reagent containers and dispensers are properly labelled and colour-coded to avoid errors and cross-contamination.

6.4.1.6 Use of Reference Samples

Two reference samples (high and low) of known concentrations are run per batch of 50 samples or less.

6.4.1.7 Reproducibility and Performance Checks

Two blanks, two replicates, and one spike are analyzed in a batch of 50 samples or less for each metal.

6.4.1.8 Calculation and Data Management

All data are generated by automatic calibration curve fitting, and necessary corrections are made by computer programs available in the AWQUALABS system. Data are automatically captured by the VAX 11/750 computer using AWQUALABS and then formatted into a report on hard copy. Each technician verifies the output, and cross-checks with the printout or chart tracings on the analytical instrument before data are transferred into the data base.

6.4.2 Precision, Accuracy, and Bias Considerations

6.4.2.1 Calculation of Standard Deviation

The standard deviation is calculated every three months using the pairs technique (Equation 2, Section 3.2).

6.4.2.2 Within-run and Between-run Standard Deviation

Within-run standard deviation is calculated using duplicates of unknown samples for each metal. Between-run standard deviation is calculated by analyzing in-house control samples A and B and studying reproducibility over a given period (see Section 6.4.3.5).

6.4.2.3 Calibration Re-run

Calibration curves are regularly plotted and the slopes compared with the range of variation normally expected (95% confidence interval). If suspected of lack of control, instrument sensitivity is checked to correct problems.

6.4.2.4 Replicate Analysis

Analysis of replicates is performed on a regular basis. Every 25th sample in a batch is reanalyzed for all parameters.

6.4.2.5 Analysis of Reference Sample and Spike Recoveries

Analysis of a spiked sample is performed and recoveries are calculated for each run on a regular basis.

6.4.3 Controls to Detect the Occurrence of Unexpected Observations

6.4.3.1 Calibration Curves

Repeatability of calibration standards should be within the normal variation defined as the 95% confidence level.

6.4.3.2 Reproducibility Within and Between Runs

Within- and between-run standards should be within the normal range of analytical variation.

6.4.3.3 Accepted Limits for Replicates

If the variation in duplicates exceeds the 95% confidence interval between and within runs, the analysis is repeated.

6.4.3.4 Accepted Limits for the Spike Recovery

Upper and lower control limits (UCL and LCL) for each parameter are determined by spiking a representative sample. The spike recoveries should fall within these limits (see Section 5.4).

6.4.3.5 Analysis of Control Samples

Synthetic composite control samples A and B (low and high levels) are analyzed on a regular basis. These control samples contain 12 metals at the following concentrations: 2 and 10 ppb for Cd, Co, Cr, Pb, Mo, Ni, and V; 5 and 20 ppb for Cu, Fe, and Zn; and 10 and 50 ppb for Mn and Al. If the

analytical results are not within the normal range of variation, then the calibration standards, reagents, and blanks are checked. Analysis is resumed only when the problem is rectified.

6.4.3.6 Certified Reference Materials

CRMs are analyzed together with every batch of fish/sediment/water on a regular basis. Results must be within acceptable limits before data on unknowns are released.

6.5 Organic Analysis Laboratory

6.5.1 Operational Considerations

6.5.1.1 Reagent Purity

The purity of reagents, solvents, adsorbents, reagent-grade water, etc. is of great importance when determining constituents in the low parts per billion or parts per trillion range. Water blanks are monitored routinely and checked for possible interfering contaminants. Analysts must ensure that the level of purity meets the requirements of their analysis. Pesticide-grade solvents and HPLC-grade solvents are concentrated and analyzed to ensure purity. All reagents used in the extraction and cleanup process are potential sources of contamination. It is mandatory that reagent blanks be run constantly for each analytical procedure, with final extracts being reduced to the same concentration normally used for the sample material. A reagent blank involves the repetition of the entire procedure without including the sample itself.

6.5.1.2 Cleaning of Glassware

Glassware must be entirely free of contamination. The cleaning procedure includes soaking in chromic acid cleaning solution, rinsing with distilled water, rinsing with acetone and hexane, and drying in an oven 80 to 100°C . Clean, dry glassware is stored in a dust-free cabinet.

6.5.1.3 Analytical Reference Standards

A reference standard must be used in all determinations requiring comparison with a chemical substance. Storage, handling, and labelling procedures must be followed to ensure the integrity of all reference standards. If a substance is not certified, the laboratory must develop and perform tests in order to assure that it is suitable for the tests intended. Reference standards are under the control of a designated staff member whose responsibility is to ensure the proper preparation and storage of analytical standard solutions.

6.5.1.4 Calibration Procedures

Instruments and equipment are calibrated according to a specific schedule with four to five standards covering low, medium, and high ranges. Response is monitored by keeping track of the

standard calibration of the instrument. Performance criteria are provided for each group of instruments to ensure that they are functioning properly. Routine maintenance is provided as specified in instrument and equipment manuals.

Written records are kept for each instrument, including description, model, serial number, date of purchase, cost, description of malfunctions and repairs, including parts replaced, date of repairs, downtime, and person performing service, routine maintenance and date and a description of performance checks, specifically the identity of the standard, the method, raw data, calculations, result, operation, and the date. A specific individual is responsible for instruments.

6.5.2 Precision, Accuracy, and Bias Considerations

6.5.2.1 Replicates

Replicate analysis in which the entire analytical method is repeated on a portion of the same sample is carried out on every 10th sample. A blank is processed for every batch of samples analyzed.

6.5.2.2 Recovery Studies (Spiked Samples)

To check whether losses have occurred during the analysis or whether interferences exist, the analyst performs a recovery check employing the addition of a known amount of substance of interest to a sample that contains no detectable amount of the substance of interest. The concentration of spikes is normally at the same level of the expected analyte concentration in that batch of samples.

If the recovery of any parameter falls outside the normal range of variation (95% confidence interval), laboratory performance for the parameter is considered unsatisfactory, and unspiked samples are suspect and not reported. At least 10% of all samples are spiked with compounds being measured.

6.5.2.3 Calculation of Standard Deviation

As part of the QC program for the laboratory, the standard deviation of the percent recovery (SD_p) is calculated after the analysis of 10 spiked water, fish, or solvent samples. The accuracy assessment is then expressed as a percent recovery using the 95% confidence interval. The accuracy assessment is updated for each parameter on a regular basis.

6.5.2.4 Within-run and Between-run Precision

Within-run and between-run precision is calculated on new methods introduced in the Organic Analysis Laboratory, and the standard deviation is calculated on a regular basis.

6.5.2.5 Quality Control Check Standards

The laboratory must demonstrate on a regular basis, through the analysis of QC check standards, that the operation of the measurement system is in control. The laboratory participates in relevant performance evaluation studies whenever possible.

6.5.2.6 Confirmation Techniques

When doubt exists over the identification of a peak on the chromatogram, confirmatory techniques such as GC with a dissimilar column, specific element detector, or mass spectrometer must be used.

6.5.3 Record Keeping in the Organic Analysis Laboratory

An integral part of good laboratory practices is to maintain adequate records of analytical activities.

In addition to numerical results of analysis, records should be kept with the chromatograms that provide supplemental operating data. The data and information supplied by the Organic Analysis Laboratory should include sample size, extent of concentration of the sample extract, injection volumes, elution cuts if column cleanup is required, all instrumental operating parameters, and identity of the GC column. Chromatograms must be clearly identified so that they may be related to the data reported to clients. Data should include quantitation using peak heights together with calculations using peak areas. The records should clearly indicate the names of the analyst and operator of the chromatographic system.

Instrument operating parameters should include column type, temperature of column, length of column, detector temperature, inlet temperature, carrier flow rate, purge flow, and chart speed.

Thorough record keeping is very important, especially when analyzing the reasons for changed analytical performance.

6.5.4 Reporting

All information pertaining to an analysis is permanently recorded in the laboratory workbook. All valid analytical results are entered manually into the computer files, and a hard copy is then checked for reporting quality, accuracy of identifying data, and analytical quality. If the data are acceptable to the Head, Organic Analysis Laboratory, a copy of the report is sent to the client. Furthermore, data requiring storage on the NAQUADAT system are formatted on magnetic tapes and forwarded to the appropriate authority for storage.

6.6 Gas Chromatography/Mass Spectrometry Laboratory

Because work carried out in the Gas Chromatography/Mass Spectrometry laboratory is an extension of what is done by the Organic Analysis Laboratory, the QA protocol practised in that facility also forms a basis for procedures used in GC/MS. Additional work must be undertaken to ensure GC/MS data are correctly acquired and interpreted and that the associated complex instrumentation is operating properly and in control.

6.6.1 System Applications

Analytical applications for GC/MS in the NWQL are as follows:

- qualitative confirmation of positive residue detections reported by the Organic Analysis Laboratory;
- quantitative confirmation of analytical data reported by the Organic Analysis Laboratory;
- qualitative identification of unknown materials eluting from GC analysis carried out in the Organic Analysis Laboratory;
- quantitative analysis for trace levels of organic contaminants for which methods amenable to routine application in the Organic Analysis Laboratory are not readily available; and
- identification and quantitation of compounds in samples collected specifically for screening of unknown contaminants and characterize pollutant profiles at selected sites.

6.6.2 System Operational Characteristics

Equipment operated by the NWQL includes low-resolution, fast-scanning quadrupole mass spectrometers interfaced directly to high-resolution, capillary gas chromatographs with a variety of injection modes. Mass spectrometer operating modes available are electron ionization and positive/negative (single or simultaneous) chemical ionization with filament or Townsend discharge. The instrumentation is computer-controlled with acquisition modes for full-scan or selected (multiple) ion monitoring.

6.6.3 System Performance

The principal components of the GC/MS systems are the gas chromatograph, the mass spectrometer, and the controlling data system. QC practices for GC are similar to those described for the Organic Analysis Laboratory. The additional protocol encompasses both mass spectrometer and data system, as they are always operated in conjunction.

The performance criteria that must be controlled are resolution, sensitivity, noise, and stability, both in the mass axis and in the intensity axis. A standard calibrant gas, FC-43 (perfluorotributylamine), is used to set up the instrument. The operating mode is selected and the appropriate parameters set according to the manufacturer's specifications (Finnigan Instruments Service Handbook 1978).

The calibrant is leaked into the instrument, and analog tuning of source lenses to optimize peak shape, resolution, sensitivity, and uniform mass transmission is carried out. This procedure is defined by the manufacturer and is partially described in Section 6.6.6.

Upon initialization, the data system (INCOS) automatically performs internal verification of all functions and produces a diagnostic report for operator review and retention. Calibration is performed by the data system using the FC-43 calibrant subsequent to tuning and acquisition of parameters specific to the intended analysis. Again, calibration diagnostic reports are produced. All of these reports are filed together with representative spectra, to serve as an historical record of system performance and to assist in isolating and diagnosing operational anomalies if and when they occur.

6.6.4 Applications Criteria

6.6.4.1 Qualitative Confirmation

Standard (recorded) mass spectrometer operating conditions applicable to the selected mode(s) of operation and GC conditions similar to those employed by the Organic Analysis Laboratory are established. Either full spectra, or, in the more common case of very low analyte levels, selected ion scans are acquired for both the analytical standard and the sample in question. Both are matched against standard library information* using tested algorithms for "purity" and "goodness of fit" on mass assignments, ion ratios, and retention index for qualitative confirmation.

6.6.4.2 Quantitative Confirmation

Quantitation is performed using a single characteristic ion and an internal standard, such as deuterium-labelled anthracene or an isotopically labelled analyte.

---

* The matching of spectra is performed against the NIH/EPA mass spectral library containing some 38 760 entries.

PURITY measures the resemblance of the currently selected data to the specified library entry. If PURITY is 1000, the set of peaks LIBRary reads from the unknown are the same in mass and locally normalized intensity as those extracted from the library. A PURITY of 800 suggest that the two compounds involved are closely related. A PURITY of 600 or more suggests that the two compounds have many fragments in common.

FIT measures the resemblance of the library spectrum to the unknown, excluding peaks in the unknown not present in the library spectrum. A high FIT (800-900) with a lower PURITY (500-600) suggest that the unknown is a mixture that includes the compound selected from the library or that the two compounds have some major substructure in common.

6.6.4.3 Qualitative Identification of Unknown Materials

This may be carried out at various levels of confidence depending upon the nature of the original request. Identification in ascending order of reliability is as follows:

(1) The full-scan, electron ionization, mass spectral data from unknown GC eluants are subjected to computerized deconvolution and compared with similarly generated mass spectral information in a library data base. "Purity" and "goodness of fit" using algorithms incorporating mass assignment and ion ratios are tested and scores assigned. Identification must be considered tentative, as many compounds have similar spectra, and the identified material may very well be a homolog of the actual contaminant.

(2) The acquired mass spectra may be examined manually by a skilled interpreter. Characteristic spectral patterns are combined, and comparison with published compendia of spectra, within specified acceptance criteria, is the basis for identification.

(3) The information derived from above may be combined with retention index/boiling point information to further increase confidence of identification. The sample is re-analyzed either in full-scan mode or in selected ion monitoring mode using five ions characteristic of the tentatively identified material.

(4) Comparison of the unknown with the authentic material using the above criteria is highly definitive. Co-injection of the unknown with the authentic material with no observed change in retention index or mass spectrum is considered ultimate confirmation.

6.6.4.4 Quantitative Analysis of Target Organic Contaminants

This section deals with target compound analysis using GC/MS detection techniques on parameters for which no suitable methods are currently available in the Organic Analysis Laboratory.

A standard library containing the compounds of interest is prepared in the multiple-ion monitoring mode using a defined set of operating conditions. Identification is based upon a unique "quantitation ion" and up to six other selected confirming ions. The standard is injected, and the library conditions are calibrated. The injection includes an internal standard, deuterium-labelled anthracene, which is also added, in the same absolute quantity, to the sample. Sample data are compared with the calibrated library data, and identification is based upon an algorithm incorporating mass assignments, ion ratios, and retention index. Quantitation is based upon the internal standard and the calibrated response factors to the target analytes.

6.6.4.5 Screening and Characterization for Unknown Contaminants

This is similar to the qualitative identification of unknown materials (see Section 6.6.4.3), except that a wider range of GC or GC/MS operating conditions may be employed in conjunction with

other chemical techniques. As this work tends to be more investigative in nature than the other categories discussed, QA protocols are necessarily less clearly defined. However, specific QA protocols are developed, established, and applied as projects of this type evolve.

6.6.5 Target Compound Analysis Performance Records

Key performance data from target compound analysis are retained over time. Response factors, spike recovery data, and detection criteria, computed and statistically updated through successive runs, are stored. QC charting is performed on the accumulated data, providing detailed information with respect to method and operator performance over time as well as early warning of impending out-of-control conditions. These charts and attendant statistical summaries are made available to data users to assist with data assessment and interpretation.

6.6.6 System Setup Procedure

6.6.6.1 INCOS Data System

(1) FC-43 calibration gas is admitted to the system such that m/z 69 is just below saturation.

(2) System parameters are adjusted to scan from m/z 60 to m/z 510 every 2 seconds.

(3) Peak intensity ratios and ion abundances are compared with historical performance records to assess mass transfer and system sensitivity; a typical spectrum with the current performance information is retained.

(4) The system calibration program is run using the newly acquired data.

(5) Diagnostics software compares the new calibration with manufacturer specifications for high mass transfer and quality of mass assignment with the mass table; an evaluation report is printed for use by the operator in deciding whether or not the system performance is adequate.

(6) System software is used to establish correct "zero" and "gain" settings according to manufacturer specifications; a graphical report is provided to enable the operator to assess performance.

(7) Upon completion of data acquisition for sample analysis, a summary diagnostic report is printed, indicating to the operator if problems were encountered with respect to selection of acquisition parameters.

6.6.6.2 Shrader Data System

(1) System scanning parameters are set up and acquisition is initiated without mass storage, the spectra being continuously displayed on the video terminal.

(2) The levels of admitted calibrant and acquisition parameters are adjusted based upon the display and accompanying diagnostic information within criteria recommended by the manufacturer.

(3) Several scans are stored and averaged, and the mean result is displayed; mass assignments for characteristic m/z values are made by the analyst, and the mass axis for the range of interest is calibrated; acquisition parameters, diagnostic information from the calibration fit, and the resulting calibration spectrum are stored.

(4) The entire calibration is compared with previous similar files to detect errors, such as mass misassignments; diagnostic data are tabulated for easy comparison over time to detect trends in system change and to highlight sporadic problems.

6.7 Automated Water Quality Laboratory Data Acquisition and Management System (AWQUALABS)

There are four general areas where the automated water quality laboratory data acquisition and management system (AWQUALABS) currently operates to support the QA practices of the NWQL:

- sample initialization;
- sample analysis;
- data entry; and
- data verification.

In addition, general system and program access is restricted according to staff responsibility, and data audit trails have been incorporated to minimize analytical error resulting from belated testing for labile (time) parameters.

It must be emphasized that these procedures have been designed to assist in carrying out effective QC and do not supplant the vital role of the analyst in this regard. At selected points throughout the analytical activity, summary reports of associated QA work are conveniently presented for evaluation by the appropriate technician, chemist, manager, or client.

Present capabilities of AWQUALABS are outlined in Sections 6.7.1.1 to 6.7.1.4.

Software is currently being developed to:

- perform a more comprehensive series of data checks; and

- operate on QC data that are generated during the analysis stages described above, and provide a definitive and ongoing assessment of laboratory performance.

These activities are outlined in Sections 6.7.2.1 to 6.7.2.4.

6.7.1 Present Activities

6.7.1.1 Sample Initialization

Certain elements of the AWQUALABS data base are predefined according to project specifications. These entries are recalled at the time of individual sample initialization either as sample file entries or to verify that operator-entered items are valid. Included are analytical methods (parameter dictionary), parameters and substrates for testing (analysis schemes or schemata), NAQUADAT station numbers and descriptions, and analysis restrictions such as detection criteria, reporting units, and significant figures.

Samples are initialized into a temporary file, not directly to the active data base. The information in this file is reviewed independently for correctness by personnel other than those responsible for initialization.

6.7.1.2 Sample Analysis

Control of laboratory analysis is exercised through a menu-driven series of programs that guide analysts systematically through the various steps necessary to carry out the following:

(1) *Identification of work to be done* — Work sheets are generated that identify samples and parameters for testing.

(2) *Analysis* — Chemical analysis and attendant intralaboratory QC are carried out either manually or by direct data acquisition from instruments.

(3) *Synchronization of sampling and baseline* — Samples are synchronized within a batch such that the data system knows whether an incoming value is to be associated with a sample, a standard, or any one of several acceptable categories of aliquot. Baseline correction is performed automatically using blanks interspersed throughout the run. All relevant activities are identified in the final "run report."

(4) *Calibration* — Calibration curves are prepared from measured standard solutions using linear, nonlinear, multipoint, or multilevel fits, depending on the requirements for analysis.

(5) *Measurement* — Sample values are interpolated from a standard calibration curve.

(6) *Verification* — Values that do not fall within predefined criteria, such as those that lie outside upper and lower standard concentration limits are flagged.

(7) *Performance* — Recovery (accuracy) and precision are computed based upon responses for aliquots of defined composition (spikes, splits, replicates).

(8) *Review* — Analytical results are presented in tabular and/or graphic form for subjective review by the analyst.

6.7.1.3 Data Entry

After the analyst is satisfied with the results of testing, the data are automatically added to the sample files:

(1) *Predefined sample files* — Predefinition of these files at the time of sample initialization prevents unwanted data, such as values for parameters that were not requested, from being inserted.

(2) *Multiple entries* — Entries for parameters that already have an associated result are flagged as rechecks and the operator informed appropriately.

(3) *Date audit* — The date of analysis is automatically entered with the analytical result.

(4) *Record keeping* — A hard-copy record of the sample file update is produced for the analyst's files. This process is rapid and free from error in transcription, the latter being a long-standing source of data quality problems in laboratories.

6.7.1.4 Data Verification

There are a number of software modules both under development and in place to perform additional data verification of samples for which analyses have been completed. There are also a number of checkpoints in the data base life history of a sample at which some form of approval must be provided by designated personnel in order for sample processing/reporting to continue.

Currently available are the following:

(1) *Ionic balance* — The difference between sums of cations and anions is calculated and compared with specified criteria for acceptability. In the case of unsatisfactory ionic balance, parameter correlations are carried out automatically, and predictions as to likely sources of error (i.e., those ions that are most probably incorrect) are identified for re-analysis.

(2) *Consistency checks* — Parameters that exhibit relative trends with respect to each other are checked and inconsistencies identified and printed. An example is verification that reported values for dissolved forms of selected parameters do not exceed those for total forms.

(3) *Correlation checks* — Those parameters for which there are established mathematical correlations are checked and inconsistencies identified and printed.

(4) *Sample acceptance* — After all tests and QA checks have been completed, interim reports are printed for final review by chemists and laboratory unit heads. If everything appears satisfactory, the chemist generates a hard-copy report for transmission to the client agency. The system periodically polls the sample data base, and, when all values for a particular sample have been reported, the sample is flagged for archiving. Computer data sets may be transmitted to clients either by magnetic tape or by direct computer–computer transfer.

6.7.2 Future Development

6.7.2.1 General Limits Screening

All data will be compared with a range of realistic and acceptable values for a specific parameter, and out-of-range points will be flagged for further attention. Exceedances will be verified, and, if analytically correct, the information will be rapidly communicated to the appropriate project leaders to permit quick response to potential problems.

6.7.2.2 Site-specific Limits Screening

Where adequate historical information for particular sites is available, a table of data limits characteristic of that location will be prepared. Data for that site will be compared with historical values for early warning of changes. Exceedances will be verified, and, if analytically correct, the information will be rapidly communicated to the appropriate project leaders to permit quick response to potential problems.

6.7.2.3 Chromatographic Decision Matrix

For chromatographic-based analytical procedures, algorithms will be developed that apply selected criteria for system response from multiple-column analysis and determine a confirmed residue quantitation. This will streamline the process of data interpretation for this type of work by reducing the extent of direct intervention by the analyst to results that fail to meet the established criteria. This activity will be carried out on a dedicated data acquisition system currently operating in the Gas Chromatography/Mass Spectrometry Laboratory.

6.7.2.4 Laboratory Performance

At the point of updating sample files in the AWQUALABS data base, all the QC information acquired to that point in time is deleted. To date, it has been the responsibility of the analyst to review these data and make subjective decisions on system performance, take corrective action as deemed necessary, and carry on with the analysis. Software is being developed that will store nonsample data (standards, blanks, spikes, replicates) in a separate file. These data will be operated on to provide an ongoing assessment of laboratory performance for a specific analytical procedure. The activity will be cumulative and ultimately yield:

- laboratory performance over time with respect to a particular procedure, including long-term definition of detection criteria;
- individual analyst performance over time; and
- early diagnosis of method problems.

In addition, sample files will be associated with the intralaboratory QC information derived from the same time period as the analysis, as well as with any interlaboratory studies and check samples that may have been in process in the same time frame. This latter activity will provide capability for complete data specification for all analytical work carried out by the NWQL and assess performance both of individual staff for specific tests and of the tests themselves within the laboratory as a whole.

6.8 References

Environment Canada. 1979. Water Quality Analytical Methods Manual. Water Quality Branch, Inland Waters Directorate, August.

Finnigan Instruments Service Handbook. 1978. Nova 3 Data System.

7.0 GUIDELINES FOR GOOD LABORATORY PRACTICE

One of the most important activities of analytical laboratories is the generation of data. To a large extent, the success of such laboratories can be measured by the quality of their products (i.e., quality of their data). In order to ensure good-quality data, every factor that contributes to the success of a laboratory must be optimized and maintained in good working order. The techniques by which success can be achieved are called Good Laboratory Practice. These practices involve all aspects that make the laboratory effective in accomplishing its specific goals.

The following sections summarize the accepted criteria for Good Laboratory Practice and are designed as guidelines to help NWQL staff perform in the most effective manner and to ensure the greatest possibility of the production of good data. If properly used, these guidelines will improve laboratory performance; if performance is already satisfactory, they will help to maintain it. Furthermore, these guidelines provide for the adoption of a mechanism by which problems may be identified and solved in the event of noncompliance to accepted practices.

7.1 Laboratory Facilities and Safety

7.1.1 Facilities and Safety Equipment

(1) The laboratory facilities should provide an environment that is well lighted, clean, uncongested, and with proper climate control (i.e., temperature and ventilation).

(2) Benches and instrument areas should be supplied with the proper electrical outlets to accommodate instruments with varying power requirements. See Treasury Board Standard 3-3 (Treasury Board 1978, pp. 47–65).

(3) The laboratory should be equipped with properly located eye washes, showers, and breathing apparatus.

(4) Laboratories should be equipped with fire extinguishers, fire alarms, and sprinkler systems.

(5) Separate fume hoods should be available for organics, mineral acids, and perchloric acid.

(6) For detailed specifications of laboratory facility design, the analyst should consult the General Laboratory Design Criteria of the Treasury Board Guide 5-1 (Treasury Board 1978, pp. 40–252).

(7) If analytical activities involve different levels of the laboratory building, a freight elevator should be available for the transportation of laboratory paraphernalia, meeting Treasury Board Standard 3-4 (Treasury Board 1978, pp. 67–74).

(8) All laboratory personnel should have access to an emergency first-aid kit, and health services should be within easy reach, complying with Treasury Board Standards 3-5 and 3-8 (Treasury Board 1978, pp. 75–85 and 97–101).

7.1.2 Safety Practices

(1) Regular monthly safety tours of laboratory premises should be conducted, and each laboratory unit should have representation on a safety committee.

(2) A safety program should be established where laboratory personnel attend periodic seminars, courses, and workshops in order to keep abreast of new developments.

(3) Good records should be kept of the state of safety devices in the laboratory and of any safety-related events in order to ensure the efficiency of the safety program. See Treasury Board Procedure 4-1 (Treasury Board 1978, pp. 183–187).

(4) All laboratory equipment should be regularly checked and its status recorded, in order to prevent hazardous conditions.

(5) Periodic health examinations should be carried out on all personnel involved with toxic chemicals and substances with microbial contamination. See Treasury Board Standard 3-13 (Treasury Board 1978, pp. 139–146).

(6) All personnel should use protective gear for the eyes, face, head, skin, hands, feet, legs, and the respiratory system. See Treasury Board Standard 3-14 (Treasury Board 1978, pp. 147–154).

(7) Lab coats and utensils used inside the laboratory should be restricted to this area to prevent the transport of contamination to other areas.

(8) Smoking and eating should not be permitted inside the laboratory.

(9) When acid or solvent bottles are transported, they should be carried in the proper protective container, such as a rubber bucket.

(10) Separate waste containers should be used for regular laboratory garbage and broken glass.

(11) All contaminated materials to be discarded (e.g., broken glass, paper towels, disposable gloves) should be rinsed with water prior to being placed in waste containers in order to protect waste handlers from toxic chemicals and acids.

(12) Special containers must be utilized for organic waste solvents and should be placed in a properly ventilated area such as a fume hood. Procedures for the handling and disposal of

laboratory wastes should comply with Chapter 6 of Treasury Board Guide 5-1 (Treasury Board 1978, pp. 284–288).

(13) All personnel involved with the handling of dangerous substances and pesticides should comply with Treasury Board Standards 3-2 and 3-15, respectively (Treasury Board 1978, pp. 37–45 and 155–158, respectively; see also Handbook on Chemical Safety, Health and Welfare Canada 1973).

(14) Chemicals such as solvents, acids, reducing and oxidizing agents, and any substances that may not be stable in the presence of other groups of substances should be stored separately under special conditions that allow for ventilation, proper temperature control, and isolation from areas of traffic. See Section 3.5 of Treasury Board Guide 5-1 (Treasury Board 1978, pp. 265–268).

(15) The use, handling, and storage of compressed gas cylinders should comply with Section 3.2 of Treasury Board Guide 5-1 (Treasury Board 1978, pp. 257–260).

(16) Toxic volatile chemicals should be weighed and handled under ventilation and in separate quarters. The use of carcinogenic chemicals should be avoided as much as possible.

(17) Laboratories should have cleanup procedures available for spills of specific chemicals used on their premises.

(18) Laboratory staff should attend seminars on fire safety and be subjected to periodic unexpected fire drills.

(19) In the event that an accident does occur, the Treasury Board Guide to accident investigations should be consulted (Treasury Board 1978, pp. 205–228).

(20) Laboratory personnel should consult the Canada Centre for Inland Waters Laboratory Safety Manual (Environment Canada 1981) whenever appropriate safety information is required.

7.2 Staff Relations and Work Environment

(1) Laboratory heads should ensure that analysts are properly trained and qualified to perform the duties they have been assigned. Analysts should have access to the Analytical Methods Manual (Environment Canada 1979) or must have detailed written instructions on the analytical technique they are to use.

(2) Laboratory personnel should be encouraged to participate in instructive meetings, seminars, and courses to keep abreast of the latest developments in their field.

(3) A mechanism should be present by which laboratory personnel and supervisors communicate and problems can be effectively solved. A system should also be present in which suggestions and complaints can be looked at and required changes made to ensure the efficient functioning of the group.

(4) Managers and supervisors should provide feedback on the performance of analysts and on the results of QC studies in order to maintain the good caliber of work or to help make required improvements.

(5) In the interest of keeping morale high, every effort should be made to provide continuous incentive programs.

(6) An organization chart should be available, showing lines of authority and areas of responsibility.

(7) Laboratory management should ensure that copies of the Handbook of Occupational Health and Safety (Treasury Board 1978) are available to all personnel. Furthermore, individual copies of the Treasury Board Standards, Guides, and Procedures should be present at the pertinent locations and with the appropriate people so that the required information can be obtained quickly.

7.3 Chemicals, Reagents, and Standards

(1) Chemicals, solvents, and standards should be dated upon receipt and discarded when their shelf lives expire.

(2) The level of purity of reagents, solvents, and chemicals should be determined by analyzing appropriate blanks with every set of analyses. Every new batch of chemicals should be checked in the above fashion.

(3) The grade of chemicals or reagents used in an analysis should be as specified in the analytical method.

(4) A good source of deionized distilled water should be available, suitable for analytical use. For organic methods, either XAD-2 cleaned water or the equivalent may be necessary.

(5) Chemical standards employed in analyses should always be analytical grade. Standards from at least two sources of suppliers should be obtained to check one against the other. For organic parameters, it is necessary to check the purity of the standards, as they are not always readily available in a pure form, as is the case with inorganic standards.

(6) Analytical-grade standards should always be kept in a cool and dry place and must be dried before weighing.

(7) For safety-related aspects of the handling, storage, and disposal of chemicals, consult Section 7.1.2.

7.4 Apparatus

(1) Glassware and plasticware should be cleaned with the appropriate cleaner specified in the analytical method.

(2) For parameters that are prone to contamination, a separate set of glassware should be allocated.

(3) Volumetric glassware that is deemed critical to analytical determination should be periodically calibrated.

(4) Preventive care should be practised on all instruments.

(5) Proper optimization and calibration of instrumental conditions should be practised on a regular basis. Instrumentation should always be kept free of contamination and in good working condition. The instrument manual should be kept within easy reach of the system (see Chapter 8).

(6) Analytical instruments should not be kept in the proximity of corrosive fumes and chemicals. If contamination occurs, the instrument should be cleaned and checked immediately.

(7) Only instruments that meet the standards set in the analytical method should be used for analyses.

(8) Analytical balances should be periodically calibrated with certified weights and always kept in a clean condition. Similarly, analytical meters such as pH meters and thermometers should be calibrated and kept clean.

(9) A systematic performance and maintenance checklist should be available for all major instruments and must be used on a regular basis (see Chapter 8).

(10) Records of all maintenance and performance checks should be kept and should be available for inspection.

(11) There should be one person responsible for the care and maintenance of every instrument on the premises.

7.5 Methodology

(1) The techniques employed for sampling, preservation, and analysis should adhere to the recommended methods in the Analytical Methods Manual (Environment Canada 1979), Sampling for Water Quality (Environment Canada 1983), or the current operational procedures with the NWQL. Because all methods in these manuals are NAQUADAT-coded, alterations in the analytical procedures require a change in the code (see Chapter 4). Furthermore, nonadherence to the recommended procedures may in certain cases result in noncompatibility of data sets.

(2) When methods are changed, the changes must be validated and authorized as per protocol (see Chapter 4), and the reasons for the changes and substantiating evidence must be documented. The new procedures should then be written up in the standard format and submitted for consideration for inclusion in the Analytical Methods Manual (Environment Canada 1979).

(3) Sample bottles must be properly cleaned with the recommended cleaner to prevent contamination (see Section 6.2).

(4) For many parameters, bottle type, preservation, storage conditions, and length of storage are critical and must be closely controlled. See Table I of the Analytical Methods Manual (Environment Canada 1979) for the recommended conditions.

(5) For those parameters that may be easily contaminated, a separate set of containers should be set aside and reused.

(6) Samples must be well homogenized prior to subsampling so that a representative sample may be readily obtained. The analyst should ensure that the technique employed for homogenization does indeed affect the required result.

(7) The chapters on QC procedures and practices (Chapters 5 and 6) should be used as a guide to the QA activities of the laboratory. The laboratory must have a comprehensive QC program to ensure that the data generated meet required standards and that satisfactory performance is maintained.

(8) Because chromatographic techniques are not always specific, the analysis of organic parameters should be routinely confirmed by alternative techniques. Confirmation techniques include mass spectrometry, chemical derivatization, and multicolumn gas, liquid, column, and thin layer chromatography.

7.6 Data Recording, Reporting, and Documenting

(1) The following should be recorded for every sample analyzed, and the records must be available for inspection:

- date, place, time, and conditions of sampling, sample identification number, and name of sampler;
- time of preservation and preservation technique used;
- date of analysis and name of analyst;
- method employed for analysis; and
- results of analysis, including charts, calibration graphs, and records of calculations.

(2) Samples should be well identified throughout the whole period of time from sampling to analysis and calculations.

(3) The analyst should use all analytical values obtained in a measurement. Taking the best two out of three results is frowned on, as such practices may change the conclusions substantially. However, data sets may be modified if statistically based rejection techniques are used.

(4) The accepted rules associated with significant figures should be adhered to, so that the method of calculation is uniform and standardized throughout the laboratory's data collection and dissemination activities. The number of significant figures retained in computation and presentation of data should be as described in the ASTM manuals (American Society for Testing and Materials 1976, 1985).

(5) Quality assessment forms (see Chapter 9) should be used as the main tool for documentation of QC data. Evaluation of these data and appropriate feedback to laboratory managers, supervisors, and analysts should ensure an effective QC program.

7.7 References

American Society for Testing and Materials. 1976. ASTM Manual on Presentation of Data and Control Chart Analysis. Special Technical Publication 15-D.

American Society for Testing and Materials. 1985. ASTM Manual on Quality Control of Materials. Prepared by ASTM Committee E-11 on Quality Control of Materials, Special Technical Publication 15-C, January, Philadelphia, Pa.

Environment Canada. 1979. Water Quality Analytical Methods Manual. Water Quality Branch, Inland Waters Directorate, August.

Environment Canada. 1981. Laboratory Safety Manual. Canada Centre for Inland Waters, March.

Environment Canada. 1983. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate.

Health and Welfare Canada. 1973. Handbook on Chemical Safety — In the Aerial Application of Chemical Materials. Cat. No. H31-1273, Information Canada, Ottawa, Ont.

Treasury Board. 1978. Handbook of Occupational Health and Safety. 2nd ed. Cat. No. BT45-3/1978, Printing and Publishing, Supply and Services Canada, Hull, Que.

8.0 GUIDELINES FOR INSTRUMENT PERFORMANCE CONTROL CHECKS

The following tables provide a summary of the instrument performance control checks for systems commonly in use at the NWQL. When adhered to, these procedures should ensure that instrumental bias and variability are minimized and controlled within the accepted levels. Furthermore, they provide a framework that standardizes the maintenance, optimization, and calibration activities carried out by the analyst. Because most analytical systems at the NWQL are used at sub-parts per billion levels, it is essential to keep instrumentation clean and properly standardized so that reproducibility and accuracy of data are kept uniform and within the accepted control limits.

Table 1. Atomic Spectroscopy Performance Control Checks

Instrument	Frequency of checking	Component to check	Procedure
Atomic absorption spectrophotometer (flame)	Daily	Hollow cathode or electrodeless discharge lamps	Warm up before use to get stable current. Optimize lamp position by use of energy meter
	Daily	Wavelength	Dial the recommended wavelength for the metal of interest and fine-tune by maximizing energy meter.
	Daily	Slit	Set slit as per operator's manual.
	Daily	Cloud chamber	Wash with distilled water and acetone and dry.
	Daily	Burner	Align burner using vertical and horizontal settings. Adjust position to peak absorbance.
		Burner O-ring	Inspect visually and replace if necessary.
		Nebulizer end cap	Inspect visually.
	Daily	NO_2 burner	Use scraping sticks to clean carbon deposits.
	Daily	Drain level and drain tube condition	Inspect visually and adjust or replace as necessary.
	Daily	Gas flow meter and supply	Inspect pressure gauge to ensure appropriate setting and adjust gas flow meter as required.
	Daily	Nebulizer	Adjust by rotating knurled nut. Fine-tune by maximizing sensitivity.
	Daily	External standard	Aspirate a 5 ppm Cu standard into the flame and adjust the nebulizer to maximize the absorption reading. If value is less than 0.250, the nebulizer should be reassembled.

Table 1. Continued

Instrument	Frequency of checking	Component to check	Procedure
Atomic absorption spectrophotometer (flame) (cont'd)	Daily	Sensitivity	Aspirate standard solutions into flame and compare sensitivity against previous results.
	Weekly or when required	Nebulizer	Disassemble and wash with distilled water.
	Monthly	Lamp window	Clean with very soft lens wiping cloth.
	Quarterly	Exterior instrument cover	Inspect visually and clean with soap and water as necessary.
	Quarterly	Burner head	Wash in dilute HCl.
	Annually or when specifications are not met	Lamp output	Check with Watt meter.
Atomic absorption spectrophotometer (graphite furnace)		Instrument parameters	See flame AAS.
	Daily	Graphite tube	See operator's manual for model.
	Daily	Argon gas supply	Inspect gauge reading.
	Daily	Temperature compensator	Rotate adjusting pot.
	Daily	Cooling water supply	Inspect visually for correct flow, leaks, and proper operation.
	As required	Graphite cones	See operator's manual for model.
Automated cold vapor analysis of mercury	Daily	Whole system	Check lamp energy. Check for abnormal operation of parts, e.g., autosampler, pump. After use, aspirate the following solutions in the given order:

Table 1. Continued

Instrument	Frequency of checking	Component to check	Procedure
Automated cold vapor analysis of mercury (cont'd)			(1) 3% hydroxylamine SO_4. (2) 10% NaOH for 5 minutes. (3) distilled water for 2 minutes. (4) 5% H_2SO_4 for 5 minutes. (5) distilled water for 5 minutes.
	Monthly	Pump tubes	Replace.
	Monthly	Gas–liquid separator	Clean with 10% NaOH.
	Monthly	UV cell	Clean with water and acetone or water and ethanol.
	Monthly	Column	Clean with 10% NaOH.
	Monthly	Glass coils	Clean with 10% NaOH.
	Quarterly	Transmission tubes	Replace.
Flame photometer (for atomic emission analysis of Na and K)	Daily	Sensitivity	Aspirate standard solutions into the flame and compare sensitivity against previous results.
	Monthly	Pump tubes	Replace pump tubes monthly or sooner if excessively used.
	Semiannually	Transmission lines	Replace reagent, waste, and lines to inspection port.
	Daily	Manifold	Rinse the system for 10 minutes with distilled deionized water and then aspirate air for 2 minutes to dry the flame photometer.
Inductively coupled argon plasma spectrophotometer (ARL Model ICPQ)	Daily	Spray chamber	Rinse with water.
	Daily	Sensitivity check	With 1 ppm standard (As, Se, Sb, or 16 metals).

Table 1. Continued

Instrument	Frequency of checking	Component to check	Procedure
Inductively coupled argon plasma spectrophotometer (ARL Model ICPQ) (cont'd)	Daily to weekly	Profile point for simultaneous spectrometer	Run standard around position using "$ STTS."
	When torch moved or changed	Torch position	Optimize horizontal position using "$ STTS."
	Quarterly	Computer and instrument check	Service contract.
Inductively coupled argon plasma spectrophotometer (ARL Model 3580)	Daily	Instrument parameters, e.g., vacuum, temperature	Type "LOG (R)."
	Daily	Sensitivity check	With 1 ppm standard (16 metals).
	Daily	Spray chamber	Rinse with distilled water.
	Daily to weekly	Profile point for simultaneous spectrometer	Run standard for specified channels using "IPOF" program.
		Position sequential spectrometer	Type "pos" (DO NOT USE COMPUTER UNTIL FINISHED OR 10 MINUTES).
	Weekly	User files (all except .RES & SCN) to floppy disk	Type "BACKUP."
	Weekly	Copy .RES & SCN files to floppy disk	Initialize disk and then copy specified files.
	Quarterly	Computer and instrument check	Service contract.

Table 2. Molecular Spectroscopy Performance Control Checks

Instrument	Frequency of checking	Component to check	Procedure
Technicon auto-analyzer system for total phosphorus, NO_3-NO_2, NO_2, SRP, NH_3, TKN, DIC, DOC	Each run	Accuracy and reproducibility	The accuracy and reproducibility are checked with reference samples in duplicate at the beginning and the end of each tray.
	Each run	Calibration curve	Run a set of standard solutions (two or more) and plot the calibration curve. The reproducibility of the calibration curve should be within the acceptance limits based on standard deviation.
	Biweekly	Pump tubes	Usually the pump tubes are replaced every two weeks, but in some cases they might be replaced earlier. (Refer to flow diagram of each method to replace the pump tubes.)
	Monthly	Colorimeter (optical peaking)	Follow instructions in operator's manual to optically peak the colorimeter.
	Monthly	Interference filters and optical system	The filters and lenses are removed from the colorimeter and cleaned with methanol and lens-cleaning paper.
Spectrophotometer for chlorophyll analysis	Daily	Interferences	One blank and one wash sample are run every eight to 10 samples.
	Weekly	Calibration	Instrument is calibrated with a fresh standard every 300 samples or approximately once a week when in use. Standard is run in triplicate.

Table 3. Chromatography Performance Control Checks

Instrument	Frequency of checking	Component to check	Procedure
Ion chromatograph	Daily	Instrument calibration	Calibrate as per Methods Manual.
	Daily	Accuracy	Run reference or control samples.
	Biweekly	Pump	Oil and inspect springs.
	Biweekly	Column	Check analytical retention times.
	Biweekly	Controller	Run diagnostics on software.
	Biweekly	Detector	Check sensitivity with standard solutions.
	Annually	Whole system	Complete maintenance check.
Gas chromatograph electron capture detector	Daily	Peak area/height resolution reproducibility	Instrument is conditioned with a standard, then calibrated. Standards are repeated every five samples to update calibration.
	Monthly	ECD contamination	Detector output and background monitored for contamination. ECDs are thermally cleaned over the weekend once per month.
	Monthly	Resolution	Known PCB standard injected and compared with previous month.
Gas chromatograph flame ionization detector	Annually or as needed	FID contamination	Detector cleaned manually using HP kit.
	Daily	Peak area/height resolution reproducibility	Instrument is conditioned with a standard, then calibrated. Standards are repeated every five samples to update calibration.

Table 3. Continued

Instrument	Frequency of checking	Component to check	Procedure
High-pressure liquid chromatograph	Daily	Peak area/height resolution reproducibility	Instrument is calibrated using triplicate standards. Further standards are analyzed every three to five samples for recalibration.
	Daily	Contamination	An instrument blank is run prior to calibration.
	Annually or as needed	Contamination/ interference	Guard column is installed in system and changed when necessary or annually.
Fluorescence detector	Monthly	Contamination	Blanks are run through the detector to check for background.
	Annually	Contamination	General cleaning and maintenance of the detector are carried out according to manual.
CHN analyzer	Daily	Interferences	Blanks are run in duplicate every day prior to sample analysis.
	Daily	Peak area/height resolution reproducibility	Instrument is calibrated in duplicate with a known standard every morning before samples are run. A natural standard (tomato leaves) is run prior to samples.
	Monthly	Combustion	The combustion tube is changed every one to two months depending on use.
	Quarterly	Reduction	Reduction tube is changed every two to three months depending on use.
	Annually	Column	The chromatographic column is changed as needed.

Table 3. Continued

Instrument	Frequency of checking	Component to check	Procedure
CHN analyzer (cont'd)	Annually	Precision	Replicates are analyzed (large number) at least yearly to determine precision of instrument and operator

Table 4. Mass Spectrometry Performance Control Checks

Instrument	Frequency of checking	Component to check	Procedure
Mass spectrometer	Daily	Vacuum	Monitor readings on instrument panel; should show high vacuum in manifold of 10^{-6} torr or less with 0.32 mm x 25 m cap column at 10 psi head pressure.
	Daily	Temperature	Monitor readings on instrument panel. Manifold to be set at 100°C, source to be at 15 x 10°C for EI fragmentation.
	Daily	Calibration	Acquire 5 scans of FC-43 calibration gas over a mass range of 60 to 510 amu/1.5 seconds. Run the CALI program, then run the FIT program. If the FIT error is less than 75 mu, the calibration is acceptable.
	Daily	Noise	With no input to the preamp and the preamp zeroed, there will be fewer than five noise spikes seen in a 3-second measurement; each has less than two ADC counts.
	Weekly	Zeroing	With filament source and EM turned off, set acquisition light until it begins to glow. Set data system with maximum sensitivity, minimum area = 0. Acquire file ZERO, monitor the acquisition using the GAIN command.
	Daily	Data system	Data capture set for 100 peaks per scan per second.
	As required	Data system — standards	Optimum conditions set up for compounds of interest when confirmations are required.

Table 4. Continued

Instrument	Frequency of checking	Component to check	Procedure
Mass spectrometer (cont'd)	As required	Data system — NBS	For general screening of full-scan acquisition, NBS library of 38 000 compounds used for comparison.
Ion source filament	As required	Filament	Remove, clean assembly, and weld new filament in place.
Gases	As required	Gas cylinder	Replace when tank pressure drops below 500 psi.
Gas lines	Yearly	Gas lines	Leak check with snoop, tighten connections, or replace fittings to obtain good seal.
Gas chromatograph attachment on mass spectrometer	Daily	Gas	Check pressure and flow rates, adjust as required.
	As required	Injector	Change septum, clean glass liner, check column connection.
	As required	Column	Replace when parameters dictate different phase or when loss of sensitivity, resolution, or excessive column bleed poses problem.
Mass spectrometer (vacuum pumps)	Daily	Oil	Check level and temperature.
	Monthly	Recirculating coolant	Check water level and operation.
	Semiannually	Oil	Change oil in mechanical pumps and check for leaks. Change oil for bearings in turbo pumps.

Table 5. Miscellaneous Instrument Performance Control Checks

Instrument	Frequency of checking	Component to check	Procedure
Balances	Daily	Pan	Clean with brush. If needed, wash with water and dry with paper cloth.
	Daily	Balance level	Ensure balance is level by monitoring bubble in the level adjustment.
	Annually	Accuracy	Calibrate using certified standard weights.
	Annually	System	Service contract.
pH meters	When in use	Accuracy	Calibrate pH meter with at least two buffer solutions. The expected pH values should be as close as possible to that of the buffer.
		Probe	Keep probe submerged in distilled water at all times.

9.0 QUALITY ASSESSMENT FORMS

In Section 3.1, quality assessment is defined as the overall system of activities that ensures that QC is being performed effectively. QC data must be evaluated and audited in a documented form that will allow managers, supervisors, and analysts to make appropriate judgments on the success of the QC activities. The following quality assessment forms are the instruments by which this documentation is carried out at the NWQL. The information from these forms is compiled on a regular basis and used to ensure that an effective QA program is maintained.

NATIONAL WATER QUALITY LABORATORY

SECTION HEAD'S
QUALITY ASSESSMENT FORM
(Semiannual)

Name _____

Section _____

Programs supported _____

No. of projects supported _____

F.Y. _____

P.Y.'s FTC _____ Term _____ Contract _____

Budgets(k) O&M _____ Capital _____ Other _____

1. Identify human resource utilization and distribution of time by type of activity.

Laboratory Area of Analysis	Person Years Utilized					
	Analysis	QC	Meth. Adap.	Field Supp.	Other	Total
Phys./Nut./MI						
Metals						
Organometals						
Biochemicals						
Biocides						
Industrial Chemicals						
Special						
Total						
% of P.Y. Allocation						

2. List frequency of analytical data checked against historical values, and/or detailed analytical audit of samples.

Date	Project No.	Sample No.	Problem Parameter	Action

3. Identify frequency of client complaints about analytical or other services provided by the NWQL.

Date	Parameter/Service	Problem	Action

4. Indicate number of new analytical methods adapted, documented, implemented, or approved in your area of analysis.

Date	Method Name	Adapted	Documented	Implemented	Approved

5. List training sessions attended by your staff, such as seminars, workshops, courses, or conferences.

Date	Type of Activity	Benefit	Recommendation

6. Indicate unsafe laboratory facilities or work practices identified by the safety committee and action taken to resolve them.

Date	Unsafe Facility/Practice	Action

Signature of Section Head

Signature of Supervisor

NATIONAL WATER QUALITY LABORATORY

UNIT HEAD'S
QUALITY ASSESSMENT FORM
(Quarterly)

Name _____

Unit _____

Section _____

F.Y. _____ Quarter: 1 _____ 2 _____ 3 _____ 4 _____ No. of staff _____

1. Name the interlaboratory QC studies in which your laboratory participated.

Study Name	Date	No. of Samples	No. of Tests	Performance

2. List all new analytical methods that were characterized and documented in your laboratory during this period of assessment.

Date	Method

3. List the analytical methods that were adapted and coded during this period of assessment.

Date	Method

4. Indicate instrument breakdowns or malfunctions and actions taken to correct problem.

Instrument	Date	Downtime	Problem	Action Taken

5. List the frequency of verification of analytical results (raw data to final results) as well as analytical data checked against historical values.

Date	Project	Parameter/Lab. No. Range	Problem (if any)	Action Taken

6. What are the multioperator precision and accuracy of the group of tests performed in your unit during this period of assessment?

Test	Precision	Accuracy (% Recovery or SRM)

7. Identify frequency of safety checks of your laboratory facility and actions resulting from them.

Date	Observation	Action Taken

Signature of Unit Head

Signature of Supervisor

NATIONAL WATER QUALITY LABORATORY

ANALYST'S
QUALITY ASSESSMENT FORM
(Monthly)

Name _____

Unit _____

Section _____

F.Y. ____ Month ____ Days Worked ____ Hours Worked _____

1.(a) Indicate the variability of blanks and standards processed during this period of assessment. (Note: The units in which the following are calculated will depend on the area of analysis, e.g., absorbance units, concentration units, specific detector response.)

		Precision (SD)			
Test	Concentration	Blanks	Standards	Internal Spikes	Surrogates

1.(b) Indicate the precision and accuracy of the group of tests you performed on samples.

	Precision			Accuracy (% Recovery and/or SRM)				
Test	Mean	Sample Dupl.	SRM	Mean	High	Low	Found	Theoretical

2.(a) INORGANIC LABORATORIES:

List the number of warning limit violations for the following QC samples in your area of analysis (i.e., ± 2 SD).

Parameter	Spikes	Duplicates	Rechecks	Blanks	SRM/Control Sample

2.(b) ORGANIC LABORATORIES:

Indicate number and type of QC samples carried out with your analyses, and specify number of outlier determinations (± 2 SD).

Parameter	Tests(#)	Blanks	Internal Spikes	Surrogates

3. INORGANIC LABORATORIES:

(a) For which parameter(s) do you use control charts?

(b) How many times did control charts indicate your analytical system to be out of control?

Date	Parameter	Source of Problem	Remedy Taken

4.(a) How often and in what manner did you perform a sensitivity check on your analytical instrument?

4.(b) Indicate the following information on stock and/or intermediate standard solutions you prepared.

Analyte	Date Prepared	% Deviation from Old Standard or % Purity

5. Indicate frequency of laboratory chemicals/reagents found to be impure and action taken to resolve problem.

Date	Chemical/Reagent	Action Taken

6. Describe method and frequency of housecleaning procedures you used in your analytical work (e.g., glassware, instruments, fume hoods, benches, etc.).

Signature of Analyst

Signature of Supervisor

NATIONAL WATER QUALITY LABORATORY

CONTAINER PREPARATION AND SAMPLE INITIALIZATION QUALITY ASSESSMENT FORM (Monthly)

Name _____

Unit _____

Section _____

F.Y. ____ Month ____ Days Worked ____ Hours Worked _____

CONTAINER PREPARATION

1. List number of sample containers prepared this month and sent for QC testing.

Container Type	Containers Prepared	QC Submissions

2. List number of batches (clean containers) rejected from last month's workload, due to inadequate quality.

Container Type	No. of Batches	No. of Containers

SAMPLE INITIALIZATION

1. No. of batches submitted for analysis. _____
2. Total no. of submission forms processed. _____
3. No. of submission forms with incorrect information. _____
4. Project numbers with incorrect data on submission form or with inadequate sample.

Project	Describe Problem	Remedy to Problem

5. Verification of accuracy of sample identification information in data base:

NOTE: This section is to be filled in by the audit officer responsible for verifying the integrity of the sample logging information.

Details of Event	No. of Events	Action Taken

_____ Signature of Employee

_____ Signature of Supervisor

Part IV

A MANUAL FOR EFFECTIVE INTERLABORATORY QUALITY ASSURANCE

K.I. Aspila

Contents

Page

LIST OF CONTRIBUTORS viii

PREFACE x

ACKNOWLEDGMENTS xi

1.0 THE ADMINISTRATION OF QUALITY IV-1
- 1.1 The management process IV-1
- 1.2 The interlaboratory study (external QA) IV-3
- 1.3 Limitations of intralaboratory QC IV-4
- 1.4 References IV-7

2.0 PREREQUISITES FOR AN INTERLABORATORY QA STUDY IV-9
- 2.1 Management issues IV-9
- 2.2 Technical issues IV-9
- 2.3 References IV-10

3.0 DESIGNING AN INTERLABORATORY QA STUDY IV-11
- 3.1 Simple designs IV-11
- 3.2 Multisample designs IV-11
- 3.3 Frequency of studies IV-12
- 3.4 References IV-12

4.0 PREPARATION OF CONTAINERS IV-13
- 4.1 Aqueous samples for inorganic parameters IV-13
- 4.2 Aqueous samples for organic parameters IV-13
- 4.3 Sediment bottles IV-14
- 4.4 Bottles for fish homogenates IV-14
- 4.5 Ampules IV-14
- 4.6 References IV-15

5.0 THE ROLE OF CERTIFIED REFERENCE MATERIALS AND STANDARDS IN EXTERNAL QA PROGRAMS IV-16
- 5.1 The external QA program IV-16
- 5.2 Certified reference materials IV-16
- 5.3 Importance of CRMs IV-17
- 5.4 The CRM program at the National Water Research Institute IV-18
  - 5.4.1 Sediment CRMs for inorganics IV-18
  - 5.4.2 Sediment CRMs for toxic organics IV-21

Contents (Continued)

Page

5.4.3 Aqueous CRMs (inorganic parameters) IV-21
5.5 References IV-21

6.0 PREPARATION OF TEST SAMPLES IV-24
6.1 Precipitation and surface waters (lakes and streams) IV-24
6.2 Preparation of dry sediments IV-27
6.3 Preparation of wet sediments for toxic organics IV-28

7.0 DISTRIBUTION OF THE INTERLABORATORY QA STUDY IV-29
7.1 General IV-29
7.2 Documentation IV-29
7.3 Packaging IV-29
7.3.1 Water samples IV-30
7.3.2 Fish or wildlife reference samples IV-30
7.4 Transport procedures IV-31
7.5 Storage and inventory control IV-31
7.6 References IV-32

8.0 THE NWRI AQC DATA BASE MANAGEMENT SYSTEM IV-33
8.1 AQCPROC (QA procedure file) IV-33
8.2 RAWDAT1 (raw data summary) IV-34
8.3 MEANPRT (summary of data) IV-34
8.4 YOUDN21 (summary of bias and flags) IV-37
8.4.1 Ranking to discern bias IV-37
8.4.2 Flagging results IV-42
8.5 System 2000—data base management system IV-42
8.6 MEDIAN1 (S2K): Interlaboratory median values IV-44
8.7 DIAN2 (S2K): Track record of an RM/CRM IV-48
8.8 LABCOMP (S2K): Performance within a study IV-48
8.9 FLGTBL (S2K): Performance of a group of laboratories based on frequency of bias and flags IV-51
8.10 APPRAIS (S2K): Automated appraisals IV-51
8.11 AQC programs in development IV-56
8.11.1 Precision functions IV-56
8.11.2 Charge balance and conductance ratios IV-58
8.12 References IV-58

Contents (Continued)

Page

9.0 THE FEDERAL—PROVINCIAL AND PRAIRIE PROVINCES WATER BOARD QA PROGRAMS IV-59
9.1 General IV-59
9.2 Study design IV-59
9.3 Data evaluation IV-61
9.3.1 The 10% or one standard deviation rule IV-64
9.3.2 Grubbs' outliers (FP and PPWB QA programs) IV-64
9.4 Laboratory performance IV-64
9.5 Impact of the FP and PPWB QA studies IV-67
9.6 References IV-67

10.0 OTHER EVALUATION TECHNIQUES IV-69
10.1 Outliers IV-69
10.2 The Youden paired-sample approach IV-70
10.3 References IV-70

11.0 THE REPORTING PROCESS (FEEDBACK) IV-73

12.0 REMEDIAL ACTION TO QA REPORT IV-75
12.1 Intralaboratory action IV-75
12.2 Management action IV-75
12.3 Users of data IV-76

GLOSSARY OF NWRI EXTERNAL QA PROGRAMS IV-77

Tables

		Page
1.	External QA programs at NWRI	IV-4
2.	Development and preparation of a CRM	IV-19
3.	Sediment CRMs for toxic organics	IV-20
4.	Sediment CRMs for inorganics	IV-20
5.	Aqueous CRMs for inorganics	IV-20
6.	Preservation of samples (aqueous)	IV-27
7.	A portion of the AQCPROC dictionary	IV-34
8.	RAWDAT1: raw data summary (for a laboratory file)	IV-35
9.	MEANPRT: Data summary for one parameter	IV-36
10.	YOUDN21 output	IV-38
11.	A typical interlaboratory study design	IV-40
12.	An example of ranking results to discern bias	IV-40
13.	Data base definition (System 2000)	IV-46
14.	MEDIAN1: Summary of interlaboratory median values	IV-47
15.	MEDIAN2: Comparison of interlaboratory median values between studies	IV-49
16.	LABCOMP: Comparison of laboratory performance (within a study)	IV-50
17.	FLGTBL: Comparison of laboratory performance over several studies (LRTAP)	IV-52
18.	Criteria used to prepare statements for the automated appraisals	IV-55
19.	Laboratory appraisal (an example)	IV-56
20.	Data report sheet, Federal–Provincial and Prairie Provinces Water Board quality assurance programs	IV-61
21.	Data summary, Federal–Provincial and Prairie Provinces Water Board quality assurance programs	IV-62
22.	Summary of flagged results for the Federal–Provincial QA program	IV-63

Figures

Page

1. A conceptual framework on quality IV-2
2. Some typical types of laboratory performance revealed by external QA studies IV-5
3. Schematic diagram for sample production IV-25
4. Examples of biased data IV-41
5. A typical precision function IV-43
6. An example of flagging criteria IV-43
7. AQC data base (System 2000) IV-45
8. Impact of external QA studies on laboratory performance IV-54
9. Precision function for calcium (LRTAP QA program) IV-57
10. Flow chart of activities to implement the FP and PPWB QA programs IV-60
11. Evaluation of data based on the 10% and one standard deviation rule for FP and PPWB QA programs IV-65
12. An example of a statistical outlier (Grubbs) IV-66
13. Youden paired-sample plot for samples 5 and 6 . . . . . IV-71
14. Comparison of the design value with median values in paired samples 4 and 5 determined by various methodologies IV-72

List of Contributors

Haig Agemian
National Water Quality Laboratory
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

H. Alkema
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

N. Arafat
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

K.I. Aspila
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

V. Cheam
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

James E. Gaskin
Water Quality Branch
Inland Waters Directorate
Conservation & Protection
Environment Canada
Ottawa, Ontario
K1A 0H3

H.B. Lee
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

J. Lee
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

K. Miles
Computing and Programming Services Section
Research Support Division
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

Y.D. Stokker
Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

J. Clark
International Joint Commission
Great Lakes Regional Office
100 Ouellette Ave.
8th Floor
Windsor, Ontario
N9A 6T3

R. White
International Joint Commission
Great Lakes Regional Office
100 Ouellette Ave.
8th Floor
Windsor, Ontario
N9A 6T3

A. Zingaro
Computing and Programming Services Section
Research Support Division
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road
P.O. Box 5050
Burlington, Ontario
L7R 4A6

Preface

A key element in the quality management strategies of environmental programs is external quality assurance (QA). The principal element of external QA is the interlaboratory comparison study. This manual focuses on the interlaboratory study and provides insight on:

- how such studies relate to quality management, QA, and quality control (QC);
- how such studies develop and employ reference materials and certified reference materials; and,
- how effective intercomparison studies are designed, prepared, distributed, and interpreted.

On the issue of interpretation, this manual provides an overview on an analytical QC data base management system that is essential to the administration of large arrays of QA data.

Acknowledgments

We acknowledge Mr. Pat Lively who in the early years of Environment Canada was instrumental in implementing the first interlaboratory studies program.

A sincere thanks is extended to Mr. R.E. White and Dr. John Clark of the International Joint Commission's Great Lakes Regional Office for their development of the computerized techniques for addressing bias in large round robins. A similar thanks is provided to Don King, QA Officer, Ontario Ministry of Environment, for his valuable and indirect contributions and for his support of the external QA programs.

The authors appreciate the patience and support provided by Mr. A.S.Y. Chau, QA Project Chief, and Dr. John Lawrence, Director, Research and Applications Branch, National Water Research Institute.

A special thanks is extended to Dr. J. Gaskin for the initial review and to Ms. Norma Snelling for patiently typing this manuscript.

The authors acknowledge Mr. A.S.Y. Chau for outlining the contents of this manual and for his continual scientific input, discussions, and suggestions during the writing and coordinating of this manual.

A Manual for Effective Interlaboratory Quality Assurance

K.I. Aspila

1.0 THE ADMINISTRATION OF QUALITY

1.1 The Management Process

The basic element contributing to a successful and effective environmental program is a management structure that is receptive to quality-related issues and that partakes actively in the management of quality (Lawrence *et al.* 1982). The organization will be successful only if management is able to direct responsibility, be accountable, and have traceability on all matters pertaining to quality (Hunter 1980). This process is referred to as quality management, and senior management must define and implement a quality management plan in order to be successful (King 1985, 1986). This plan includes the assignment of tasks, protocols, and procedures that verify that the facility can achieve the level of product quality required and that defined quality is indeed being achieved.

An effective quality management plan includes designated quality assurance (QA) officers whose primary role is to ensure management that a QA program is in place (Brossman *et al.* 1982; Hart 1988). This program, implemented by management through a QA implementation plan, is a planned schedule of activities that assures managers that a quality control (QC) program is in place and is being carried out effectively. This QA program within an agency is external to the laboratory and the environmental monitoring and surveillance projects. It is a management program that verifies through a neutral audit process that field, laboratory, and data handling systems can achieve and at all times are achieving their data quality objectives. One specific form of an audit inherent to a QA implementation plan is the interlaboratory study. Such studies are neutral, third-party evaluations of laboratory performance. This manual addresses these interlaboratory studies.

Another element of management's role and responsibility for quality-related issues is the QC program (Agemian 1988). This control program includes all technical efforts practised in the field, laboratory, and data handling systems that verify and document that data quality objectives are being maintained continuously. Such control programs are monitored by the QA office under the directives of the QA implementation plan. In brief, quality control is simply those planned systems of technical activities within a laboratory that document product quality.

A flow diagram depicting the management mechanisms to implement, monitor, verify, and document quality is given in Figure 1.

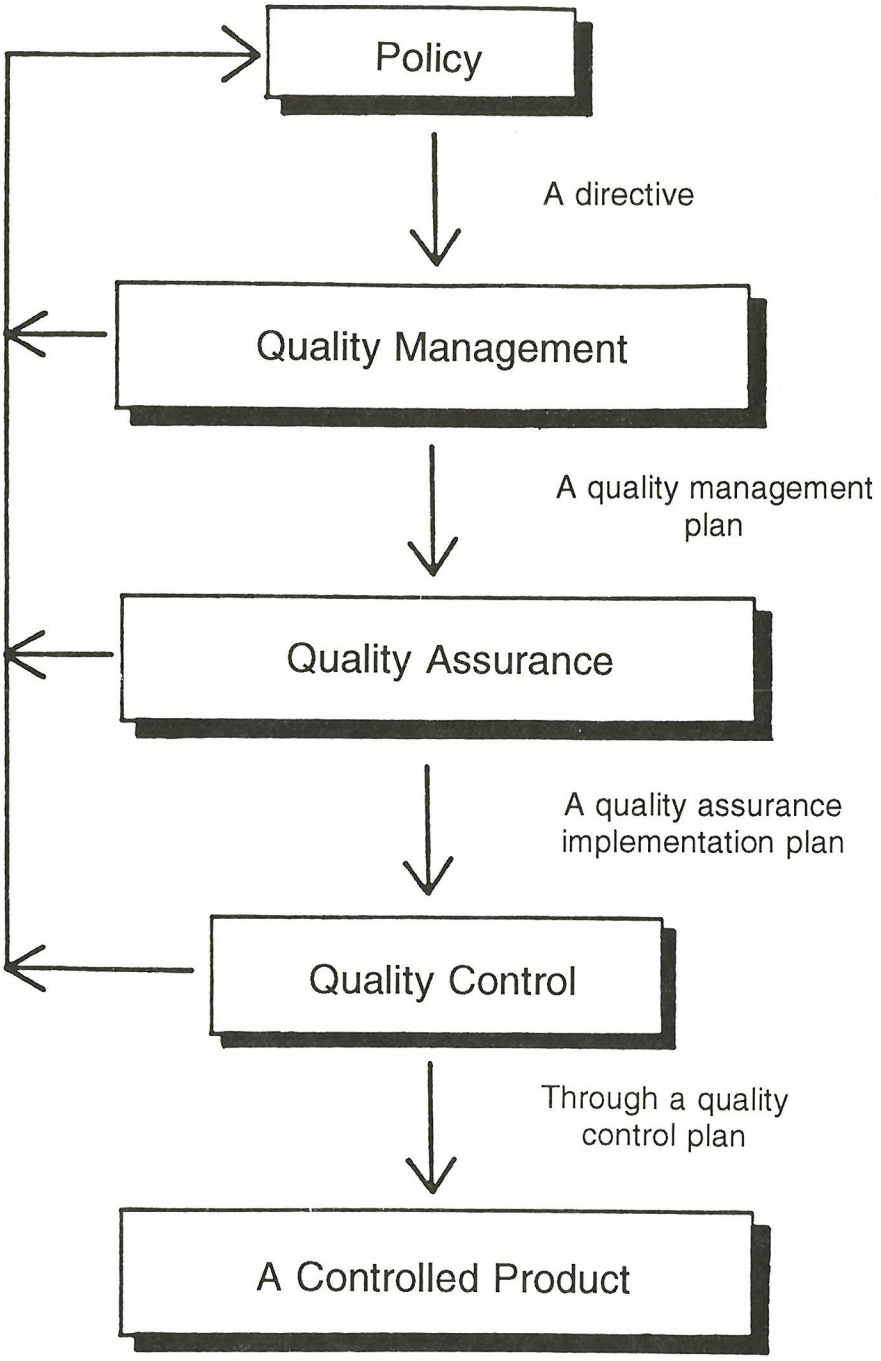


Figure 1. A conceptual framework on quality.

1.2 The Interlaboratory Study (External QA)

An interlaboratory study normally consists of providing an identical set of several test samples to various laboratories for the analysis of specific constituents (Aspila *et al.* 1985). Results reported are analyzed (Youden 1969; Clark 1981), and a report is prepared. If the objectives of the study, its design, and the evaluation techniques are clear, concise, and well thought out, then these reports can be very informative to both the participants and the management they represent (Youden and Steiner 1975; American Society for Testing and Materials 1983).

The design of a study must be carefully established in order to meet the requirements or objectives outlined in the quality management plan. If a laboratory or field method is not well established, the study may be designed to obtain information on how well that method performs by one or more analysts. On the other hand, the studies may be designed to verify if the same or different methods can produce the same results (i.e., are they comparable?). If the methods are comparable, then the design may be established to verify if application of the method is controlled. For complex methods such as those required for trace toxic organics, the design of a study may be specialized to address a certain area, such as the calibration standard or extraction efficiency. These studies are collaborative studies (Youden and Steiner 1975) that evaluate methods and very often require a very detailed format (Taylor 1987).

A successful and well-designed interlaboratory study can provide valuable feedback to analysts, laboratory managers, and data users. For instance, the studies may identify:

- overall precision and bias within a laboratory or between laboratories;
- percent recovery of a constituent;
- erratic performance;
- measurement systems that are out of control;
- measurement systems that have significant baseline errors (poor blank corrections);
- complete failure of a method (not suitable for analysis of substrate);
- operational blunders;
- complete inadequacy of intralaboratory QC;
- inadequacy of internal laboratory standards;

- complete adequacy (or inadequacy) of two or more laboratory measurement systems to allow inference that these two systems will produce compatible data bases that are adequate for interagency use; and

- a neutral, third-party assessment of the overall performance of a laboratory (a vital statistic when contract laboratories are under review by management).

The interlaboratory study is an element of great significance in QA. Under the framework of a quality management plan, interlaboratory study reports must inform the appropriate levels of management on the status of quality and control with the clear understanding that authority and power exist to implement corrective action if performance is substandard. This management process must include the project heads and project managers, whose use of data and whose whole data bases may be affected. Project leaders are assumed to retain close routine liaison with their laboratory data producers. Perhaps the most important information transfer is to the laboratory managers via the interlaboratory study evaluation report. The transfer of performance statements must be swift, as most issues are laboratory measurement issues and, if serious, must be quickly addressed to prevent the data base from being tainted with bad data. Various types of poor and excellent performance in an interlaboratory study are given in Figure 2. These examples present the situations that external QA programs must review, assess, and report on. External QA programs at the National Water Research Institute (NWRI) are highlighted in Table 1.

Table 1. External QA Programs at NWRI

Program*	No. of labs	Clients
LRTAP	102	U.S.-Canada acid rain labs
IJC	140	Great Lakes surveillance
UGLCCS	16	Bi-national (connecting channels)
National	137	Canada (national program)
FP and PPWB	18	Federal-provincial program
FICP (CAPCO)	40	Pesticide labs
Eulerian	8	U.S.-Canada (acid rain)
National Dioxin QA	20	Commercial and federal

* Refer to the glossary at end of Part IV for explanations of NWRI external QA programs.

1.3 Limitations of Intralaboratory QC

A single laboratory working in total isolation is simply unable to verify that its methods' applications, standards, and product output are adequate for its data users. Moreover, it is unable to show comparability with its peers when two or more different laboratories are to merge their data bases. Management at all levels of authority must accept this reality and must react in supporting external QA to authenticate the effectiveness of intralaboratory control measures and allow different laboratories the opportunity of merging their data bases for common use.

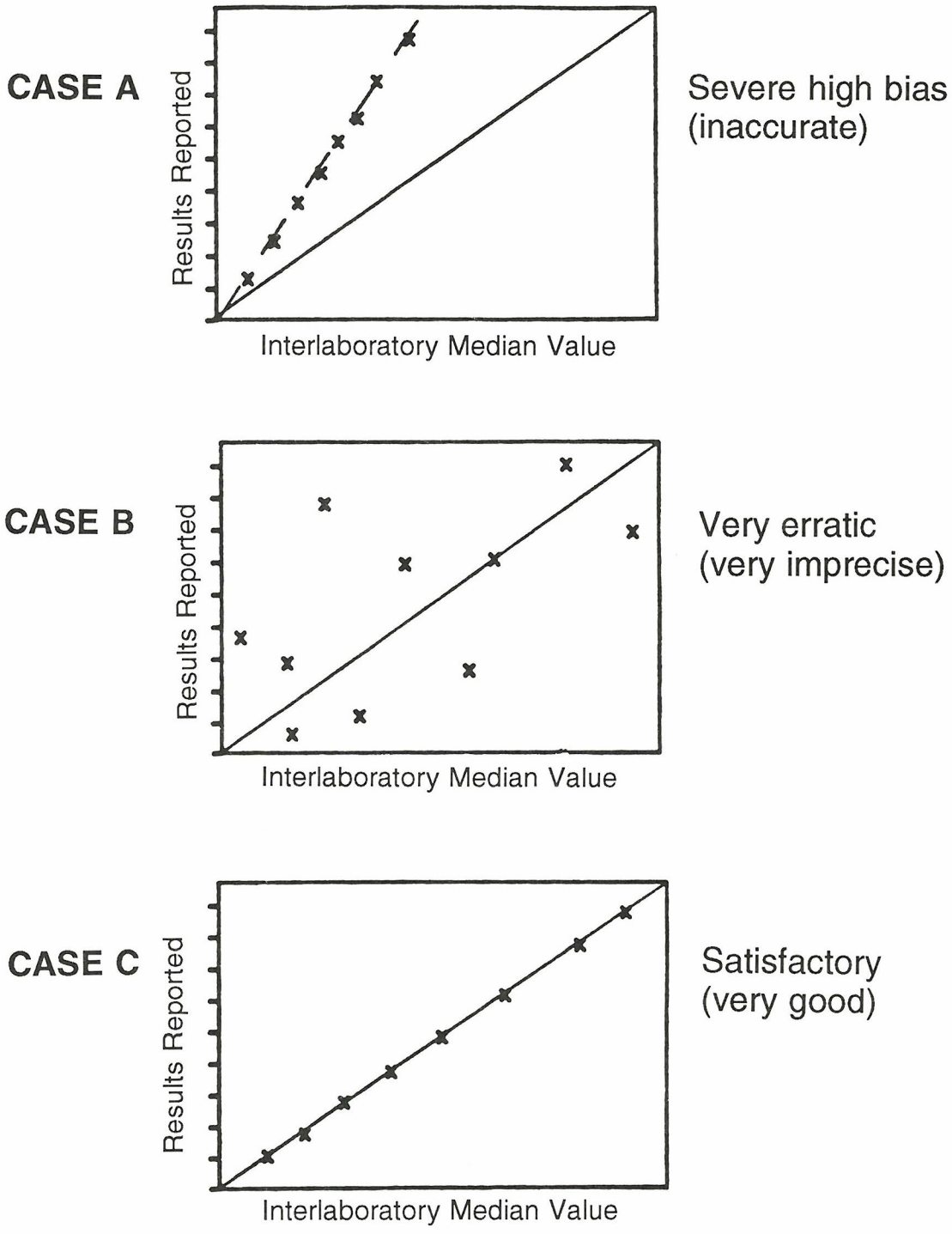
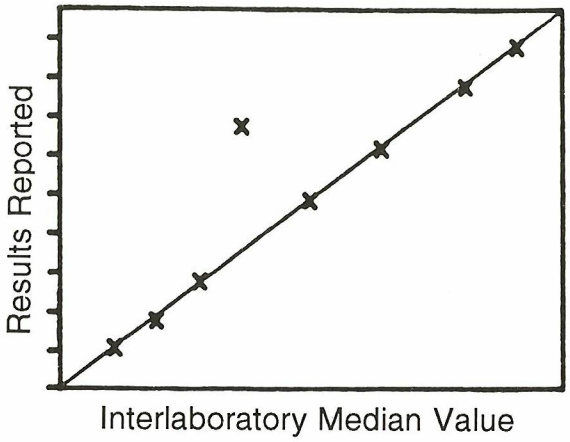


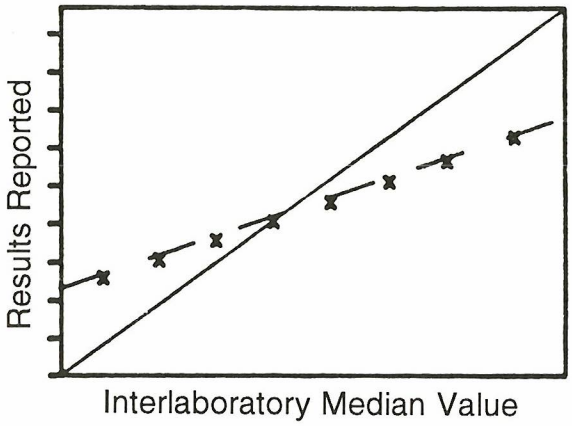
Figure 2. Some typical types of laboratory performance revealed by external QA studies.

CASE D



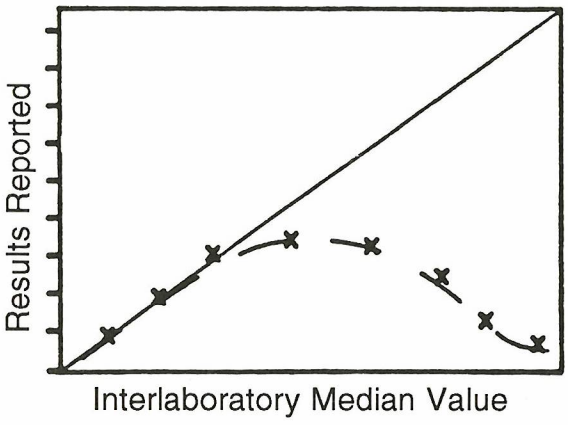
Out of Control

CASE E



Biased with
Blank Problem

CASE F



Methods Failure

Figure 2. Continued.

The most serious failure of an intralaboratory QC program is its inability to authenticate the validity of calibration standards. This is especially true for toxic organics that are prone to solvent evaporation, degradation, and uncertainty of purity of stock standards. Variability of supply and instability of diluted standards are also in question. A major concern is that when the laboratory calibration is erroneous, then so will be the resulting data. The magnitude of the calibration bias may be very small or extremely large but in either case can be swiftly detected by a suitable intercomparison study.

A second serious deficiency of a within-laboratory QC program lies in its inability to address verification of the control on the test sample blanks. For example, if a laboratory addressing a colorimetric analysis of phosphorus uses dirty water to prepare calibration standards, its instrument zero baseline will appear right but will be technically wrong. A simple interlaboratory study using clean water as a blank and very low level standards will quickly reveal any existing anomaly. A clean test sample in an impure matrix will actually yield a negative instrument response and will be translated into a negative concentration on test samples. This case example, which sometimes occurs, serves to reveal the deficiency of relying only on intralaboratory QC, the merits of an intercomparison study, and the valuable contribution made if all laboratories were stimulated into reporting all calculated values whether positive, zero, or negative.

A third major deficiency of the sole reliance on intralaboratory QC is that it is unable to provide management and program data users with any information on the comparability of data between different laboratories. Documented intralaboratory statistics may draw inferences on potential levels of comparability, but it is the demonstrated performances of laboratories that will provide the assurance.

Awarding of analytical contracts should not rely solely on intralaboratory QC information. Management that approves contracts for environmental measurements should remain vigilant and recognize that performance of a laboratory should be assessed on both written documentation as well as demonstrated performance in the analysis of test samples through an ongoing external QA program. This is critical to the contract selection process. The multi-laboratory, multi-sample intercomparison study can quickly reveal peer group performance and verify suitability of a contract laboratory. Sole reliance on intralaboratory information is quite unacceptable.

1.4 References

Agemian, H. 1988. Quality Assurance in the National Water Quality Laboratory. National Water Quality Laboratory, Water Quality Branch, Burlington, Ont.

American Society for Testing and Materials. 1983. Intralaboratory Quality Control Procedures and a Discussion Reporting Low-Level Data, A Standard Practice. (ASTM-D19), D4210-83, Philadelphia, Pa.

Aspila, K.I., R.E. White, and J.L. Clark. 1985. Quality assurance aspects of the International Joint Commission Great Lakes Monitoring Program. Am. Soc. Test. Mater. Spec. Tech. Publ. 867: 407–421.

Brossman, M.W., T.J. Hoogleem, and R.C. Splinter. 1982. Quality assurance project plans: a key to effective cooperative monitoring programs. Am. Soc. Test. Mater. Spec. Tech. Publ. 867: 53–61.

Clark, J.L. 1981. Evaluation of performance of laboratories determining water quality constituents through natural water samples whose true values are unknown. *In* Summary of Conference Presentations, Envirometrics 81, April 8–10, Alexandria, Va., pp. 54–55.

Hart, D. 1988. Quality Assurance Coordination of Great Lakes Surveillance. In a report prepared by Beak Consultants Ltd., Brampton, Ont., for Inland Waters Directorate (Ontario Region), Environment Canada, Burlington, Ont.

Hunter, J.S. 1980. National system of scientific measurement. Science 210: 869–874.

King, D. 1985. Quality Management Plan. Laboratory Services Branch, Ontario Ministry of Environment.

King, D. 1986. Quality Assurance Policy and Guidelines. Laboratory Services Branch, Ontario Ministry of Environment.

Lawrence, J., A.S.Y. Chau, and K.I. Aspila. 1982. Analytical quality assurance: key to reliable environmental data. Can. Res. 15(7): 35–47.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Chelsea, Mich.

Youden, W.J. 1969. Ranking laboratories by round-robin tests. Natl. Bur. Stand. (U.S.) Spec. Publ. 300(1): 165-9–169-13.

Youden, W.J., and E.H. Steiner. 1975. Statistical Manual of the Association of Official Analytical Chemists. AOAC, Washington, D.C.

2.0 PREREQUISITES FOR AN INTERLABORATORY QA STUDY

Prior to the inception, development, and implementation of an interlaboratory QA study, a number of management and technical issues must be addressed.

2.1 Management Issues

When a small- or large-scale environmental program is conceived and implemented, management has the responsibility of assuring that the program yields credible and traceable data (King 1982). One element of the management plan is the external QA program and the associated interlaboratory studies (Hunter 1980; Lawrence *et al.* 1982; Aspila *et al.* 1985). To be successfully implemented, the external QA program must have the complete support of management. Areas that must be addressed include:

- assigning responsible and qualified QA persons to develop the interlaboratory program;
- providing resources and adequate lead time to develop the reference materials essential and pertinent for laboratory analysis;
- providing guidance documents to outline the data quality objectives of the environmental program;
- allowing the QA project personnel developing the study to participate in the subgroups or committees when the environmental program is multijurisdictional; and
- identifying and defining all laboratories and associated managers involved in both the environmental program and the QA assessment process.

2.2 Technical Issues

When the quality management process has been implemented (through the policy directive), the interlaboratory QA infrastructure must be developed and the QA plan implemented. The plan must incorporate the ongoing production of essential reference materials suitable for developing certified reference materials (CRMs), a suitable means for rapid analysis of all reference materials to verify stability, and the necessary logistics to deliver intercomparative studies. When data are returned, the results must be swiftly and correctly interpreted in order to provide timely and essential advice to laboratories, managers, and data users.

On a technical basis, the implementation of an effective QA program requires:

- an adequate budget to develop and deliver QA studies;

- suitable physical resources, such as a cold storage for large volumes of water, soils, sediments, biota, and vegetation;

- a well-equipped and suitable technical support facility to acquire, synthesize, and house large environmental reference materials;

- skilled chemists who are knowledgeable in various aspects of QA and the essential concepts of statistics, computers, and chemistry;

- a well-equipped laboratory system of proven competence that is able to produce and process data in a timely manner;

- a computer facility and network with programming capability and procedures in place for rapid data input and retrieval and timely output of reports; and

- secretarial support to assist in the routine output of reports.

2.3 References

Aspila, K.I., R.E. White, and J.L. Clark. 1985. Quality assurance aspects of the International Joint Commission Great Lakes Monitoring Program. Am. Soc. Test. Mater. Spec. Tech. Publ. 867: 407–421.

Hunter, J.S. 1980. National system of scientific measurement. Science 210: 869–874.

King, D. 1982. Credibility, the Consequence of Quality Assurance. Laboratory Services Branch, Ontario Ministry of Environment, Rexdale, Ont.

Lawrence, J., A.S.Y. Chau, and K.I. Aspila. 1982. Analytical quality assurance: key to reliable environmental data. Can. Res. 15(7): 35–37.

3.0 DESIGNING AN INTERLABORATORY QA STUDY

When management, through the quality management plan, has authorized the development of an intercomparison studies program, considerable care must be taken in its design (Ku 1969; Taylor 1987). It is often necessary to maximize benefits to the client and keep the development costs at a minimum. Studies may be designed simply to verify calibration standards or to address whole analytical systems. They may be designed to intensively compare analytical methodologies or simply to determine the level of comparability of a number of different laboratories that use different methods to supply data to a large program data base. Some studies may have a very simple design, whereas others are quite complex. The decision to select a particular study design is partially influenced by the study objective and the costs related to the study design and operational stages.

3.1 Simple Designs

A study with a simple design contains very few samples and requires a minimum sample workup. Such studies may include:

- a single-ampule standard solution ready for direct analysis;
- several-ampule standard solutions for direct analysis or requiring appropriate dilution;
- one or two natural samples; and
- one or two natural samples so highly characterized as to be classified as CRMs.

Simple studies have merit in that they are relatively inexpensive to prepare and inexpensive for the participating laboratories. Unfortunately, they are not always very informative and may provide little information on the full characteristics of a laboratory's calibration program (e.g., baseline corrections or curvature of the calibration). Simple studies by techniques such as Youden's pairs are able to provide visual information implying the possible source of error (random or systematic error). To this end, they provide inexpensive information at low cost. Their value is improved when the samples are CRMs or true-value calibration standards.

3.2 Multi-sample Designs

The most popular and the most effective study design is usually one that contains many samples. For these studies, the samples selected cover the entire routine concentration range of the equipment used by clients, and each sample contains many constituents. Such designs have been popular in the LRTAP, Eulerian, and IJC Great Lakes QA programs. There are normally 10 samples, most of them having an extensive history of use in previous studies. For each study, several new samples are introduced. When possible, synthetic samples (waters) or fortified samples (e.g., fish homogenates or wet sediments) are included.

The interpretation of data for the multisample approach is more complex and is addressed in Chapter 8.

3.3 Frequency of Studies

When a QA program is being designed to meet the needs of an environmental plan, some thought needs to be given to the frequency of the external QA studies. If the environmental program has a well-developed quality management/assurance/control structure in place, the question is mundane, as the process will provide the necessary assurance. For example, the management, managers, and users of data would have already set their data quality objectives, and the level of protection and the need for verification of proven performance would already have been discussed and verified. Unfortunately, this ideal situation is seldom, if ever, achieved.

New programs involving new laboratories (contract facilities) should always have a plan of action in place before external studies are conducted to verify that the participating laboratories can indeed meet minimum specifications. Too often this is not done, and program staff and users of data are left with less-than-adequate benchmarks to work with.

When a laboratory or several laboratories are engaged in a long-term monitoring program, it is important to verify performance on a continuous basis; for an annual field program, the external laboratory audit performance should be made at the beginning, the middle, and near the end of the analytical season. This may mean three studies per year and, for critical programs, perhaps even monthly audits. Clear evidence of how an external audit program (external QA) can demonstrate performance is given in Chapter 8.

3.4 References

Ku, H.H. (ed.) 1969. Precision measurement and calibration. Natl. Bur. Stand. (U.S.) Spec. Publ. 300(1).

4.0 PREPARATION OF CONTAINERS

An essential prerequisite to the successful preparation of interlaboratory study test samples is the use of sample containers, closures, and equipment that will not influence or contaminate the substrate. For many constituents at ultratrace concentrations, the container can be a source of contamination, and efforts must be made to prevent or limit this possibility. The following is a description of how sample containers should be prepared (Environment Canada 1983; Hunt and Wilson 1986).

4.1 Aqueous Samples for Inorganic Parameters

Containers used for aqueous samples are normally plastic (conventional or high-density polyethylene, polycarbonate, polystyrene, or Teflon). For major ions, nutrients, and physical parameters residing in moderate to hard waters, rigorous cleaning is not always required. New bottles must be subjected to QC checks using distilled water (or distilled deionized water) blanks to verify the absence of contamination. Similarly, for high concentrations (>250 μg/L) of trace metals (e.g., Fe, Mn, Zn, Cu, Ni, Co, Pb, Cd, As), the use of new bottles, without cleaning, may be accceptable, but caution must be exercised.

Containers for soft surface waters and rainfall samples in which the constituent concentrations are usually very low require special attention in the control of blank values. Experience has shown that chromic acid can be adequate for bottle washing. If this powerful acid is used, one must be very meticulous in washing the closures, as acid residues may adhere to the bevelled edges of the closures. A more cautious approach in the preparation of bottles for soft-water samples is to use the bottles after they have been left filled with distilled deionized water for several months. Prior to use, a representative number of bottles must undergo QC testing to ensure that no interfering substance is present. Because bottles are purchased in large quantities, a single production lot should be requested. This action ensures that all bottles have an equal chance of containing the same contaminant, should any be present.

4.2 Aqueous Samples for Organic Parameters

Glass bottles are generally used for studies involving toxic organics in water; the 40-oz. whiskey bottle is normally the bottle of choice. These bottles require rigorous cleaning. The following is a suitable procedure:

(1) Fill bottles with a hot detergent solution and leave for 2 hours (a suitable detergent solution is Liqui-Nox in hot tap water).

(2) Brush the insides of the bottles to dislodge particulates.

(3) Rinse three times with hot tap water to remove all traces of detergent.

(4) Rinse three times with distilled deionized water.

(5) Rinse the inside and outside of bottles with reagent-grade acetone.

(6) Oven-dry the bottles for 2 hours at 200°C.

(7) Plastic closures may be cleaned less rigorously, as the test samples do not come in contact with the plastic because of the use of clean Teflon as a sealing liner.

4.3 Sediment Bottles

Bottles used for sediments that contain inorganic constituents may not require as thorough a cleaning as bottles used for organic parameters. A quick check of a representative set of bottles using a clean acid wash solution may reveal no appreciable metal residue (compared with the potential metal content of the sediment). If there is uncertainty regarding the contamination of the bottles (with respect to the parameters of interest), all bottles should be acid-washed by filling the bottles with nitric acid for a minimum period of 24 hours. Closures should be treated in the same manner. The extent to which the washing is carried out will depend in part on the anticipated level(s) of the parameter of interest and on the requirements of the QA program.

Bottles used as containers for sediment samples destined for toxic organic studies need to be treated in a manner similar to that of the whiskey bottles used for water samples (see Section 4.2). Polyethylene closures lined with solvent-washed aluminum foil are used to prevent contamination from the plastic lining of the closures. The aluminum foil may be washed with two solvents: first acetone, and then hexane. An alternative cleaning process for the foil is oven-drying at 150°C for 12 hours to drive off the organic contaminants.

4.4 Bottles for Fish Homogenates

Bottles for fish (15- to 50-mL ointment jars) are treated in the same manner as sediment containers.

4.5 Ampules

Glass ampules (borosilicate-score break) used for toxic organic standards are generally required in 10 to 100 gross quantities. Their application with respect to injection-ready dilute standards requires that care be taken to ensure they are clean. A suitable cleaning procedure is as follows:

(1) Stand 50 to 100 ampules upright in a large, clean glass beaker.

(2) Carefully fill each ampule with acetone.

(3) Fill the beaker with acetone to ensure all ampules are filled and fully submerged.

(4) Place the beaker into an ultrasonic cleaner for 5 minutes.

(5) Empty the ampules of acetone and dry them in the oven at 200°C for 2 hours.

4.6 References

Environment Canada. 1983. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ont.

Hunt, D.T.E., and A.L. Wilson. 1986. The Chemical Analysis of Water. 2nd ed. The Royal Society of Chemistry, London, U.K.

5.0 THE ROLE OF CERTIFIED REFERENCE MATERIALS AND STANDARDS IN EXTERNAL QA PROGRAMS

Nearly every phase in environmental protection and pollution control depends on the identification and measurement of pollutants in the environment. Millions of dollars are involved in generating analytical data (e.g., expensive sampling trips, manpower, equipment for analysis and data interpretation). There are even greater financial implications when decisions are based on analytical data generated (e.g., sewage treatment process changes, plant modification or construction of new facilities, import and export of food). Indeed, as pointed out by Uriano and Gravatt (1977), "never before have so many critical decisions involving health, safety and economics depended on the quality of chemical analytical data." Therefore, assurance of environmental data quality is an extremely important aspect in the effort to ensure the quality of the environment and health of the public (ACS Committee on Environmental Involvement 1980). QA must be an integral part of analytical and data interpretation activities, as questionable data result in questionable interpretation and subsequently in questionable decisions or conclusions.

5.1 The External QA Program

An effective interlaboratory QA program should involve various activities to assist analytical laboratories in generating reliable data. Distribution of test samples to participants and generation of data reports are two important areas of the program. Interlaboratory QA programs consist of many research and investigative activities, such as sample preservation, sample handling, and analytical methodology. Validated procedures and methods are used in analytical laboratories for interlaboratory QC studies, and a repository for reference analytical standards for distribution to regional analytical laboratories is necessary for the operation of the program. In addition, research is conducted to develop certified reference materials (CRMs) and reference materials (RMs) for both organic and inorganic parameters in environmental substrates. This enhances the effectiveness of the intra- and interlaboratory QC programs.

Development and application of CRMs constitute some of the most important activities in an effective QA program. Without standard or certified reference materials, accuracy of an analytical method cannot be determined (Sherma 1976).

5.2 Certified Reference Materials

CRMs are stable, homogeneous, and well-characterized reference materials prepared in quantity and having essentially identical or very similar matrices to the field program materials in order to eliminate or minimize matrix effects between reference and test samples. The assigned values are obtained by repetitive analysis by several operators and different methodologies in one or more qualified laboratories of known precision and accuracy. Application of CRMs should enable the user to evaluate and calibrate the whole measurement process rather than just a part of it (May *et al.* 1979).

RMs are similar to CRMs except they are less rigorously characterized. They are the forerunners of CRMs.

Ideally, all CRMs should be made from naturally contaminated samples to reflect actual environmental situations. In practice, it is not always feasible to prepare CRMs from naturally contaminated samples, because resource and time restraints often limit the number of CRMs that can be developed and produced at a given time, and because sites of suitable natural samples may not be accessible.

Alternatives, such as a fortified material using natural samples that are not contaminated with the parameters of interest, may be used. However, spiking a matrix has several inherent weaknesses, notably the following:

(1) The spiked compounds may not become well mixed and integrated into the sample matrix.

(2) The recovery of the spike is often not a measure of the true recovery of the endogenous compounds from the real samples. It is quite common for spiked recoveries to be higher than the corresponding recoveries from real samples.

Spiked CRMs using solvent- or reagent-grade samples such as distilled water not only inherit all the pitfalls of spiking but do not take into account the influence of the sample matrix; for example, distilled water is significantly different from natural water in many aspects of water quality. The limitations of spiking have been discussed by many authors. For example, Trautmann (1971) and Browman (1971) reviewed the problem and limitations of spiking, respectively, and Albro (1981) recently reemphasized the weaknesses of spiking to obtain recovery information.

5.3 Importance of CRMs

The importance of suitable CRMs for the assurance of data reliability cannot be overly emphasized. The success of interlaboratory QA programs "depends upon the availability and use of high quality CRM and standard reference samples" (Bretthauer *et al.* 1975). Interlaboratory programs that are not based on CRMs can give a measure of between-laboratory precision only, not accuracy (Yolken 1975). The need to identify both precision and accuracy in the control of a measurement system is well recognized. According to Hunter (1980), no measurement system can be truly under statistical control without measures of both its precision and its intralaboratory bias (a reflection of accuracy).

In an intralaboratory QC program, the application of suitable CRMs as control samples can increase the program's effectiveness. In the absence of CRMs, accuracy of the analysis is determined by spiking the analyte(s) into blank samples and determining their percent recovery. As noted earlier, this widely practised technique is a poor compromise, because spike recovery is not a measure of the recovery of the endogenous compounds. Its use results in misleading information and generates false confidence in both the accuracy of the method used and the data obtained.

Interlaboratory QC studies using suitable CRMs are important in data interpretation (e.g., trend analysis of pollutants, monitoring and surveillance programs involving data from various agencies). Lacking QA and the interlaboratory calibration, data users cannot correlate data between methods or between laboratories. Because of this, data users have found it difficult to identify trends even when such trends exist (Kirch 1974).

5.4 The CRM Program at the National Water Research Institute

Procedures for the preparation of CRMs depend on the type of sample matrix and parameters that are involved. A simplified scheme is outlined in Table 2. In the case of waters and sediments, sample sites are selected and test samples are analyzed to determine background concentration levels. A large sample is then collected. Depending on the requirements and physical characteristics of the sample, the sample is freeze-dried and blended or mixed. Homogeneity of the bulk samples is checked. After subsampling, between- and within-subsample homogeneity is checked by repetitive analyses. The reference values of the parameters to be certified are obtained by replicate analyses using multi/operators, multilaboratories, and a minimum of two independent methods. The precision and accuracy of the chosen analytical methods and analysts are predetermined and monitored during analysis. The total number of analytical data used to generate the design value varies but on the average is about 200.

Research into the development of CRMs (Tables 3 to 5) is one of the more important activities in the QA program at the NWRI. Strategies and activities for organic parameters are different from those for inorganic parameters, and thus the program is divided into two key areas: inorganic CRMs and organic CRMs. Environmental substrates such as water, sediment, and biota are the matrices of interest. Those CRMs that are available through the QA program at NWRI are listed in Tables 3, 4, and 5.

5.4.1 Sediment CRMs for Inorganics

The development and preparation of CRMs for inorganic parameters in sediment are less complex than those for organic parameters, mainly because there are fewer variables to affect the preparation; as well, there are literature analogies on the preparation of rock and soil CRMs and the recently developed National Bureau of Standards (NBS) river sediment for several trace metals. Because of matrix differences, a rock or soil CRM is not completely suitable for sediment work, even when the certified parameters are of the same type and at similar levels. For the same substrate (e.g., sediment), there are considerable matrix variations in different locations resulting from different geological, biological, and human activities. Some analytical measurement systems are sensitive to matrix variation and can also be sensitive to concentration levels. For QA purposes, it is therefore desirable to have at least one CRM with characteristics similar to those encountered in the routine test samples.

Table 2. Development and Preparation of a CRM

Step	Procedure
1	*Preliminary investigations* • spiking and homogenization techniques • long-term storage facilities and stability studies • preservation requirements • selection of containers • freeze-drying, grinding, screening, and subsampling
2	*Selection of sampling site* • sample type, location • resource and space restraints • technical considerations • historical data • in-house analysis to confirm suitable parameters exist in small-scale field study
3	*Sample collection* • arranging • collecting 1000 L of water; 2 tonnes of wet sediment
4	*Sample handling and preparation* • freeze-drying, crushing, homogenization, bottling
5	*Certification* • in-house analysis using two or more standard methods • external analysis using different methods • QA studies • data reduction
6	*Maintaining the registry* • inventory control • monitoring sample integrity • external QA studies • internal QC applications

Table 3. Sediment CRMs for Toxic Organics

CRM/RM identification	Sediment source	Year initiated	Parameters*
EC-1	Hamilton Harbour	1978	PCB, PAH
EC-2	Lake Ontario	1980	PCB†, PAH, Chlorobenzenes†
EC-3	Niagara River plume	1982	PCB†, PAH†, Chlorobenzenes†
EC-4	Toronto Harbour	1983	PCB†, PAH†
EC-5	Humber River	1983	PCB†, PAH†
EC-6	Lake Erie	1986	PCB†, PAH†
EC-7	Lake St. Clair	1987	PCB†, PAH†

* Concentration levels of organic parameters range from 0.01 to 25.0 μg/g and vary for the different CRMs and RMs.

† Certification is in progress for some parameters.

Table 4. Sediment CRMs for Inorganics

CRM/RM identification	Sediment source	Year initiated	Parameters
WQB-1	Lake Ontario	1974	As, Se, Hg
WQB-2	Lake Ontario	1974	As, Se, Hg, trace metals*
WQB-3	Hamilton Harbour	1980	As, Se, Hg, trace metals*
SUD-1	Sudbury	1982	Trace metals*
TH-1	Toronto Harbour	1983	Trace metals*
HR-1	Humber River	1983	Trace metals*

* Certification is in progress for some parameters.

Table 5. Aqueous CRMs for Inorganics

CRM/RM identification	Year initiated	Initial volume (L)	Parameters	Concentration range (mg/L)
CM-ION-91	1981	1000	Major ions	0.05–13.5
CM-ION-92	1980	1000	Major ions	0.03–106
CM-ION-93	1981	1000	Trace metals	0.01–1.0
CM-ION-94*	1984	1000	SO_4	2.8
CM-ION-95	1987	1000	Major ions	

* Coloured water.

5.4.2 Sediment CRMs for Toxic Organics

Unlike the situation for inorganic parameters, development and preparation of CRMs for organic parameters are very complex and require considerable in-house research to provide background information (Chau and Thomson 1979; Chau *et al.* 1979; Chau and Lee 1980; Chau 1983; Cheam and Chau 1984; Lee *et al.* 1986, 1987; Lee and Chau 1987; Cheam *et al.* 1988). There is very little literature on the subject. Furthermore, because samples from different locations may have different matrices, each sample needs to be dealt with individually.

The following are key areas that need to be investigated before a CRM for organics can be developed:

- homogeneous techniques;
- spiking technique (if needed);
- long-term storage conditions;
- preservation techniques; and
- choice of methods and procedures for certification of levels.

Sediment reference materials that are currently certified or undergoing certification are listed in Tables 3 and 4.

5.4.3 Aqueous CRMs (Inorganic Parameters)

A list of several aqueous CRMs is given in Table 5. Most of these samples are from natural sources, and each has been used in several large interlaboratory studies.

5.5 References

ACS Committee on Environmental Involvement. 1980. Guidelines for data acquisition and data quality evaluation in environmental chemistry. Anal. Chem. 53: 2242.

Albro, P.W. 1981. The referee. J. Assoc. Off. Anal. Chem. 4.

Browman, M.G. 1971. Extraction and Analysis of Organochlorine Insecticides from Lake Sediments. Ph.D. Thesis, University of Wisconsin, Madison, Wis.

Bretthauer, E.W., G.B. Morgan, and R.E. Jaquish. 1975. Standard reference materials and environmental monitoring. *In* Standard Reference Materials and Meaningful Measurements. Edited by R.W. Seward. U.S. Government Printing Office, Washington, D.C.

Chau, A.S.Y. 1983. Standard reference materials in environmental quality assurance. Water Qual. Bull. 8: 26.

Chau, A.S.Y., and H.B. Lee. 1980. Analytical reference materials. III. Preparation and homogeneity test of large quantities of wet and dry sediment reference materials for long term polychlorinated biphenyl quality control studies. J. Assoc. Off. Anal. Chem. 63: 947.

Chau, A.S.Y., and R. Thomson. 1979. Analytical reference materials. I. Preparation and purification of photomirex. J. Assoc. Off. Anal. Chem. 62: 1302.

Chau, A.S.Y., J. Carron, and H.B. Lee. 1979. Analytical reference materials. II. Preparation and sample integrity of homogeneous fortified wet sediment for polychlorinated biphenyl quality control studies. J. Assoc. Off. Anal. Chem. 62: 1312.

Cheam, V., and A.S.Y. Chau. 1984. Analytical reference materials. IV. Development and certification of the first Great Lakes sediment reference material for arsenic, selenium and mercury. Analyst 109: 775.

Cheam, V., K.I. Aspila, and A.S.Y. Chau. 1988. Analytical reference materials. VIII. Development and certification of a new Great Lakes sediment reference material for eight trace metals. Natl. Water Res. Inst. Contrib. No. 88-71, April.

Hunter, J.S. 1980. The national system of scientific measurement. Science 210: 869.

Kirch, J.B. 1974. Study of Users' Requirements for Water Quality Branch Data and Recommendations on Improvement of the Data Dissemination Service. Draft report. Water Quality Branch, Inland Waters Directorate, Environment Canada.

Lee, H.B. and A.S.Y. Chau. 1987. Analytical reference materials. VII. Development and certification of a new Great Lakes sediment reference material for total polychlorinated biphenyls. Analyst 112: 37.

Lee, H.B., R.L. Hong-You, and A.S.Y. Chau. 1986. Analytical reference materials. V. Development of a sediment reference material for chlorobenzenes and hexachlorobutadiene. Analyst 111: 81.

Lee, H.B., G. Dookhran, and A.S.Y. Chau. 1987. Analytical reference materials. VI. Development and certification of a sediment reference material for selected polynuclear aromatic hydrocarbons. Analyst 112: 31.

May, W.E., J.M. Brown, S.N. Chesler, F. Guenther, L.R. Hilpert, H.S. Hertz, and S.A. Wise. 1979. Natl. Bur. Stand. (U.S.) Spec. Publ. 519.

Sherma, J. 1976. Manual of Analytical Quality Control for Pesticides and Related Compounds. EPA-600/1-76-017, U.S. Environmental Protection Agency.

Trautmann, W.L. 1971. The Persistence and Analysis of Organochlorine Pesticides in Soils and Sediments. Ph.D. Thesis, University of Wisconsin, Madison, Wis.

Uriano, G.A., and C.C. Gravatt. 1977. The role of reference materials and reference methods in chemical analysis. CRC Crit. Rev. Anal. Chem. 361: 361–411.

Yolken, H.T. 1975. The National Standard Reference Materials Programs in the USA. Natl. Bur. Stand. (U.S.) Spec. Publ. 408.

6.0 PREPARATION OF TEST SAMPLES

A study consisting of 10 test samples and 70 laboratories involves the preparation of almost 1000 individual bottles (or ampules). The primary goal of this effort is to establish a rapid procedure capable of producing identical aliquots so that each laboratory gets a very similar portion of the stock (Fig. 3). For aqueous distilled water standards, organic standards, surface waters, or precipitation (all properly prefiltered using a 0.45-μm filter), the issue of subsampling to yield homogeneous identical test samples is not critical, because all constituents are dissolved and normally quite uniformly dispersed. Dry or wet sediments, fish, and unprocessed natural waters provide special problems. All efforts must be made to verify that all subsamples prepared from a large stock are identical.

Large samples (1 to 10 kg of fish or sediments) and 25 to 200 L of water are normally required for the development of a reference material and should yield 100 to 500 subsamples. It is wise to have a moderate excess, because these test samples are required to monitor stability and homogeneity and can be used in subsequent studies. Larger samples—200 kg of sediments or 3000 L of water—should be considered if the substrate is slated for certification.

The following is a discussion on the procedures required.

6.1 Precipitation and Surface Waters (Lakes and Streams)

The collection of bulk rain waters can be handled by large rain sampling devices (1 to 50 m^2), either large plastic-covered greenhouse rooftops or 1- to 2-m^2 buckets. In preparing large rain water samples, it is more critical to be efficient in the collection of a large sample (50 to 500 L) than to be specific with respect to site selection or sampling protocols.

As with rain, lake or stream water can be collected in bulk by numerous methods, including submersible pumps or plastic pails. Efforts should be made to collect the specific type of water that represents the environmental program being addressed.

Containment vessels (e.g., plastic bottles or tanks) for the storage of bulk reference waters can be large in size (100 to 200 L) and in numbers. Volumes of 200 L are suitable for use over a few years. The collection of 1000- or 2000-L units can prove very useful when there are adequate resources and storage space available.

Once the bulk waters are acquired, several handling procedures are carried out. These include:

- centrifuging the stock to remove particulate matter;
- pasteurizing the water by autoclaving at 80 to 90°C (a temperature sensor is required to verify 80°C has been reached and maintained for about 10 minutes);

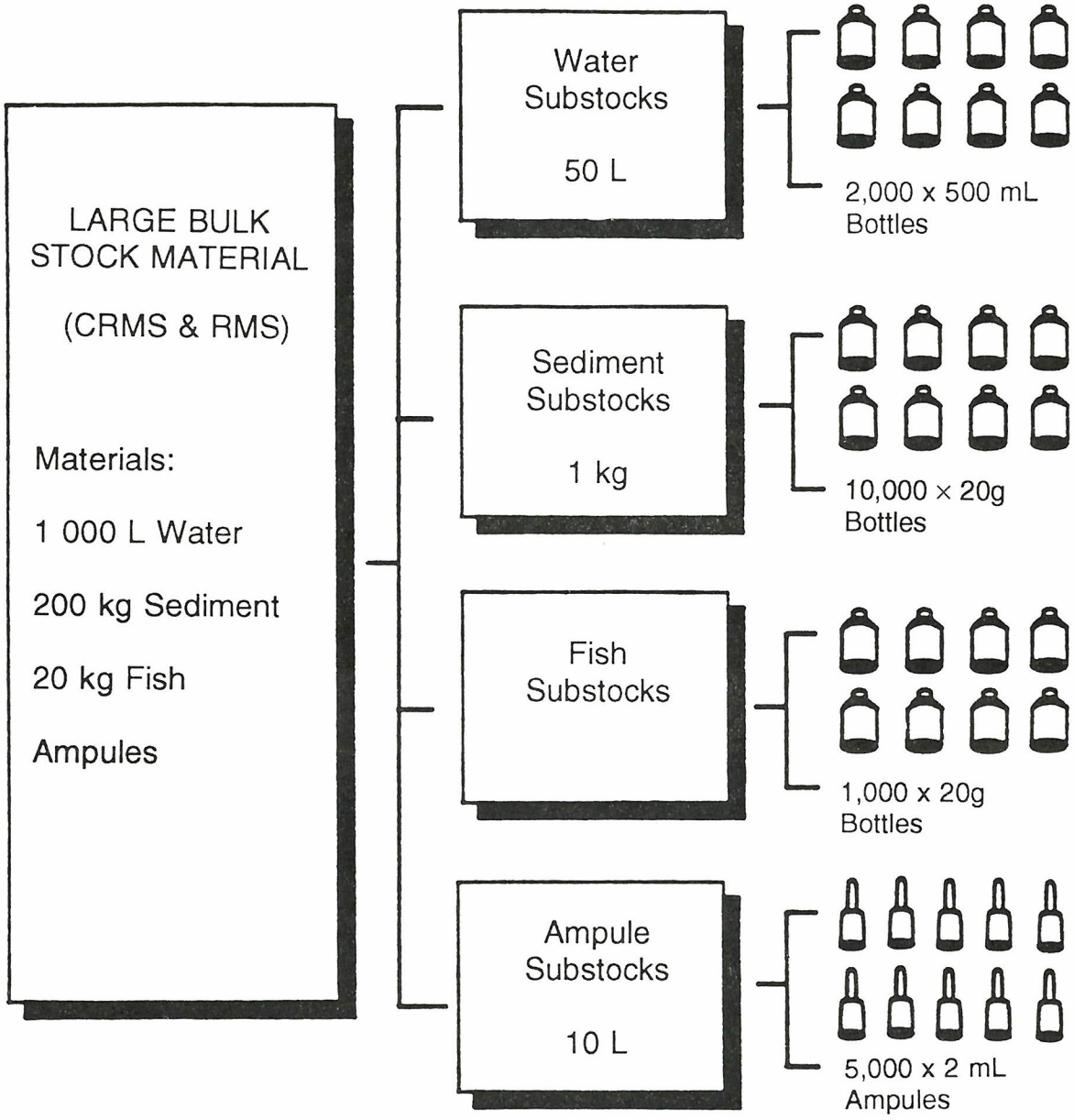


Figure 3. Schematic diagram for sample production.

- combining the specific aliquots of water into the large barrel or tank that will serve as the long-term containment vessel; and

- establishing an analytical monitoring program to verify long-term stability of the chemical constituents.

Pasteurization is not always a critical step but should be considered if there is prior knowledge that bacteria are present that would affect the nutrient equilibrium (ammonia to nitrate or vice versa). Experience in the NWRI's various QA programs over a 15-year period has shown that many clean, unpreserved lake waters may not always require pasteurizing. Some rain waters with very high ammonia content and infested with insects may not need to be pasteurized. An alternative approach is to collect the water sample, centrifuge, and then allow the sample to remain dormant for about six months to a year. This storage time allows bacterial action to reach either completion or a steady-state equilibrium. The degree of processing of water prior to its use relates to its intended use. Experience, knowledge, and project objectives may often provide directional guidance.

Once the bulk waters are documented, stored, and characterized by internal analysis, they are ready for subsampling. For most water studies, the substock required is 40 to 50 L (a clean, 14-gallon container will suffice). There are two approaches that can be used in removing the sample:

- syphoning with a glass tube and clean plastic tubing; or

- rigorously agitating the bulk water with a high-speed stirring device or with a high-volume pump that pumps the water in and out of the tank or barrel.

If the second option is used, the 40- to 50-L substock can be slowly syphoned during the rigorous mixing process. The decision to use either option depends on the absence or presence of particulate matter in the water. A centrifuged water has most constituents of interest in the dissolved phase, and subsampling can be handled by a simple syphon.

When the substock has been transferred and contained in a clean vessel, small aliquots are simply removed by gravity feed into clean empty bottles. This transfer is normally done while the 40- to 50-L stock is mixed rigorously by a stirring device. It remains the choice of the QA person as to whether all test sample bottles are rinsed prior to filling.

The primary objective in subsampling is to have all test samples identical. To achieve this, the analyst should always be aware of airborne contamination from such mundane sources as dandruff (Se, Zn), clothing dust (Zn), deodorants (Al), floor dust, concrete dust, dry skin, and tissue paper (Zn).

Once prepared, the bottles must be labelled and appropriately stored if unstable. If the production run is large, considerable care must be taken to ensure that unlabelled bottles for one sample do not intermingle with unlabelled bottles for another sample.

Knowledge of sample stability is essential when producing aqueous reference samples. Preservatives are often required to ensure stability. A list of some stabilizing agents is given in Table 6.

Table 6. Preservation of Samples (Aqueous)

Constituent	Bottle	Preservative
Trace metals	0.50-L polyethylene	0.2% HNO_3
As and Se	0.50-L polyethylene	0.2% H_2SO_4
Mercury	100-mL flint glass or Pyrex	1% H_2SO_4 + 0.05% Cr_2O_7
Silver	0.5-L amber polyethylene	0.4% EDTA
Li, Be, Sb	0.5-L polyethylene	0.2% HNO_3
Total phosphorus	0.1-L glass	0.2 or 0.3% H_2SO_4
Orthophosphate	0.1-L glass	Sterilize (autoclave), store at 4°C
Nitrilotriacetic acid (NTA)	100-mL polyethylene	1.85% formaldehyde (4°C)
Turbidity	100-mL glass	Store at 4°C
Major ions, nutrients, and physical parameters	0.5-L polyethylene	Store at 4°C

6.2 Preparation of Dry Sediments

The preparation of dry reference sediments in large quantities can be a major undertaking. The following are guidelines:

(1) Identify site location by reviewing all available environmental information.

(2) Identify and assemble resources, such as sampling devices, vessels, and crew.

(3) Collect about 1000 kg of wet sediment (or as much as can be practically handled).

(4) Deep-freeze the collection pails (a large freezer is required) for at least one week.

(5) Drill a large number of small holes into each pail.

(6) Allow each pail to thaw slowly (this can release about 40 to 80% of the water out of the sample and usually takes 3 to 5 days).

(7) Air-dry the sediment with large fans.

(8) Freeze-dry the sediment (a large freeze-drier is required).

(9) Pass the dried sediment through a suitable crushing device (ball mill or roller mill).

(10) Sieve through 200-mesh screens (higher-mesh screens are preferred to enhance homogeneity of the final product).

(11) Homogenize the stock sediment (a large sealed mixing device may be required).

(12) Ensure homogeneity of the bulk material by analyzing various aliquots for the parameter of interest.

(13) Bottle the sediment (an automated dispenser is essential if there are thousands of bottles).

The above relates to processing large reference sediments that have initial weights of at least 100 kg. Smaller samples (10 to 100 kg wet) are less difficult to handle and can be prepared at the laboratory bench. The use of a programmable freeze-dryer is useful to finalize the drying, because it yields a fine powder requiring little crushing for silt or clay sediments.

6.3 Preparation of Wet Sediments for Toxic Organics

Large numbers of wet sediment reference samples are more difficult to prepare than dry sediments because of the separation of suspended sediments from the predominantly aqueous medium.

When preparing wet samples, some experimentation is required to adjust the percentage of water to allow a wet sediment to be mixed and maintained in a uniform and homogeneous slurry. A large blender with variable speed blades is suggested. If the sample materials cannot be maintained in a slurry state, then water must be either removed or added. When successful, one needs to rapidly scoop subsamples into glass vessels until the stock sample is exhausted. Large commercial systems such as those used in the pharmaceutical industry are available for adding water or removing measured aliquots during the mixing process.

When wet or dry sediment reference samples are produced, it is very important to ascertain that the bottled subsamples are all taken from a well-mixed homogeneous stock mixture.

7.0 DISTRIBUTION OF THE INTERLABORATORY QA STUDY

An efficient and cost-effective distribution of a large intercomparison study (100 sample sets) requires a well-organized facility containing appropriate supplies.

7.1 General

Prior to any distribution of interlaboratory samples, a letter of introduction and questionnaires should be sent to all prospective laboratories and management personnel to alert potential participants of a pending study. These should include a brief description of the study and any special treatment to be applied to the samples. All efforts should be made to identify the appropriate date of distribution and a required completion date. For some studies, as in cases of perishable test samples, the delivery date may be firmly set. A deadline for reporting data may also be firm.

7.2 Documentation

Documentation is essential to the interlaboratory study and very often must be prepared several days in advance of the sample distribution date. The paperwork includes covering letters, instruction forms, and report forms for results and for methods-related comments. If the study is method-specific, the paperwork may include specific instructions and procedures that must be followed. For some programs, guidelines may be given on how to report low-level concentrations and the need to report all calculated values.

For large studies, specific documentation for a given laboratory may be replicated and provided with the box of test samples. In such cases, one set of documents is sent to the laboratory manager and one to the analyst; other copies may be sent to appropriate and responsible program managers.

When test samples are distributed internationally, it is essential to prepare documentation and packaging suitable for import and export regulations.

Verification of the receipt of the shipment of test samples in good condition will require the enclosure of a confirming letter and report form with a self-addressed envelope.

7.3 Packaging

The packaging of test samples must be secure and able to withstand considerable abuse. The package should be designed to allow it to be thrown several metres, bounced off walls, or rolled down several flights of stairs. If it withstands these shocks, it will withstand travel by most carriers. For toxic samples, all packages must comply with Workplace Hazardous Materials Implementation System (WHMIS) and Transport of Dangerous Goods regulations.

7.3.1 Water Samples

As a guideline, glass bottles should be packaged in blocks of styrofoam in which holes have been cut to contain individual glass bottles. The blocks of styrofoam should be sealed in plastic. The plastic bag (heat-sealed) will contain the liquid should the glass bottle break. The blocks of styrofoam should then be placed into a larger box and encased on all sides with 2 to 5 cm of loose-fitting styrofoam chips. The chips serve as a buffer if the cardboard box is punctured.

Plastic bottles containing water can be packaged as above. All containers should be sealed tightly and verified as "leakproof." Secure, airtight plastic bags (heat-sealed) are essential for containing any liquid if closures are accidentally broken.

As a precaution, any package that is heat-sealed airtight with a plastic bag should have a small puncture to release air and prevent the package from exploding and disintegrating the box when air transport and depressurization are anticipated. Some cargo planes pressurize only to about 8000 feet in altitude.

The shipment of water samples in the winter, when heated carrier service is unavailable, may require special treatment. For instance, half-full glass bottles may not break, but the chemistry of the liquid in them may change on freezing. These matters may require review in the design and development of the study. Likewise, in the heat of summer, delivery truck temperatures that may exceed 45°C are an issue that must be recognized and verified as not influencing test samples. It is imperative that the QA chemist consult the carrier on all aspects of the anticipated shipping environment to ensure the continual integrity of test samples.

7.3.2 Fish or Wildlife Reference Samples

Whole-fish homogenates must be shipped in the frozen state and must remain frozen during transit to reduce sample degradation. Large freezer boxes can be prepared from foam slabs (5 to 8 cm thick). The custom-made inserts can be used to line large cardboard boxes on all sides. The fish bottles (25 to 100 mL) need to be secured in boxes and placed in the large freezer box. A large number of freezer bags are normally required to retain heat and to keep the small mass of fish frozen. It is advisable that a test run be conducted in which the temperature of the frozen contents is monitored over 72 hours to verify that the box used is adequately insulated and can prevent the product from thawing. A thin thermocouple and recording thermometer (-40 to +10°C) are required.

An alternative to homemade insulated boxes is commercially available boxes. They are, however, much more expensive and will require tests to verify that samples remain frozen over the anticipated travel time.

Coordination of the shipment of frozen biological tissue is essential. These samples must be transported on a specific date to be in total transit time of no more than 3 days, and the recipient must be alerted in advance, as the box must not be allowed to remain unrefrigerated on receipt. These

boxes also require special labelling as "perishable goods" and should have the phone numbers of specific employees visibly recorded on them, so that employees in the shipping and receiving section can alert the laboratory to take special action.

7.4 Transport Procedures

The shipment of test samples (small boxes) may be by mail, via commercial couriers, or sometimes by personal delivery. The decision is normally a balance between cost and time. The mail service is the least expensive but may not be the most appropriate, as there are restrictions on the mailing of some chemicals.

Private couriers are the first choice if resources are available. A single box (1 kg) may cost $5 for local service. Transcontinental service may exceed $200 per box for the larger boxes of fish homogenates. A single study distributed to 50 to 100 laboratories may have a shipping bill of $2000 to $6000. The fee structure is determined by weight and destination and whether ground transport or air transport is required. Perishables should always be sent by air, and the shipment should normally be moved on a Monday or Tuesday to guarantee same-week delivery.

Logistics involved with the transportation of test samples should be discussed with representatives of the carrier service, whether it is the postal service or private couriers. There are restrictions (see references in Section 7.6), and paperwork is required for some carriers. Close consultation is necessary, as the shipper or agent is ultimately responsible for any damage. For example, inappropriate labelling, such as "natural tap water—for cyanide analysis," may cause undue concern and is very misleading. A package containing a steel cylinder encasing a Teflon insert should not be identified as a "Teflon bomb"; this label will unnecessarily bring in the bomb squad.

Transport of toxic chemicals, even at trace concentrations, requires review with all the appropriate authorities (International Civil Aviation Organization 1983).

7.5 Storage and Inventory Control

The distribution of any study (e.g., 10 different samples to over 70 laboratories) will deplete a portion of the stock supply of reference materials and the sample supply prepared for the study. The consumption of the supply of remaining materials must be controlled through cataloguing and the maintenance of inventories.

A study in progress very often requires spare samples as replacement for damaged goods or for special follow-up by laboratories requesting assistance to verify problem areas that may be corrected. The development, maintenance, and inventory of extra sets of samples are wise and precautionary measures.

Chapter 11 addresses a data base management system that allows for a routine administration of inventory of stock materials as well as analytical data.

7.6 References

Amendment to Canada Labour Code. 1987. Statutes of Canada, Chapter 30, Part II. Canadian Government Publishing Centre, Supply and Services Canada, Ottawa, Ont.

Control Products Regulations. SOR/88-64, Canada Gazette, Part II.

Hazardous Products Act. 1987. Statutes of Canada, Chapter 30, Part I.

International Civil Aviation Organization. 1983. Technical Instructions for the Safe Transport of Dangerous Goods by Air. Place de l'aviation internationale, Montreal, Que.

8.0 THE NWRI AQC DATA BASE MANAGEMENT SYSTEM

Over the course of 15 years of external QA development (IJC, National, and LRTAP), K.I. Aspila of the QA group at NWRI, with the programming support of Karon Miles (Computing and Programming Services Section, NWRI), developed a series of software programs to address QA data and issues on the mainframe computer. This software has proved suitable for addressing large studies (40 laboratories, 10 samples, and 20 constituents). The highlight program (YOUDN21) was made possible through the development work of Dr. John Clark and Mr. R. White of the International Joint Commission's Great Lakes Regional Office (Clark 1981; American Society for Testing and Materials 1983; Aspila *et al.* 1985). Through their insight, the nonparametric techniques of Youden for bias assessment (Youden 1969; Youden and Steiner 1975) were developed and computerized to allow the assessment of bias in large data arrays (100 laboratories, 100 samples).

As the QA programs at NWRI increased in size and in number, it became apparent that transfer of information between programs and different studies would be required and indeed was essential. The original software produced single outputs (flat files) for studies in isolation. To improve the retrieval of study-to-study data and to make long-term projections more efficient, Karon Miles developed the essential data base structure on System 2000 (S2K) (a commercial data base software package). This chapter describes many evaluation procedures now recognized as essential for large external QA studies.

8.1 AQCPROC (QA Procedure File)

AQCPROC is the acronym for "analytical quality control procedures." In brief, it contains the Network Operating System (NOS) commands that are needed to execute many different software packages maintained on the NWRI mainframe computer (Cyber 180/310A). Several of the essential computer programs used for handling larger studies for the IJC, Ocean Dumping, Eulerian, LRTAP terrestrial, and LRTAP aquatic QA programs are described in the following sections. Example outputs are presented.

In the early phase of software development, the AQCPROC had flat file outputs. The output results of computer programs such as RAWDAT1, MEANPRT, and YOUDN21 were produced in isolation. They have since been extended with S2K retrievals to produce information across different studies using programs such as MEDIAN2 and FLGTBL. These programs have proved very useful and are highlighted in the following sections.

None of these programs is possible without a dictionary to translate parameter codes into memories. A small portion of the dictionary is given in Table 7. It is structured to have elements listed in ascending atomic numbers.

Table 7. A Portion of the AQCPROC Dictionary

Code	Parameter	Shortform	Units
07092	Nitrate + nitrite	$NO_3 + NO_2$	mg N/L
07102	Organic nitrogen	ORGAN N	% N
07192	Ammonia	NH_3	mg N/L
07202	Nitrate	NO_3	µg N/g
07292	Total nitrogen UV	TOT. N UV	mg N/L
07392	Total Kjeldahl nitrogen	TKN	mg N/L
07492	Nitrate-IC	NO_3	mg N/L
07592	Nitrate-non IC	NO_3	mg N/L
08002	Oxygen	O	% O
09002	Fluorine	F	µg F/g
09092	Fluoride	F	mg/L
1100P	Sodium in plants	Na (ODW)	µg/g (ODW)
11001	Sodium	Na	% Na
11091	Sodium	Na	mg/L
11092	Sodium in sediments	Na	µg/g
1200P	Magnesium in plants	Mg (ODW)	mg Mg/g (ODW)
12001	Magnesium	Mg	% Mg
12091	Magnesium	Mg	mg/L
12092	Magnesium in sediments	Mg	µg/g
1300P	Aluminum in plants	Al (ODW)	µg/g (ODW)
13001	Aluminum	Al	% Al

8.2 RAWDAT1 (Raw Data Summary)

This is the first program that is initiated after the results have been entered. An example of output is shown in Table 8. The primary role of RAWDAT1 is to search for errors in the data files. No output is achieved until all laboratory and parameter codes, method summary, and data values are correctly in place. Once verified, the output (Table 8) is produced. This output can then be proofread for transcription errors. A nominal study of 60 laboratories yields 60 pages of data (one page for each laboratory).

Future plans to effect data entry with the use of electronic scanners or electronic mail systems may reduce input errors. Scanners that are currently available will be able to read pages of data directly into electronic files located in personal computers, and those data can then be downloaded into the mainframe data base. Other options available to streamline data entry make use of floppy diskettes created by each laboratory for reading data directly into a personal computer and then into the mainframe computer. These methods are currently under review.

8.3 MEANPRT (Summary of Data)

The MEANPRT program provides the type of output shown in Table 9. Its primary use is to overview raw data in a convenient summary format during the course of a study and, perhaps more importantly, to scan the raw data for blunders (e.g., wrong units) due to either input errors

Table 8. RAWDAT1: Raw Data Summary (for a Laboratory File)

STUDY: INTERLABORATORY STUDY NO. 20 MAJOR IONS IN WATER
PRINTOUT PREPARED: 89/04/12
LAB: L004

			SAMPLE NO.									
Parameter	Code	Method	1	2	3	4	5	6	7	8	9	10
Colour	Hazen unit	02022	3	31	3	31	0	0	0	60	3	6
Spec. cond.	μmho/cm	02041	10.4	17.9	60.0	67.2	12.9	10.4	8.9	22.4	34.2	52.8
pH units	pH units	10301	5.2	6.7	4.3	6.4	5.0	5.5	5.1	5.8	6.4	7.2
DOC	mg C/L	06104	0.4	3.0	1.1	3.6	0.2	0.3	0.2	4.7	1.8	2.4
Alk.-FE (pH 4.5)	mg/L $CaCO_3$	10101		3.0		2.2				0.7	2.6	12.4
Gran. alk.	mg/L $CaCO_3$	10110	0.08	3.35	-2.59	2.48	-0.48	0.08	-0.28	0.82	2.88	12.89
DIC	mg C/L	06180		0.4		0.2				0.3	0.4	2.2
$NO_3 + NO_2$	mg N/L	07112	0.12	0.13	1.75	0.27	0.08	0.26	0.18	0.04	0.15	0.11
NH_3	mg N/L	07555	0.073	0.007	0.250	0.004	0.050	0.208	0.029	0.057	0.018	0.029
Na	mg/L	11105	0.67	1.08	0.16	5.7	0.79	0.13		0.49	0.94	0.91
Mg	mg/L	12102	0.07	0.31	0.92	0.71	0.05	0.07	0.15	0.45	0.69	1.56
Reac. Si	mg Si/L	14105	0.05	1.94	0.22	2.54	0.02	0.03	0.02	0.51	0.77	1.04
SO_4 (non-IC)	mg/L	16306	0.9	3.8	7.0	9.2	1.4	2.2	1.9	7.8	8.2	91
Cl (non-IC)	mg/L	17203	1.22	1.34	4.30	11.1	1.39	0.27	0.09	0.64	1.34	1.03
K	mg/L	19102	0.08	0.14	0.14	0.73				0.28	0.48	0.67
Ca	mg/L	20103	0.18	1.64	3.99	4.48	0.43	0.72	0.52	2.12	3.17	5.8

Table 9. MEANPRT: Data Summary for One Parameter

LRTAP STUDY NO. 20 MAJOR IONS IN WATER
PRINTOUT PREPARED: 89/03/14
PARAMETER: CHLORINE NON IC METHODE (mg/L)

	Sample results									
Lab	1	2	3	4	5	6	7	8	9	10
L002	1.2	1.4	4.4	10.9	1.4	0.3			1.3	1.0
L004	1.22	1.34	4.30	11.1	1.39	0.27	0.09	0.64	1.34	1.03
L006	1.26	1.34	4.34	11.30	1.41	0.316	0.088	0.762	1.35	1.06
L010	1.1	1.2	3.8	10.7	1.2	T 0.5	T 0.5	0.7	1.3	1.0
L023	1.20	1.40	4.30	10	1.40	0.35	W 0.10	0.80	1.30	1.10
L024	1.17	1.06	4.22	11.00	1.35	0.26	0.08	0.33	1.21	0.89
L025	1.0	1.0	3.80	11.0	1.3	T 0.2	W 0.1	T 0.6	1.2	T 0.9
L045	1.19	1.22	4.18	10.9	1.48	0.30	0.10	0.64	1.32	0.95
L049	1.2	1.2	4.0	11.1	1.7	0.5	0.2	0.5	1.2	0.9
L063	1.10	1.20	4.20	11.3	1.27	0.28	0.13	0.67	1.20	0.90
L067	1.0	1.1	3.9	11.0	1.2	0.3	T 0.1	0.6	1.1	0.8
L085	1.09	1.02	4.01	11.5	1.27	0.27	0.07	0.37	1.18	0.89
L086	1.0	1.2	3.8	10.5	1.6	0.4	W 0.2	0.6	1.2	0.9
L089	1.11	1.16	4.30	10.7	1.30	0.30	0.06	0.33	1.17	0.91
Total labs reporting	14	14	14	14	14	14	14	14	14	14
Total labs used	14	14	14	14	14	14	14	14	14	14
Mean	1.13143	1.20286	4.11071	10.92587	1.37643	0.32471	0.14180	0.58015	1.24071	0.94500
Std. dev.	0.08743	0.13100	0.21925	0.37505	0.14281	0.08683	0.13190	0.15490	0.07580	0.08178
Median	1.14000	1.20000	4.19000	11.00000	1.37000	0.30000	0.09500	0.60000	1.20500	0.90500

Note: The T code has the following meaning: "Value reported is less than criterion of detection." The use of this code warns the data user that the individual datum with which it is associated does not, in the judgment of the laboratory that did the analysis, differ significantly from 0.

The W code has the following meaning: "Value observed in less than lowest value reportable under T code." This code is used when a positive value is not observed or calculated for a result. In these cases, the lowest reportable value is observable is reported with the W (American Society for Testing and Materials 1983).

or transcription errors by the analyst. The combination of output from RAWDAT1 and MEANPRT facilitates a smooth search for errors prior to finalizing a report to analysts and updating results in the data base.

8.4 YOUDN21 (Summary of Bias and Flags)

This procedure program, although simple to initiate as a batch job, is by far the most sophisticated in construction. An example of the output is given in Table 10. For a 20-parameter study involving 50 laboratories, the output can approach 80 to 100 pages. The example in Table 10 is for only one parameter for a small group of laboratories.

The key elements in the YOUDN21 output are discerning whether or not a laboratory data set is biased and whether an individual result reported by a laboratory is sufficiently different from the median to warrant the flagging of the individual result. The following is a brief overview on bias and flagging.

8.4.1 Ranking to Discern Bias

The Youden bias assessment technique is a nonparametric process in which a matrix of results (e.g., 10 samples, 50 laboratories) is converted into a matrix of ranks. Each sample (with, say, 50 results) is ranked such that the lowest result is assigned rank 1, the second lowest is assigned rank 2, and so on. The highest result is rank 50 if there are 50 laboratories. When laboratories report equal values, then the rank assigned is an average. Examples are provided by Youden (1969) (see also Tables 10, 11, and 12).

The next step in the ranking process is to review the total laboratory rank (sum the ranks) or the average rank. The immediate impact is recognition that some laboratories have a unique ability to rank very high or very low. The question to resolve is whether these anomalous high or low ranks are rare events (less than 5% chance of occurring). To evaluate if bias exists, one needs to use a traditional hypothesis test. First, it is assumed that no bias exists. The next step is to calculate the probability of total ranks from the matrix of ranks (e.g., 10 by 50). This calculation (found in gambling handbooks) is synonymous to calculating the probability of scores when 10 dice (samples) are thrown and each dice has 50 sides (50 laboratories). The probabilities of interest are the very high and very low scores. When extreme scores (very high or low ranks) are found in the matrix of ranks with occurrence probabilities of less than 5% of the time, then the null hypothesis is rejected and the laboratory data set is declared biased. The risk in declaring a laboratory biased when it is not is one chance in 20 (5%).

A description of this process is given in Tables 10 to 12. The example is derived from a LRTAP study. Youden's original work (Youden 1969) describes total ranks for which a matrix of critical ranks was calculated manually. The probability calculations described above and developed by Clark are parallel to those of Youden. Both methods provide very informative statements when appraising interlaboratory results.

Table 10. YOUNDN21 Output

PARAMETER 16001, SULPHATES NON IC METHODS
LRTAP STUDY NO. 20 MAJOR IONS IN WATER

LOWER LIMIT FOR USE OF BASIC ACCEPTABLE ERROR = 2.00 BASIC ACCEPTABLE ERROR = 0.50 CONCENTRATION ERROR INCREMENT = 0.2000
LABORATORIES YET TO REPORT: L014C, L009, L022, L052, L098
LABORATORY RESULTS OMITTED ARE NONE

	Sample*																	
	1		2		3		4		5		6		7		8		9		10	
Lab No.	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank	Reported value	Rank
L002	0.9	3.5		0.00		0.00		0.00	1.8	9.50	1.8	2.50	1.7	3.50	6.6	7.00	7.9	5.00	8.5	3.50
L003	1.0	9.00	2.1	4.00	6.5	3.50	7.7	3.00	1.6	7.00	1.95	6.00	1.7	3.50	6.1	4.00	8.05	6.00	8.6	5.50
L004	0.9	3.50	3.8 EH	9.00	7.0	7.00	9.2	8.00	1.4	4.00	2.2	10.00	1.9	8.50	7.8	10.00	8.2	8.50	91 EH	10.00
L006	0.984	7.00	2.22	6.00	6.84	6.00	7.72	4.00	1.62	8.00	1.93	5.00	1.73	5.00	6.76	8.00	8.30	10.00	8.94	9.00
L023	0.95	5.00	2.85 H	8.00	6.45	2.00	7.55	2.00	1.50	5.00	2.00	8.00	1.80	7.00	6.45	6.00	7.80	4.00	8.50	3.50
L049	1.0	9.00	1.5 L	1.00	4.8 L	1.00	13.5 EH	9.00	1.3	2.00	1.5	1.00	1.5	1.00	4.0 VL	1.00	5.2 VL	1.00	7.0	1.00
L063	0.983	6.00	1.920	3.00	6.647	5.00	8.380	6.00	1.323	3.00	1.830	4.00	1.733	6.00	6.050	3.00	7.600	2.00	8.320	2.00
L067	1.0	9.00	2.3	7.00	7.2	8.50	7.3	1.00	1.8	9.50	2.0	8.00	2.00	10.00	6.4	5.00	8.2	8.50	8.8	8.00
L085	0.852	2.00	1.68	2.00	7.2	8.50	7.74	5.00	1.59	6.00	1.80	2.50	1.62	2.00	5.94	2.00	8.13	7.00	8.64	7.00
L086	0.8	1.00	2.2	5.00	6.5	3.50	8.5	7.00	1.2	1.00	2.0	8.00	1.9	8.50	6.8	9.00	7.7	3.00	8.6	5.50
Median Conc.	0.967		2.200		6.647		7.740		1.545		1.940		1.732		6.425		7.975		8.600	

Table 10. Continued.

(In order of lab code)

Lab No.	Total rank	Average rank	No. of samples ranked	Summary of flagging		Method coding
L002	34.50	4.929	7			
L003	51.50	5.150	10			16304
L004	78.50	7.850	10	EH EH	Biaised high	16306
L006	68.00	6.800	10			Technicon
L023	50.50	5.050	10	H		Methylthymol
L049	27.00	2.700	10	L L EH VL VL	Biased low	Colorimetric
L063	40.00	4.000	10			Turbidimetric
L067	74.50	7.450	10			Auto MIB
L085	44.00	4.400	10			ICPS
L086	51.50	5.150	10			AA-MIB
Overall average:		5.361				

(In order of total rank)

Lab No.	Total rank	Average rank	No. of samples ranked	Summary of flagging		Method coding
L049	27.00	2.700	10	L L EH VL VL	Biased low	Colorimetrie
L063	40.00	4.000	10			Turbidimetric
L085	44.00	4.400	10			ICPS
L002	34.50	4.929	7			
L023	50.50	5.050	10	H		Methylthymol
L003	51.50	5.150	10			16304
L086	51.50	5.150	10			AA-MIB
L006	68.00	6.800	10			Technicon
L067	74.50	7.450	10			Auto MIB
L004	78.50	7.850	10	EH EH	Biased low	16306
Overall average:		5.361				

* EH = extremely high; H = high; L = low; VL = very low.

Table 11. A Typical Interlaboratory Study Design

Lab code	Sample No. 1	2	3	9	10	n
A						
B						
C			e.g., 10 samples			
D			56 laboratories			
.			23 different constituents			
.						
K			concentrations range from 1 to 100 units			
.						
M						

Table 12. An Example of Ranking Results to Discern Bias

Lab code	Rank on sample results 1	2	3	. . .	8	9	10	Total rank	Average rank
A	12	18	41		20	26	28	250	25
B	43	58	49		45	57	59	550	55
C	3	2	1		5	4	2	20	2
.	.	.	.		.	.	.	.	.
.	.	.	.		.	.	.	.	.
M	23	22	20		21	24	25	240	24

Note: Lowest result (on a sample) = rank 1;
second lowest result = 2; and
highest result is highest rank = no. of labs.

Nonparametric techniques are powerful procedures for discerning small systematic errors in calibrations. In some cases, the difference is valid but is so slight that some laboratories are unable to react and adjust their calibration to remove the slight difference between their standards and the error implied from the interlaboratory study evaluation. Some laboratories that achieve considerable precision and a statistical control in determinations associated with blanks and secondary blanks have had the ability to adjust calibration 2 or 3% using backup verification from standard reference materials (e.g., for NO_3 and SO_4). Gross bias (10 to 30% error) is sometimes so severe in some laboratories that bias assessment by Youden's ranking method does not need to be applied, because a graphical format or simple review of the matrix of results is visually adequate (Fig. 4).

Assessment of bias by the Youden method is not always informative. A typical set of results in Figure 4 serves as an example for discussion. The data set in this example has a blank problem, and the calibration is also biased. To correct these shortcomings, the performance of a laboratory needs to be complemented by a flagging system or by an inference using graphics as a visual aid.

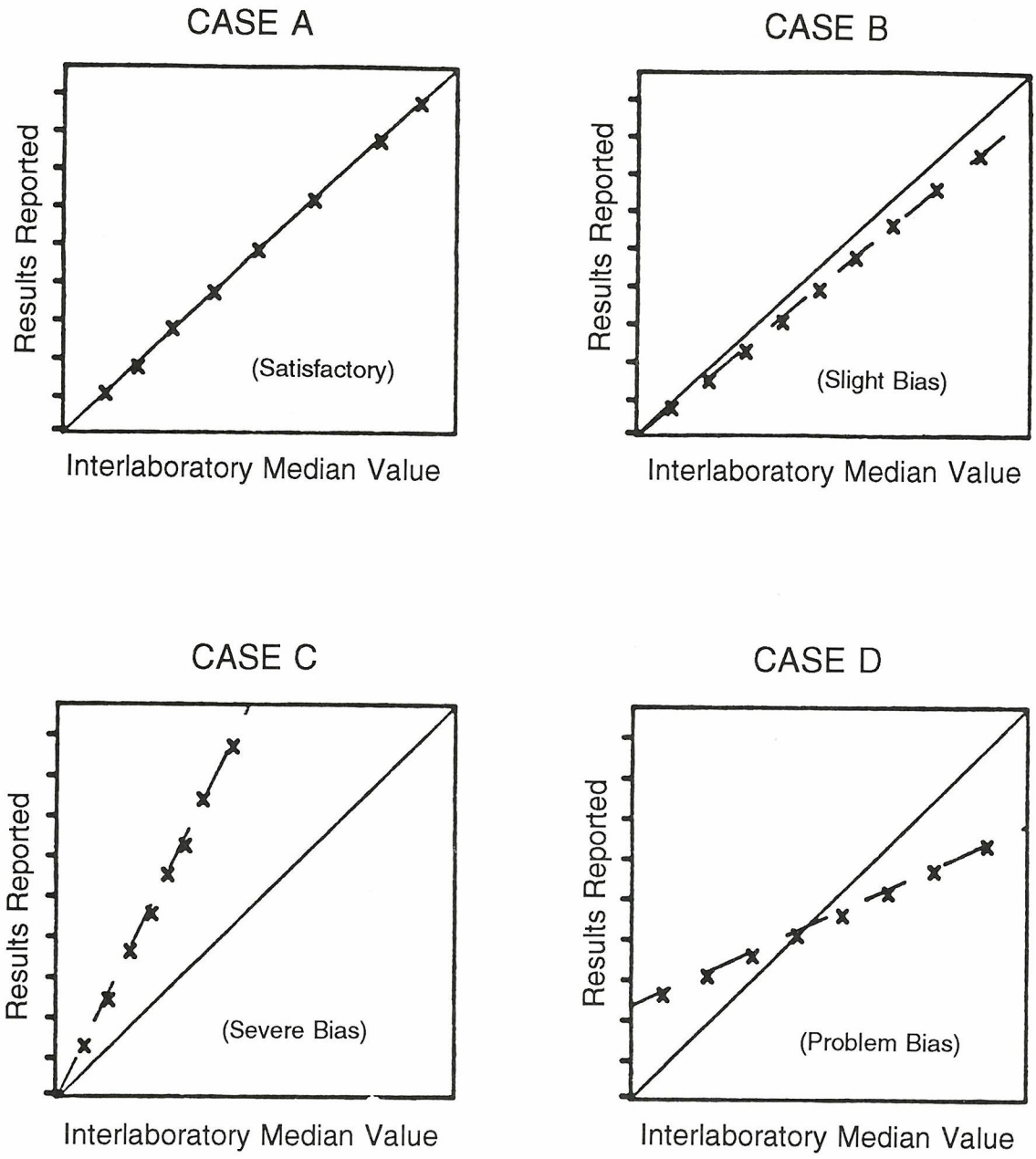


Figure 4. Examples of biased data.

8.4.2 Flagging Results

The Youden bias assessment in many large studies can successfully address and discern the presence of inaccuracy in the laboratory measurement process. The rigour with which this method identifies inaccuracies is of course clouded when there is a serious blank error or if the entire group of laboratories is in error. It is rare that an entire group of laboratories is wrong (for large studies), but vigilance and review must be maintained when difficult substrates and constituents are under review (e.g., toxic organics in fish or sediments).

To complement bias assessment, large or small studies can use a flagging procedure that identifies a laboratory result as very high or very low. The flagging process and the bias assessment are two different and separate evaluation procedures. Flagging is critical, because some laboratories are imprecise; as such, the degree of bias cannot be easily determined, as there are, on average, very high and very low results. Figure 4 provides examples.

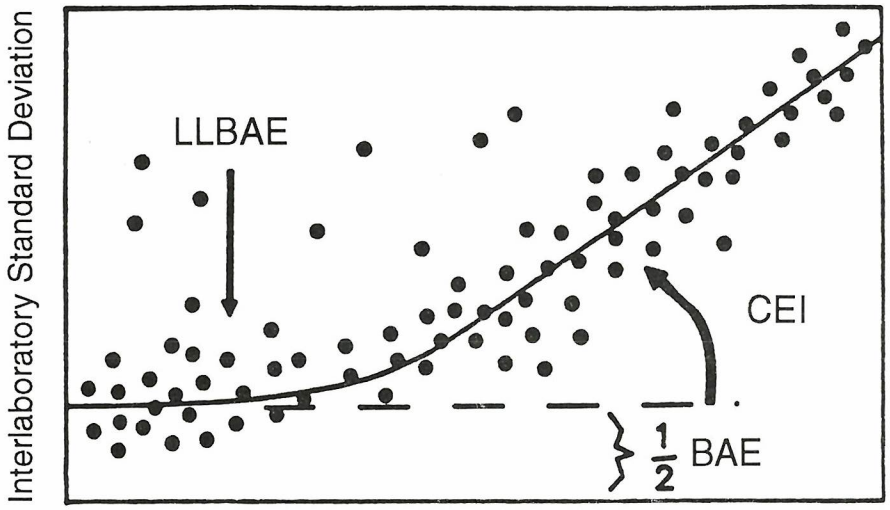
A formula to flag individual results on a sample within a study has been developed for many traditional constituents. Experience has shown that within any study covering a concentration range of one or two orders of magnitude, the interlaboratory standard deviation varies and increases almost linearly from low concentration to high concentration (Fig. 5).

Figure 5 allows for construction of a simple formula for flagging. Three variables are required to decide if a reported result deviates sufficiently from an interlaboratory median to warrant a flag (high or low). The first is the basic acceptable error (BAE), which is the allowable deviation fixed over all concentrations. The second is the lower limit for use of basic acceptable error (LLBAE), which is the concentration at which the acceptable deviation (reported result minus the median) begins to increase. The rate of increase, similar to the slope of the precision function (Fig. 5), is referred to as the concentration error increment (CEI). These three variables are given in Figure 6.

The observed precision function and the flagging formula are closely related. The principal issue to resolve is the values assigned to the BAE, LLBAE, and CEI. Some trial and error may be required if the correct precision function is unknown. The median is chosen as a target, because medians are more robust than average or mean values, which are often influenced by extreme results. Criteria chosen can be adjusted so that some (10 to 30%) of all results reported are flagged either H (high) or L (low). When results are very different, they can be flagged VH (very high) or VL (very low). These extremes can be arbitrarily recognized when the results reported are more than one and a half times the acceptable deviation. A third flag—EL (extremely low) or EH (extremely high)—is assigned if the deviation is more than two times the acceptable deviation.

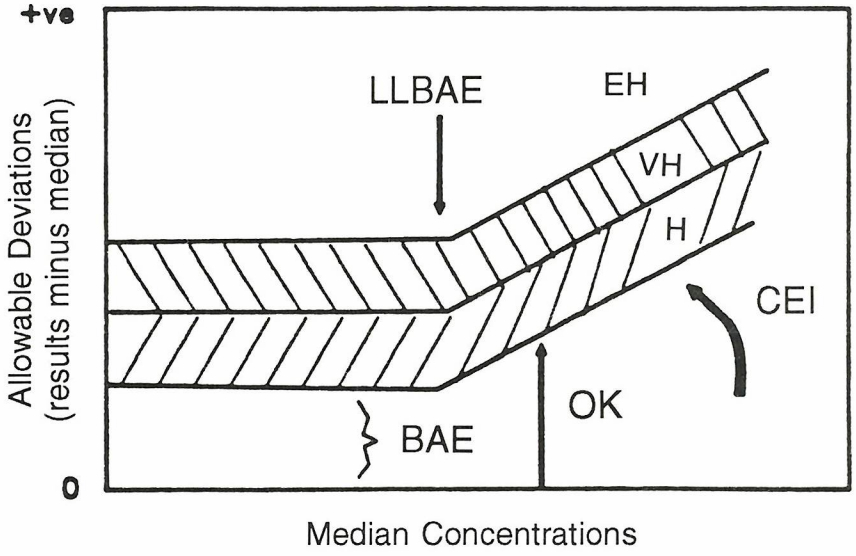
8.5 System 2000—Data Base Management System

The three procedure programs (RAWDAT1, MEANPRT, and YOUDN21) provide a rapid data assessment for large or small studies. They have been and will continue to be utilized for large NWRI



BAE = Basic Acceptable Error
LLBAE = Lower Limit for Use of BAE
CEI = Concentration Error Increment

Figure 5. A typical precision function.



NOTE: Flags assigned are EH, VH, H

Figure 6. An example of flagging criteria

studies (error search). This flat file processing is, however, unable to provide rapid electronic study-to-study information. The S2K data base management system was designed to allow a search for almost every conceivable piece of information entered as data and much of the calculated information generated by the very successful YOUDN21 program. It was recognized early in the system's development that one could create files to:

- confirm swiftly the stability of reference samples (RMs and CRMs);
- seek specific information to identify laboratory performance (bias and flags) over all studies;
- obtain hard evidence very swiftly on improvements or change in laboratory performance; and
- make accessible all data or calculated data entered for any laboratory on any study over all programs.

The structure of the data base is given graphically in Figure 7 and in narrative format in Table 13. The components on laboratory-specific information (e.g., names, addresses, phone numbers) are included to create a fully automated report-writing system in which covering and informational letters (sometimes as many as 200 letters) can be swiftly generated and in which appraisals and a vast amount of support data (hundreds of pages per laboratory) are handled. This part of the system relating to merging, letters, names, appraisals, and support data is still under development.

The AQC data management system is now being used extensively in the NWRI QA programs (LRTAP, LRTAPP, Eulerian, Ocean Dumping, and for some elements of the National and International Joint Commission QA programs). The data base is populated when RAWDAT1 creates the data format, a program PLSUPDA makes the program, study, and laboratory information available, and the program DATAUP updates the data onto the S2K data base. All pertinent data can then be accessed and retrieved. Various programs are outlined in the following sections to illustrate some key outputs.

The inherent value of the AQC data base is in its administrative control of the vast amount of data generated, be it on large reference materials that are transient (two to five studies) or in large reference materials set for certification (10 to 100 studies). The data base currently has about a quarter of a million QA results and within a few years should contain several million individual laboratory data.

8.6 MEDIAN1 (S2K): Interlaboratory Median Values

This program is simple and is performed after a study is completed or whenever a summary copy of an earlier study is required. The program simply abstracts interlaboratory median values for all parameters on all samples for a specified study. A typical output is given in Table 14. Although partly administrative, the program serves a value in providing a capsule summary of data for a large study and, internally, a quick means of selecting previous samples when preparing for a new study. This type of output is provided in study reports.

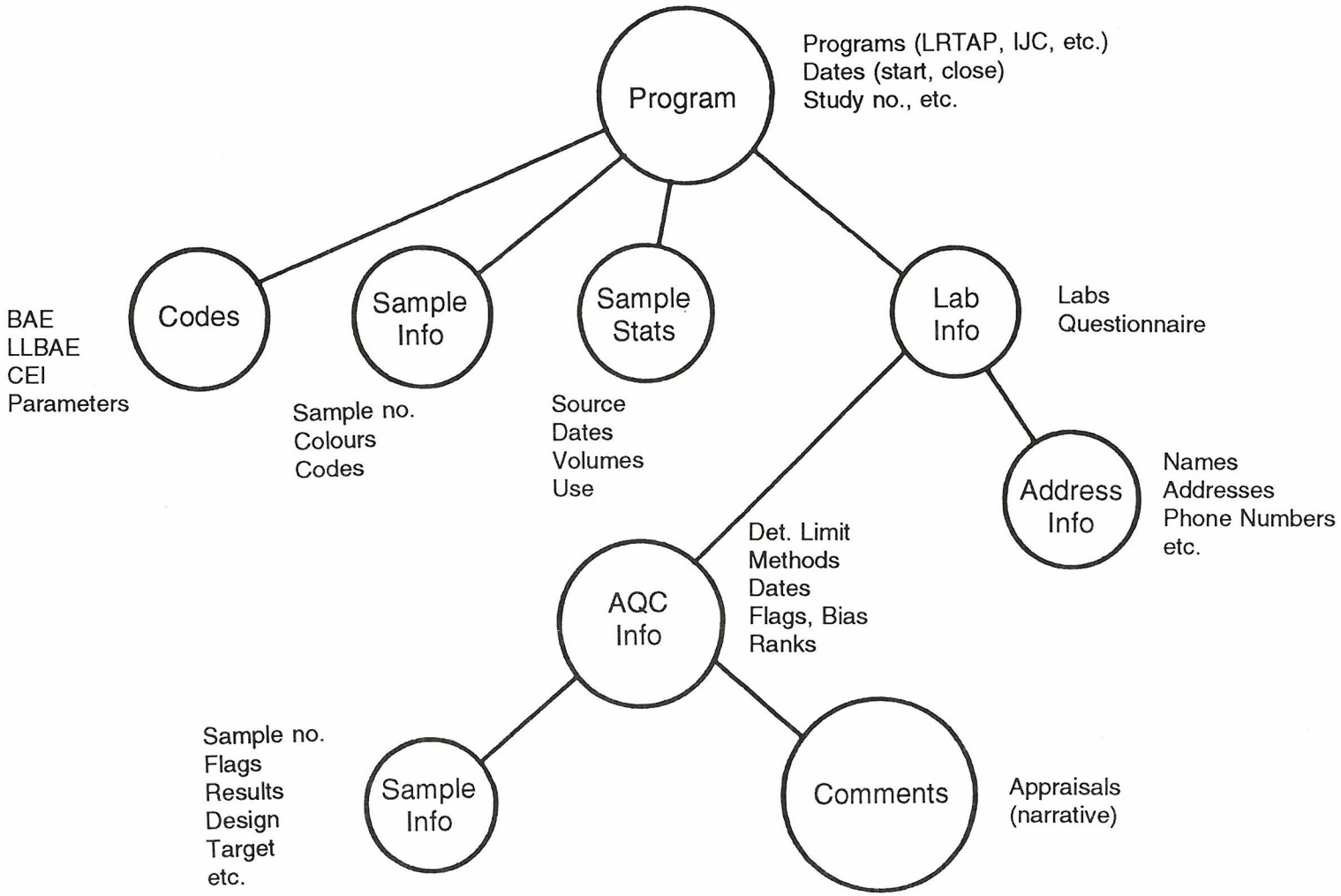


Figure 7. AQC Data Base (System 2000).

Table 13. Data Base Definition (System 2000)

```
1*      PROGRAM (CHAR. X(10));
2*      STUDY (NON-KEY CHAR X (40));
3*      STUDY CODE (CHAR XXXX);
4*      STARTDATE (NON-KEY DATE);
5*      STOPDATE (NON-KEY DATE)
100*    CODE INFO (RECORD);
101*    CODEA (NON-KEY CHAR X(5) IN 100);
102*    BAE (NON-KEY DECIMAL NUMBER 999.999 IN 100);
103*    CEI (NON-KEY DECIMAL NUMBER 999.999 IN 100);
104*    LLBAE (NON-KEY DECIMAL NUMBER 999.999 IN 100);
105*    AVER RANK1 (NON-KEY DECIMAL NUMBER 9999.999 IN 100);
200*    SAMPLE INVEN (RECORD);
201*    SAMPLE NO. (NON-KEY INTEGER NUMBER 99 IN 200);
202*    QAMS CODE (NON-KEY CHAR X(12) IN 200);
203*    SAMPLE COLOUR ID (NON-KEY CHAR X (10) IN 200);
250*    SAMPLE STATS (RECORD);
251*    QCODE (NON-KEY CHAR X(12) IN 250);
252*    SOURCE (NON-KEY CHAR X(50) IN 250);
253*    DATE OF PREPARATION (NON-KEY DATE IN 250);
254*    INITIAL VOLUME (NON-KEY DECIMAL NUMBER 9999.99 IN 250);
255*    VOLUME ON HAND (NON-KEY DECIMAL NUMBER 9999.99 IN 250);
256*    UNITS (NON-KEY CHAR XXX IN 250);
300*    LAB INFO (RECORD);
301*    LAB CODE (CHAR X(5) IN 300);
302*    DATE QUESTIONNAIRE (NON-KEY DATE IN 300);
303*    TAKING PART (NON-KEY CHART X IN 300);
400*    AQC INFO (RECORD IN 300);
401*    PARM CODE (CHAR X(5) IN 400);
402*    DETECTION STATED (NON-KEY CHAR X(10) IN 400);
403*    METHOD (NON-KEY CHAR X(18) IN 400);
404*    DATE OF ENTRY (NON-KEY DATE IN 400);
405*    OTHER INFO (NON-KEY CHAR X(18) IN 400);
406*    BIAS FLAG (NON-KEY TEXT X(42) IN 400);
407*    BIAS STATEMENT (CHAR XXXX IN 400);
408*    TOTAL RANK (NON-KEY DECIMAL NUMBER 999.99 IN 400);
409*    NO. OF SAMPLES RANK (NON-KEY INTEGER NUMBER 99 IN 400);
500*    SAMPLE INFO (RECORD IN 400);
501*    SAMPLE NUMBER (NON-KEY INTEGER NUMBER 99 IN 500);
502*    FLAG (NON-KEY CHAR X IN 500);
503*    VALUE (NON-KEY X(8) IN 500);
504*    NO. OF DECIMALS (NON-KEY INTEGER NUMBER 9 IN 500);
505*    REPLICATE SAMPLE FLAG (NON-KEY INTEGER NUMBER 99 IN 500);
506*    DESIGNED VALUE (NON-KEY CHAR X(7) IN 500);
507*    MEDIAN (NON-KEY CHAR X(9) IN 500);
600*    COMMENTS (RECORD IN 400);
601*    LINE NUMBER (NON-KEY CHAR XX IN 600);
602*    NARRATIVE (NON-KEY CHAR X(58) IN 600);
700*    ADDRESS INFO (RECORD IN 300);
701*    ADDRESS NUMBER (CHAR XX IN 700);
702*    ADDRESS TYPE (CHAR XXXX IN 700);
703*    NAME (NON-KEY CHAR X(36) IN 700);
704*    TITLE (NON-KEY CHAR X(36) IN 700);
705*    AFFILIATION1 (NON-KEY CHAR X(36) IN 700);
706*    AFFILIATION2 (NON-KEY CHAR X(36) IN 700);
707*    ADDRESS (NON-KEY CHAR X(36) IN 700);
708*    CITY PROVINCE POST CODE (NON-KEY CHAR X(36) IN 700);
709*    COUNTRY (NON-KEY CHAR X(20) IN 700);
710*    PHONE NO. (NON-KEY CHAR X(14) IN 700);
711*    AFFILKEY (CHAR X(10) IN 700);
800*    SUBCON (RECORD IN 700);
801*    SUBSTRATE CONSTITUENT (NON-KEY CHAR X(10) IN 800);
```

Table 14. MEDIAN1: Summary of interlaboratory Median Values

Parameter		Sample Number									
		LR-PRV-13 Sample 1	LR-SSW-20 Sample 2	LR-PRC-04 Sample 3	LR-SSW-16 Sample 4	EU-07 Sample 5	EU-08 Sample 6	EU-09 Sample 7	AUD-04 Sample 8	AUD-05 Sample 9	AUD-06 Sample 10
Colour	Hazen unit	3.000	34.000	3.000	32.000	1.750	1.000	1.000	56.500	5.000	8.000
Specific conductance	µmho/cm	10.500	17.550	62.320	68.000	13.330	10.600	9.010	22.250	34.250	53.100
Gran. acidity	mg/L $CaCo_3$	1.370	1.860	4.220	1.590	1.760	1.060	1.700	1.870	1.190	1.040
Acidity to pH 8.3	mg/L $CaCO_3$	2.325	2.678	5.800	2.765	2.367	2.118	2.310	3.310	2.018	2.230
pH	pH units	5.173	6.600	4.293	6.400	4.985	5.505	5.165	5.830	6.580	7.189
Dissolved org. carbon	mg C/L	0.500	3.920	1.400	4.720	0.200	0.224	0.271	6.180	1.370	3.400
Alkalinity - fixed endpt. pH 4.5	mg/L $CaCO_3$	1.400	4.500	0.000	4.000	1.100	1.700	1.400	2.350	4.200	13.800
Alkalinity - gran. inflec, extrap.	mg/L $CaCO_3$	0.000	3.200	0.000	2.460	0.000	0.030	0.000	0.870	2.790	12.880
Gran. titra alk.	mg/L $CaCO_3$	-0.240	3.245	-2.575	2.570	-0.490	0.060	-0.265	0.940	2.880	12.880
Dissolved inorg. carbon	mg C/L	0.220	0.983	0.285	0.740	0.296	0.328	0.241	0.446	0.956	3.295
Nitrate + nitrite	mg N/L	0.110	0.124	1.539	0.249	0.070	0.250	0.160	0.027	0.120	0.083
Ammonia	mg N/L	0.079	0.008	0.290	0.010	0.051	0.232	0.022	0.040	0.019	0.022
Total Kjeldahl nitrogen	mg N/L	0.154	0.121	0.470	0.236	0.100	0.300	0.135	0.390	0.192	0.232
Sodium	mg/L	0.710	1.183	0.185	6.026	0.860	0.150	0.020	0.543	1.031	0.990
Magnesium	mg/L	0.82	0.310	0.954	0.730	0.055	0.086	0.162	0.476	0.750	1.613
Reactive silica	mg Si/L	0.053	1.940	0.230	2.550	0.022	0.039	0.027	0.520	0.770	1.050
Sulphate IC method	mg/L	0.954	1.510	6.920	7.421	1.630	1.920	1.743	5.6700	7.975	8.420
Sulphate non IC methods	mg/L	0.967	2.200	6.647	7.740	1.545	1.940	1.732	6.425	7.975	8.600
Chloride IC method	mg/L	1.220	1.140	4.100	11.000	1.390	0.280	0.062	0.356	1.260	0.940
Chloride non IC methods	mg/L	1.140	1.200	4.190	11.000	1.370	0.300	0.095	0.600	1.205	0.905
Potassium	mg/L	0.110	0.173	0.178	0.791	0.012	0.030	0.022	0.320	0.550	0.730
Calcium	mg/L	0.196	1.600	4.000	4.447	0.430	0.717	0.540	2.085	3.155	5.770

8.7 DIAN2 (S2K): Track Record of an RM/CRM

This program has exceptional value in its ability to administer a continually updated track record on the characteristics of a CRM or RM. The NWRI has many QA programs that involve or have involved many hundreds of reference materials (e.g., water, rain, sediments, vegetation, fish). Maintaining a swift electronic data summary of how each sample has performed over time is essential. A typical output is given in Table 15. The output is now available on command within minutes.

The usefulness of this tracking program is apparent when one has a data base containing over a million results for several external QA programs employing many samples. Besides being administratively useful, the program can provide a summary output that quickly confirms the stability of data over time. Knowledge and hard evidence on the stability of constituent concentrations are critical if any external QA program is to be successful.

8.8 LABCOMP (S2K): Performance Within a Study

The output from this program is derived by searching the S2K data base. An example output is given in Table 16. Its primary purpose is to provide each laboratory with a precise statement on its relative performance with its peer group within a study. It is particularly useful for large studies involving many laboratories that analyze many different constituents.

The program can isolate and also respond to all or any group parameters. A very wide choice of outputs is available. The example output (Table 16) is for a LRTAP study, and the footnotes indicate those parameters used and those that were excluded when this table was created. This program option is particularly useful when a request is made to compare a particular contract laboratory with existing program laboratories for a specific series of constituents.

The program LABCOMP ranks laboratory performance and provides a score. This score is the summation of the percentage of parameters biased and percentage of results flagged. A very low score is indicative of superior performance, whereas a very high score is indicative of poor performance.

Performance is accepted in LABCOMP as quite relative. It includes bias (which reflects accuracy) and precision (indicated by many flags). Laboratories that are severely imprecise will, if the flagging process is correctly established, have as many as half their results flagged (any flag H, L, VL, VH is counted). If half their data are flagged, the laboratories' score will be 50%. On the other hand, if a laboratory is precisely inaccurate (no flags), it may be frequently discerned as biased by the Youden technique. If six out of 10 parameters are biased, then the score will be 60%. Some laboratories are both biased and flagged and can have very high scores indicative of poor performance; corrective action is suggested in these cases.

Table 15. MEDIAN2: Comparison of Interlaboratory Median Values Between Studies

		Previous Data (Study No./Sample No.)					
		LR-PRD-10					
Parameter	Study No. / Sample No.	0012 / 8	0010 / 9	0016 / 5	0019 / 2	Avg. Med.	Std. Dev.
Colour	Hazen unit	30.000	35.000	30.000	30.000	31.250	2.500
Specific conductance	µmho/cm	35.300	35.600	36.000	35.800	35.675	0.299
Acidity to pH 8.3	mg/L $CaCO_3$	7.000	6.875	6.740	5.620	6.556	0.640
Acidity (all methods)	mg/L $CaCO_3$	6.530	6.770	-	-	6.650	0.170
pH	pH units	4.420	4.400	4.420	4.410	4.413	0.010
Nitrate + Nitrite	mg N/L	0.010	0.007	0.009	0.006	0.008	0.002
Ammonia	mg N/L	0.010	0.010	0.007	0.006	0.008	0.002
Total Kjeldahl nitrogen	mg N/L	0.151	0.178	0.160	0.140	0.157	0.016
Sodium	mg/L	0.531	0.500	0.535	0.537	0.526	0.017
Magnesium	mg/L	0.410	0.400	0.403	0.400	0.403	0.005
Reactive silica	mg Si/L	2.389	1.326	2.390	2.397	2.376	0.033
Sulphate non IC methods	mg/L	8.800	8.600	8.700	8.350	8.613	0.193
Sulphate (all methods)	mg/L	8.120	8.255	-	-	8.138	0.025
Chloride non IC methods	mg/L	0.400	0.400	0.355	0.415	0.393	0.026
Chloride (all methods)	mg/L	0.200	0.210	-	-	0.205	0.007
Potassium	mg/L	0.141	0.140	0.140	0.150	0.143	0.005
Calcium	mg/L	1.800	1.780	1.780	1.753	1.778	0.019
Sulphate IC method	mg/L	8.080	8.040	8.220	8.096	8.109	0.078
Chloride IC	mg/L	0.200	0.200	0.200	0.210	0.203	0.005
Alkalinity-gran., inflec., extrap.	mg/L $CaCO_3$	0.000	0.000	0.000	0.000	0.000	0.000
Gran. titra alk.	mg/L $CaCO_3$	-2.130	-	-2.000	-2.150	-2.093	0.081
Alkalinity-fixed endpt. pH 4.5	mg/L $CaCO_3$	0.100	0.000	0.000	0.000	0.025	0.050
Gran. acidity	mg/L $CaCO_3$	5.340	6.590	5.545	4.409	5.471	0.895
Dissolved org. carbon	mg C/L	-	-	5.699	5.550	5.625	0.105
Dissolved inorg. carbon	mg C/L	-	-	0.275	0.325	0.300	0.035

DATES: 0012 (86/04/07), 0010 (85/10/29), 0016 (87/08/24), 0019 (88/08/24)

Experience in analyzing many studies has created guidelines on performance (as viewed through the Youden bias and flagging process). Scores of over 60% are poor (maximum score is 200%: all data flagged and all parameters biased). Scores of less than 10% are satisfactory, well done, scores between 10 and 25% are satisfactory, and scores between 25 and 60% are moderate.

The visual impact to a laboratory that is graded with a high score in LABCOMP is informative. A very low score (satisfactory) creates satisfaction. A very high score (over 60%) can be very surprising to a laboratory; it is certainly stimulating and cause for immediate internal review. To this end, the LABCOMP output program has merit, and evidence would suggest that the impact for some laboratories has been constructive.

Table 16. LABCOMP: Comparison of Laboratory Performance (Within a Study)

Lab Code	BIAS			FLAGS			Sum of % bias and % flags score
	No. of parameters analyzed	No. of parameters biased	Percentage of parameters biased (%)	No. of results ranked	No. of flags assigned	Percentage of results flagged (%)	
L002	14	1	7.14	125	1	0.80	7.94
L002C	9	0	0.00	90	0	0.00	0.00
L003	12	0	0.00	100	6	6.00	6.00
L004	15	8	53.33	136	11	8.09	61.42
L005	7	0	0.00	66	0	0.00	0.00
L006	16	6	37.50	160	2	1.25	38.75
L007	10	5	50.00	100	15	15.00	65.00
L008	6	3	50.00	60	19	31.67	81.67
L010	10	4	40.00	100	21	21.00	61.00
L011	11	2	18.18	105	1	0.95	19.13
L013	6	0	0.00	54	0	0.00	0.00
L014	16	6	37.50	156	22	14.10	51.60
L019	10	4	40.00	96	9	9.38	49.38
L021	11	2	18.18	104	13	12.50	30.68
L023	13	3	23.08	117	13	11.11	34.19
L024	15	3	20.00	142	6	4.23	24.23
L025	20	5	25.00	188	1	0.53	25.53
L029	16	1	6.25	157	8	5.10	11.35
L030	9	0	0.00	86	0	0.00	0.00
L032	8	3	37.50	79	11	13.92	51.42
L033	8	3	37.50	71	3	4.23	41.73
L034	10	1	10.00	96	5	5.21	15.21
L035	6	1	16.67	60	9	15.00	31.67
L041	5	1	20.00	50	8	16.00	36.00
L043	4	1	25.00	40	0	0.00	25.00
L045	13	4	30.77	126	4	3.17	33.94
L047	11	6	54.55	92	53	57.61	112.15
L048	12	0	0.00	114	5	4.39	4.39
L049	16	10	62.50	155	31	20.00	82.50
L049B	2	0	0.00	20	0	0.00	0.00
L053	11	3	27.27	99	0	0.00	27.27
L058	13	4	30.77	130	3	2.31	33.08
L061	10	0	0.00	99	0	0.00	0.00
L063	17	2	11.76	161	1	0.62	12.39
L064	9	5	55.56	87	19	21.84	77.39
L066	10	3	30.00	100	10	10.00	40.00
L067	12	2	16.67	119	23	19.33	35.99
L069	16	2	12.50	156	4	2.56	15.06
L073	10	1	10.00	97	1	1.03	11.03
L078	2	0	0.00	20	0	0.00	0.00
L081	11	1	9.09	101	5	4.95	14.04
L082	12	5	41.67	120	8	6.67	48.33
L085	14	2	14.29	124	7	5.65	19.93
L086	13	5	38.46	128	16	12.50	50.96
L087	10	0	0.00	100	5	5.00	5.00
L088	6	1	16.67	60	0	0.00	16.67
L089	12	1	8.33	120	5	4.17	12.50
L089B	1	0	0.00	10	0	0.00	0.00
L089C	10	0	0.00	97	0	0.00	0.00
L090	11	1	9.09	106	2	1.89	10.98
L090B	1	0	0.00	10	0	0.00	0.00
L091	14	2	14.29	140	20	14.29	28.57
L092	3	0	0.00	27	4	14.81	14.81
L093	13	1	7.69	129	15	11.63	19.32
L094	14	1	7.14	135	9	6.67	13.81
L095	10	1	10.00	100	6	6.00	16.00
L096	7	4	57.14	70	27	38.57	95.71
L097	11	0	0.00	98	3	3.06	3.06

The following codes were used in the analysis: 00392 01090 01092 06002 06194 07092 07192 11091 12091 14092 16001 17001 19091 20091 16000 17000 06193 06282 06592 07392 01089
The following codes were excluded: 00292

8.9 FLGTBL (S2K): Performance of a Group of Laboratories Based on Frequency of Bias and Flags

Some studies, such as for the federal–provincial LRTAP intercomparison program, are frequent (three per year), involve laboratories of equivalent capability, use the same types of water (soft), and have criteria for flagging that have remained constant over several years. These studies also involve about 50 laboratories that, in general, analyze the same constituents. With this resource (almost 20 major ion studies), it is possible to compare the frequency with which laboratories have their data assessed as biased or flagged. In fact, it is possible to provide a track record on the performance of each laboratory over time.

Within any study, a laboratory that gets most of its parameters declared biased and has most of its results flagged is considered as giving a very poor performance. On the other hand, a laboratory with no bias and no flags may be considered satisfactory and an excellent performer. Between these two extremes lie average or moderate performances. When studies are frequent, it is possible to examine trends in the frequency of biased and flagged results (e.g., improvements over time may be observed). The program called FLGTBL helps in this effort. An example output is given in Table 17. This output is the integrated results of LABCOMP; when the output is transferred from the mainframe to a PC, performance can be graphically displayed. Because the output is created from S2K, it is possible to create an output for any group of laboratories, studies, or parameters.

A typical graphic display of performance for one laboratory abstracted from FLGTBL is given in Figure 8. The performance index in this figure (and Table 17) is the same as used in LABCOMP. It is arbitrary and may be modified when all evidence has been reviewed.

8.10 APPRAIS (S2K): Automated Appraisals

When the original YOUDN21 flat file programs were applied to large 50-laboratory, 10-sample, 20-parameter studies, a great deal of manual effort was required to prepare narrative comments on each laboratory for every parameter and each sample result. Not only was it tedious, but it was subject to human and transcription error.

With the development of the S2K data base, the preparation of an appraisal became extremely rapid, because sufficient space was built into the data base structure to store the calculated outputs. Table 18 gives the criteria developed so that when the program APPRAIS is initiated it retrieves from the data base the necessary information to formulate a written narrative. A typical narrative is defined as a "laboratory-specific appraisal" and is given in Table 19. This appraisal is attached to a covering letter accompanied by all support data (MEDIAN1, MEDIAN2, LABCOMP, FLGTBL, and YOUDN21). This critical support information is essential and is provided to each participant when a study is formally completed.

Table 17. FLGTBL: Comparison of Laboratory Performance over Several Studies (LRTP)

A.

Lab code	% Bias and % Flags on Studies						Median score	Comments
	0015	0016	0017	0018	0019	0020		
L002	27.0	0.8	7.4	12.6	10.7	7.9	9.3	Satisfactory, well done
L002C	24.6	11.1	13.5	0.0	2.2	0.0	6.7	Satisfactory, well done
L003	10.5	0.0	2.5	18.8	28.8	6.0	8.3	Satisfactory, well done
L004	23.3	67.2	14.5	4.4	37.6	61.4	30.4	Moderate
L005	0.0	17.5	14.3	41.5	33.3	0.0	15.9	Satisfactory
L006	0.6	13.1	23.2	18.4	30.7	38.8	20.8	Satisfactory
L007	27.5	59.1	13.0	17.1	38.8	65.0	33.2	Moderate
L008	41.7	-	126.7	80.0	96.0	81.7	81.7	Poor
L010	51.5	32.8	43.1	58.1	46.3	61.0	48.9	Moderate
L011	23.9	0.9	46.7	14.5	-	19.1	19.1	Satisfactory
L013	6.5	16.7	31.2	0.0	1.8	0.0	4.1	Satisfactory, well done
L014	12.5	54.0	20.6	24.0	28.8	51.6	26.4	Moderate
L014C	47.0	60.8	8.1	24.0	32.5	-	32.5	Moderate
L017	-	-	-	45.3	-	-	-	-
L019	-	2.0	24.2	28.3	0.0	49.4	24.2	Satisfactory
L020	5.3	-	25.9	-	-	-	15.6	Satisfactory
L020C	5.2	-	1.3	-	-	-	3.2	Satisfactory, well done
L021	1.3	10.1	4.9	32.0	39.2	30.7	20.4	Satisfactory
L022	39.0	-	14.4	-	37.2	-	37.2	Moderate
L023	29.1	6.1	43.5	35.3	37.1	34.2	31.7	Moderate
L024	10.0	15.1	26.3	16.2	37.2	24.2	20.2	Satisfactory
L025	31.8	39.4	26.5	15.1	21.2	25.5	26.0	Moderate
L027	56.0	51.8	1.5	-	-	-	51.8	Moderate
L029	35.5	25.1	12.1	-	13.7	11.3	13.7	Satisfactory
L030	4.1	1.4	1.3	24.0	0.0	0.0	1.3	Satisfactory, well done
L031	7.9	14.4	16.0	37.5	10.1	-	14.4	Satisfactory
L032	63.3	37.0	53.1	61.5	60.0	51.4	56.5	Moderate
L033	2.9	38.8	0.0	55.0	20.1	41.7	29.4	Moderate
L034	33.9	0.0	18.7	2.8	15.1	15.2	15.2	Satisfactory
L035	30.0	0.0	0.0	14.3	30.9	31.7	22.1	Satisfactory
L041	-	-	15.0	10.0	-	36.0	15.0	Satisfactory
L043	0.0	-	25.0	0.0	-	25.0	12.5	Satisfactory
L045	23.6	-	8.5	11.8	25.7	33.9	23.6	Satisfactory
L047	-	-	74.8	75.6	98.5	112.2	87.0	Poor
L048	14.5	-	34.1	7.7	324.2	4.4	14.5	Satisfactory
L049	62.7	-	26.9	79.7	60.1	82.5	62.7	Poor
L052	9.5	38.4	40.5	33.9	29.2	-	33.9	Moderate
L053	18.2	-	9.1	-	-	27.3	18.2	Satisfactory
L054	50.0	-	-	106.0	-	-	78.0	Poor
L056	101.5	-	-	-	-	-	-	-
L057	51.4	54.2	50.0	16.3	-	-	50.7	Moderate
L058	12.1	0.9	43.6	93.3	-	33.1	33.1	Moderate
L059	-	-	-	96.7	48.2	-	72.4	Poor
L060	45.7	-	-	-	-	-	-	-
L061	2.9	-	11.0	-	10.0	0.0	6.4	Satisfactory, well done
L063	13.6	30.8	27.1	25.0	23.2	12.4	24.1	Satisfactory
L064	73.3	16.0	35.0	34.0	20.0	77.4	34.5	Moderate
L066	-	13.7	13.4	42.5	23.0	40.0	23.0	Satisfactory
L067	21.7	53.3	67.7	51.2	-	36.0	51.2	Moderate
L069	-	-	32.9	18.3	40.8	15.1	25.6	Moderate
L073	21.3	6.8	71.4	0.0	10.0	11.0	10.5	Satisfactory
L074	51.3	44.7	61.4	63.1	27.8	-	51.3	Moderate
L078	50.0	0.0	0.0	0.0	5.0	0.0	0.0	Satisfactory, well done
L081	77.4	3.6	15.2	22.6	30.4	14.0	18.9	Satisfactory
L082	-	44.4	-	48.4	-	48.3	48.3	Moderate
L083	37.6	34.9	-	-	-	-	36.2	Moderate
L084	-	31.0	-	-	-	-	-	-
L085	-	-	35.7	34.6	-	19.9	34.6	Moderate
L086	-	65.7	54.2	53.7	60.2	51.0	54.2	Moderate
L087	-	-	-	35.3	19.8	5.0	19.8	Satisfactory
L088	-	-	-	10.0	20.0	16.7	16.7	Satisfactory
L089	-	-	-	78.8	46.7	12.5	46.7	Moderate
L089C	-	-	-	0.0	23.3	0.0	0.0	Satisfactory, well done
L090	-	-	-	18.0	10.4	11.0	11.0	Satisfactory
L091	-	-	-	31.4	25.7	28.6	28.6	Moderate
L092	-	-	-	62.4	31.3	14.8	31.3	Moderate
L093	-	-	-	65.0	42.0	19.3	42.0	Moderate
L094	-	-	-	57.3	27.4	13.8	27.4	Moderate
L095	-	-	-	75.0	-	16.0	45.5	Moderate
L096	-	-	-	-	-	95.7	-	-
L097	-	-	-	-	-	3.1	-	-

Study dates: 0015 (87/04/06), 0016 (87/08/24), 0017 (87/12/01), 0018 (88/05/02), 0019 (88/08/24), 0020 (89/01/10)

Table 17. Continued

B.

Lab code	Median score (%)	Number of studies	Lab code	Median score (%)	Number of studies
L001	-	0	L005	15.9	6
L009	-	0	L088	16.7	3
L012	-	0	L053	18.2	3
L013A	-	0	L081	18.9	6
L016	-	0	L011	19.1	5
L017	-	1	L087	19.8	3
L026	-	0	L024	20.2	6
L028	-	0	L021	20.4	6
L031C	-	1	L006	20.8	6
L037	-	0	L035	22.1	6
L038	-	0	L066	23.0	5
L039	-	0	L045	23.6	5
L040	-	0	L063	24.1	6
L042	-	0	L019	24.2	5
L044	-	0	L069	25.6	4
L046	-	0	L025	26.0	6
L050	-	0	L014	26.4	6
L051	-	0	L094	27.4	3
L056	-	1	L091	28.6	3
L060	-	1	L033	29.4	6
L068	-	0	L004	30.4	6
L070	-	0	L092	31.3	3
L071	-	0	L023	31.7	6
L077	-	0	L014C	32.5	5
L080	-	0	L058	33.1	5
L084	-	1	L007	33.2	6
L096	-	1	L052	33.9	5
L097	-	1	L064	34.5	6
L098	-	0	L085	34.6	3
L089C	0.0	3	L083	36.2	2
L078	0.0	6	L022	37.2	3
L030	1.3	6	L093	42.0	3
L020C	3.2	2	L095	45.5	2
L013	4.1	6	L089	46.7	3
L061	6.4	4	L082	48.3	3
L002C	6.7	6	L010	48.9	6
L003	8.3	6	L057	50.7	4
L002	9.3	6	L067	51.2	5
L073	10.5	6	L074	51.3	5
L090	11.0	3	L027	51.8	3
L043	12.5	4	L086	54.2	5
L029	13.7	5	L032	56.5	6
L031	14.4	5	L049	62.7	5
L048	14.5	5	L059	72.4	2
L041	15.0	3	L054	78.0	2
L034	15.2	6	L008	81.7	5
L020	15.6	2	L047	87.0	4

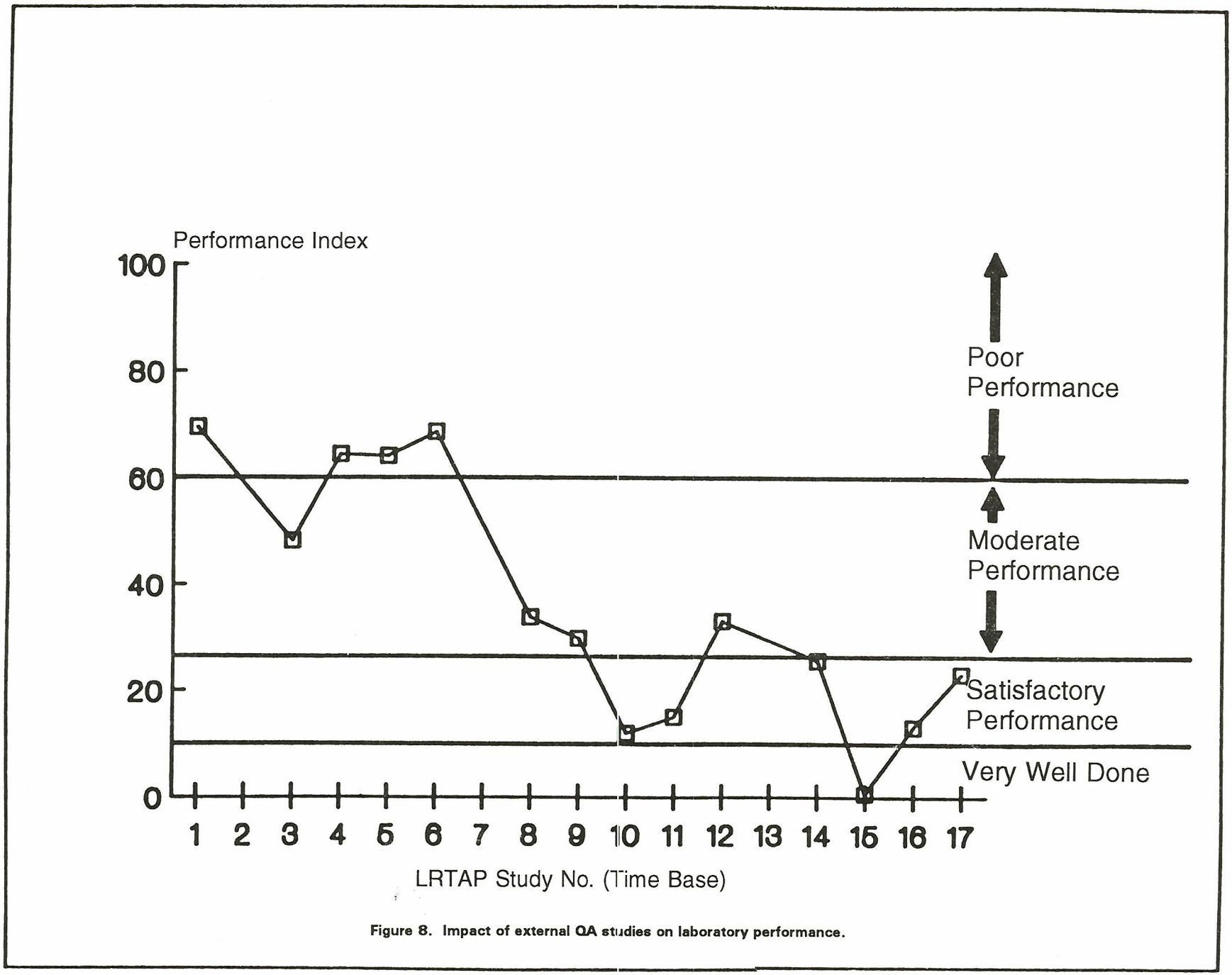


Figure 8. Impact of external QA studies on laboratory performance.

Table 18. Criteria Used to Prepare Statements for the Automated Appraisals

Status of data	Statement produced in appraisal
1(A). No flags, no bias in the data.	Satisfactory.
(B). No bias, only 1 minor flag (H or L).	Satisfactory, except for low/high flag on sample _.
2. No data or results reported by laboratory.	No results reported.
3. Data reported on less than half of the samples, no results flagged.	Insufficient data to assess bias.
4. Same as item 3, but some results are flagged.	Flagged _ on samples _, _ and flagged _ on sample _. Insufficient data to access bias.
5. No results are flagged, but data set identified as biased high or low.	Although no results are flagged, ranking indicates results are biased high/low.
6. Some results are flagged, the data set is discerned as biased.	Flagged _ on samples _, _ and flagged _ on samples _, _. Ranking indicates results are biased.
7. Some results are flagged.	Flagged _ on samples _, _ and flagged _ on sample _.
8. No bias statement, but two or three results are flagged, one is very high, the other is very low.	Flagged very high on sample _ and very low on sample _. These results are slightly erratic.
9. No bias statement, but two or more results are flagged very high and two or more results are flagged very low.	Flagged very high on samples _, _ and flagged very low on samples _, _, _. These results are erratic.
10. Results are ranked, the data set is not biased, but one result is flagged very high or very low.	Flagged _ on sample _. This extreme result suggests the measurement process is out of control.

Table 19. Laboratory Appraisal (an Example)

Parameter	Appraisal
SPECIFIC CONDUCTANCE	Satisfactory
GRAN ACIDITY	No results reported
ACIDITY TO PH 8.3	Satisfactory
PH	Satisfactory except for low on sample 4
DISSOLVED ORG CARBON	Satisfactory
ALKALINITY-FIXED ENDPFT. PH 4.5	No results reported
ALKALINITY-GRAN, INFLEC, EXTRAP	Insufficient data to assess bias
GRAN TITRA ALK	No results reported
DISSOLVED INORG CARBON	Flagged low on sample 10—ranking indicates results are biased low
NITRATE + NITRITE	Satisfactory
AMMONIA	Satisfactory
TOTAL KJELDAHL NITROGEN	No results reported
SODIUM	Flagged extremely high on sample 9 — this extremely high result suggests the measurement process is out of control
MAGNESIUM	Flagged high on samples 4, 10
REACTIVE SILICA	Flagged high on samples 6, 7 Flagged very high on sample 5
SULPHATE IC METHOD	Although no results are flagged, ranking indicates a slight bias high
SULPHATE NON IC METHODS	Flagged extremely low on sample 1—, this extremely low result suggests the measurement process is out of control
CHLORIDE IC	Satisfactory
CHLORIDE NON IC METHODS	Satisfactory
POTASSIUM	Although no results are flagged, ranking indicates a slight bias high
CALCIUM	Satisfactory

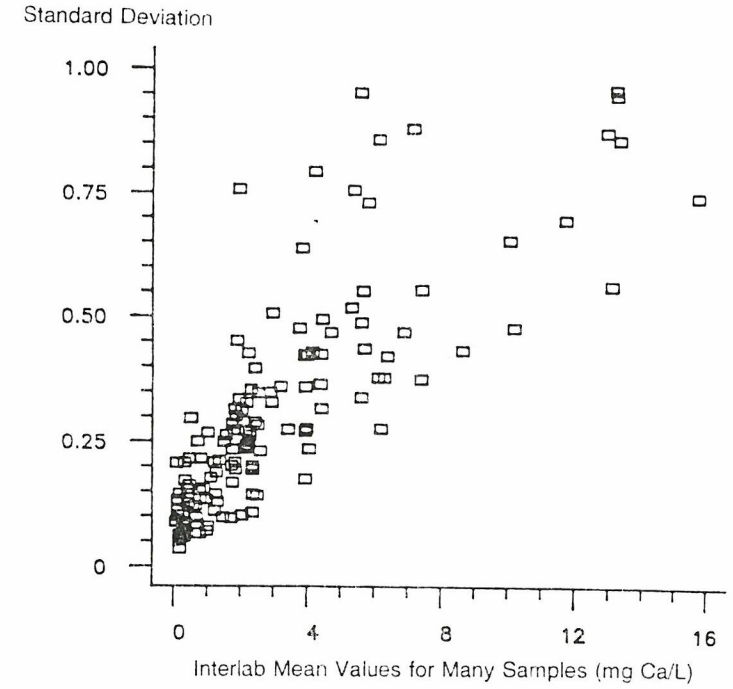
8.11 AQC Programs in Development

The above AQC data base management system has proven very effective and efficient in addressing large external QA studies. The external QA studies provided to the LRTAP, Eulerian, IJC, LRTAP terrestrial, and the National QA programs have created a data base of some significance. This resource is now being analyzed to improve the overall QA assessment program and to allow more ready access of information to analysts, program managers, and the users of data.

8.11.1 Precision Functions

The main strength of the YOUDN21 (or YOUDS2K) program has been the ability to discern bias (systematic error) and provide analysts information on precision (flags). The formula for flagging has been a series of educated guesses so that the criteria (BAE, LLBAE, and CEI) would provide reasonable distribution of flags.

The heavily populated data base system can now be used to analyze interlaboratory means and standard deviations to create more meaningful distributions of precision (the precision functions).



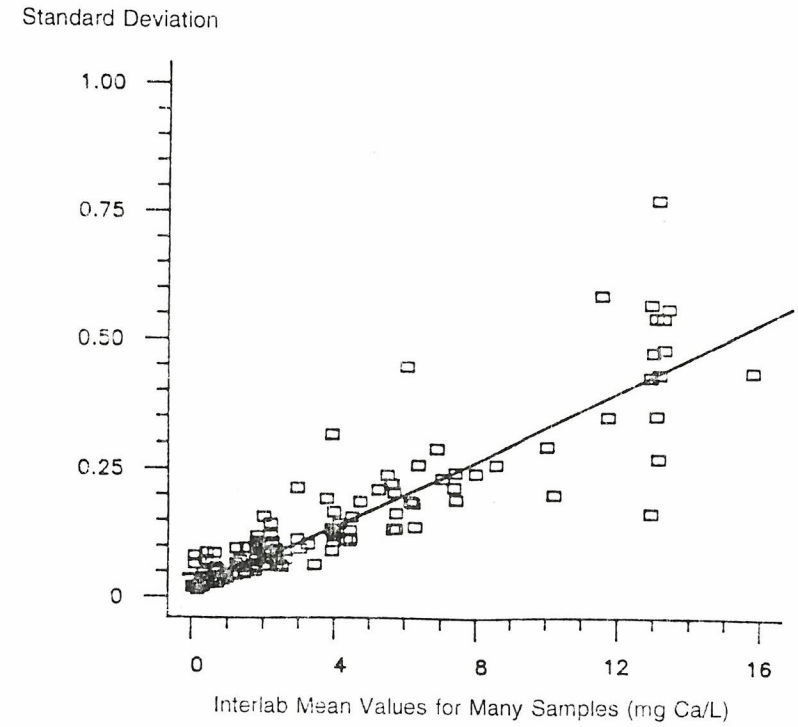


Figure 9. Precision function for calcium (LRTAP QA program) (A) raw data; (B) after removing some outliers.

This possibility can be realized because of the existence of a larger body of information on many different samples (20 to 200) contained in such programs as the Eulerian and LRTAP. A typical output that is anticipated is given in Figure 9.

The selective nature of the S2K data base can allow the graphics to be created to isolate such functions on groups of laboratories (e.g., surface water, rain water, government, contract) for selected samples or for all samples. The more serious benefit will be in the development of functions clearly stating the level of performance that satisfactory laboratories have demonstrated. All other laboratories (e.g., contract laboratories) will need to either achieve this performance or excel beyond this minimum standard. Future applications of the data system have considerable potential for serving environmental programs and the inherent quality management issues.

8.11.2 Charge Balance and Conductance Ratios

One specific intralaboratory QC procedure that all large water quality laboratories often utilize to check results is the calculation of charge balance (anions versus cations) and, for soft waters, the ratio of measured to calculated conductance.

As with precision functions, the AQC data base is sufficiently large for water studies to analyze the overall results to confirm that interlaboratory medians indeed have integrity and to create percent error or precision functions on charge balance and conductance calculations. The ability to create a distribution of percent error or uncertainties as a function of ionic strength will have merit for users of data and will give information on the criteria required for the selection of contract laboratories. This work is in progress.

8.12 References

American Society for Testing and Materials. 1983. Intralaboratory Quality Control Procedures and a Discussion Reporting Low-Level Data, A Standard Practice. (ASTM-D19), D4210-83, Philadelphia, Pa.

Aspila, K.I., R.E. White, and J.L. Clark. 1985. Quality assurance aspects of the International Joint Commission Great Lakes Monitoring Program. Am. Soc. Test. Mater. Spec. Tech. Publ. 867: 407–421.

Clark, J.L. 1981. Evaluation of performance of laboratories determining water quality constituents through natural water samples whose true values are unknown. *In* Summary of Conference Presentations, Envirometrics 81, April 8–10, Alexandria, Va., pp. 54–55.

Youden, W.J. 1969. Ranking laboratories by round-robin tests. Natl. Bur. Stand. (U.S.) Spec. Publ. 300(1): 165-9–169-13.

Youden, W.J., and E.H. Steiner. 1975. Statistical Manual of the Association of Official Analytical Chemists. AOAC, Washington, D.C.

9.0 THE FEDERAL–PROVINCIAL AND PRAIRIE PROVINCES WATER BOARD QA PROGRAMS

Chapter 8 described how large interlaboratory studies are presented and processed by various computer assessment techniques. These have served well for LRTAP aquatics, LRTAPP terrestrial, recent National, and the Eulerian QA studies. The QA programs that use alternative methods for assessing laboratory results are the Federal–Provincial (FP) and Prairie Provinces Water Board (PPWB) QA programs. The following is an overview of these two programs.

9.1 General

Under terms of the federal–provincial agreements and the agreements under the PPWB (Prairie Provinces Water Board 1981), QA programs have been implemented to assess and improve the comparability of water quality data (Dux 1986). These programs provide ongoing bimonthly studies for some 40 inorganic constituents in surface waters. Participants include eight federal and eight provincial laboratories. A laboratory from the Department of Indian and Northern Affairs also participates in these programs. The essential activities to implement the FP and PPWB QA programs are highlighted in Figure 10.

The objectives of the FP and PPWB QA programs are briefly expressed as follows:

- to detect laboratory measurement anomalies and report them quickly to laboratory managers to allow for swift remedial action;
- to ensure and define comparability and reliability of data that are eventually stored in the national data base system (NAQUADAT) in order to assist users of data;
- to provide evidence on the effectiveness of intralaboratory and interlaboratory QC procedures; and
- to provide regular reports that summarize results for laboratory and project managers.

9.2 Study Design

A single bimonthly study (Cheam and Chau 1982) consists of four or five standard reference samples of known values. Half of these samples are for trace metal analysis at two concentration levels. For the other half of the samples, the laboratories report on 25 major ions and on nutrient and physical parameters. Altogether, 100 analysis methodologies and individual results are tabulated in the data summaries. In 1988, the FP QA program was expanded to address toxic organics using injection-ready ampule standards.

A standard reporting form showing various inorganic parameters is given in Table 20.

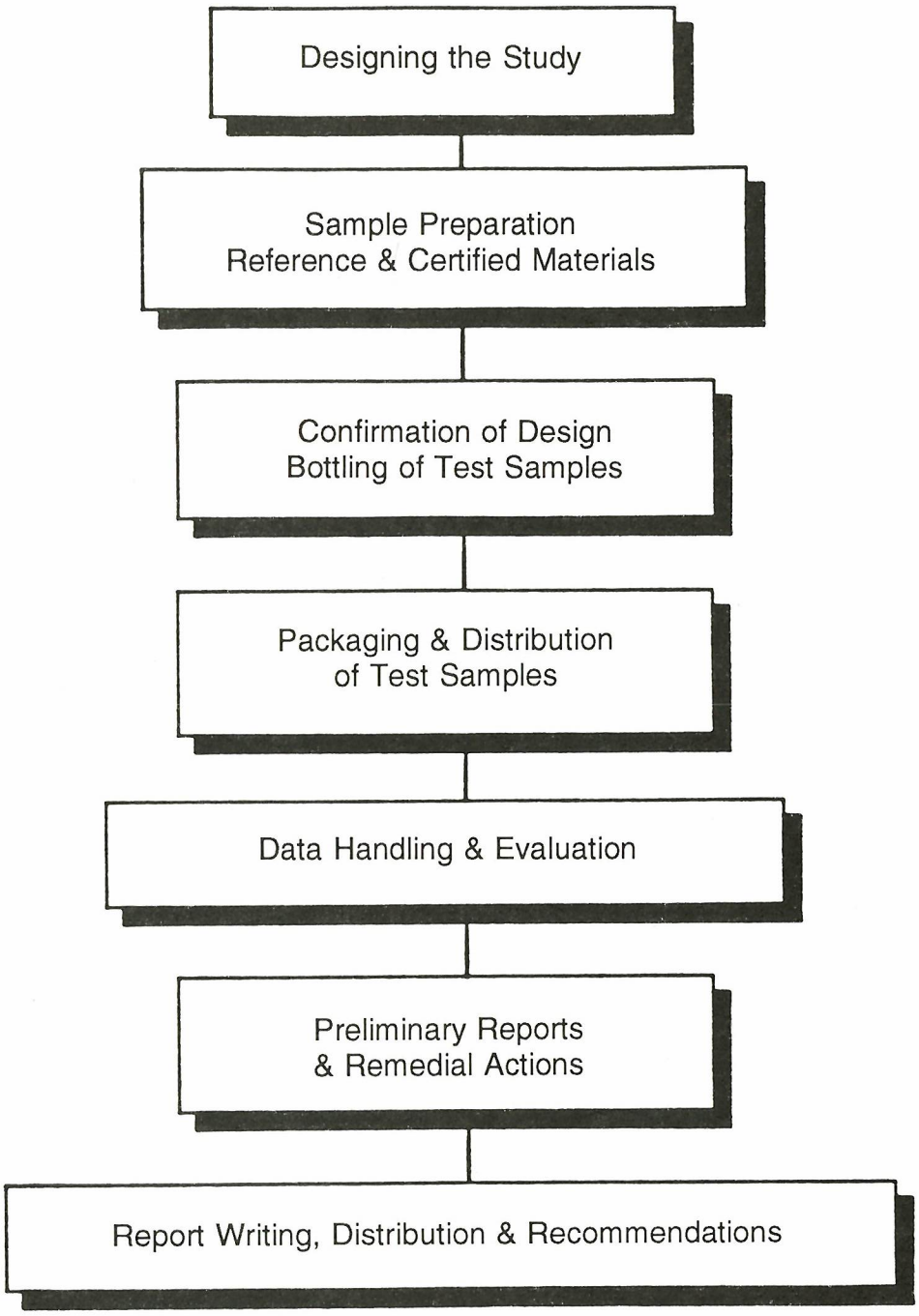


Figure 10. Flow chart of activiities to implement the FP and PPWB QA programs.

Table 20. Data Report Sheet, Federal-Provincial and Prairie Provinces Water Board Quality Assurance Programs

Lab No. _____ Study No. _____ Date _____

Major ions, nutrients	NAQUADAT code	Sample results, ppm	Metals	NAQUADAT code	Sample results, ppm
Colour (units)	020		Aluminum	13	
Specific conductance ($\mu\Omega$)	020		Vanadium	23	
Turbidity (JTU/NTU)	020		Chromium	24	
Boron	051		Manganese	25	
DOC	06		Iron	26	
DIC	06		Cobalt	27	
TKN	07		Nickel	28	
$NO_3 + NO_2$	071		Copper	29	
Ammonia	075		Zinc	30	
Total nitrogen	076		Strontium	38	
Fluoride	091		Molybdenum	42	
Total alkalinity	101		Cadmium	48	
Acidity	102		Barium	56	
pH (units)	103		Lead	82	
Total hardness	106				
Sodium	11				
Magnesium	12				
Silica (as SiO_2)	14				
Total P (as P)	15		MI calculated parameters		
Sulphate (as SO_4)	16				
Chloride	17		Sum of cations meq/L	00120	
Potassium	19		Sum of anions meq/L	00125	
Calcium	20		% difference	00110	

PLEASE REPORT ALL DATA IN PPM, UNLESS OTHER UNITS ARE INDICATED

9.3 Data Evaluation

Analyses for four or five test samples are made by each laboratory during a two-month period and results reported to the NWRI QA chemist in charge. Data are entered into the Cyber 180/830A mainframe computer, and programs are executed to format data and prepare printouts. A sample data report showing various laboratory methodologies and parameter statistics is given in Table 21. An overview is given in Table 22.

Because it is difficult to have all participants analyze all samples simultaneously, several preliminary reports are often provided during each bimonthly study. This service allows laboratories the opportunity of rapid review and corrective action if their data are indicated as inadequate. A formal report is distributed at the completion of each study. The following describes how flags are assigned to deviating results.

Table 21. Data Summary, Federal-Provincial and Prairie Provinces Water Board Quality Assurance Program

STUDY NO.	FP 35	PP 75			DATE: 01/11/88		PAGE 1
Sample 1					Trace Metals D/A. in 3.0% HNO_3)		
	13009	13111	13302	13306	13321	13322	13999
LAB	Al Tot	Al Dis	Al Ext	Al UF	Al Ext	Al Ext	Aluminum
	5% ICP	ICP DA	AAS DA	AAS OX	ICP DA	DCP DA	COMMON
1	-	-	-	1.014	-	-	1.014
2	-	-	1.1	-	-	-	1.1
3	-	-	0.980	-	1.11	-	0.980
6	1.0	-	-	-	-	-	1.0
8	-	-	1.2	-	-	-	1.2*
9	-	1.05	-	-	-	-	1.05
10	-	0.97	-	-	-	-	0.97
15	-	-	-	-	1.02	-	1.02
16	-	-	-	-	-	1.03	1.03
MEAN	1.0000	1.0100	1.0933	1.0140	1.1100	1.0300	1.0404
STD DEV	-	0.0566	0.1102	-	-	-	0.0712
REL STD	-	5.6	10.1	-	-	-	6.8
DES VAL	-	-	-	-	-	-	1.048
	25004	25011	25012	25302	25311	25321	25999
LAB	Mn Tot	Mn Tot	Mn Tot	Mn Ext	Mn Ext	Mn Ext	Manganese
	AAS GF	5% ICP	TX DCP	AAS DA	ICP DA	ICP DA	COMMON
1	-	-	-	-	0.101	-	0.101
3	-	-	-	-	-	0.099	0.099
6	-	0.095	-	-	-	-	0.095
8	0.097	-	-	-	-	-	0.097
9	-	-	-	-	-	-	0.099
10	-	-	-	-	-	-	0.094
11	-	-	-	0.093	-	-	0.093
13	-	-	-	-	0.09	-	0.09
14	-	-	-	-	-	-	-
15	-	-	-	-	0.098	-	0.098
16	-	-	0.112	-	-	-	0.112 R
MEAN	0.0970	0.0950	0.1120	0.0930	0.0940	0.0990	0.0978
STD DEV	-	-	-	-	0.0057	-	0.0060
REL STD	-	-	-	-	6.0	-	6.1
DES VAL	-	-	-	-	-	-	0.0972

NOTE: All concentration units are expressed in mg/L of each element, the exception being: Colour in relative units, Conductivity in μS/cm, Turbidity in JTU or NTU, Nitrogen analysis in N, Alkalinity & Hardness in $CaCO_3$, Silica in SiO_2, and Sulphate in SO_4.

Results flagged with the letter R have been rejected (refer to Section 9.3.2). Results flagged with an asterisk deviate from the target by more than one standard deviation (see Section 9.3.1).

Table 22. Summary of Flagged Results for the Federal-Provincial QA Program (for Studies FP 3I-32)

LAB 2	FLAGS :	NITRATE	-16%	POTASSIUM	-91% R	ZINC	25%
		NITRATE	-62% L				
	HDL :	T N DIS					
LAB 3	FLAGS :	NITRATE	-81% L	T N DIS	-89% L		
LAB 4	FLAGS :	NONE					
LAB 5	FLAGS :	D O C	73%	D I C	16%	AMMONIA	-75% L
		PH	-14%	D O C	337% R		
	HDL :	SILICA					
LAB 7	FLAGS :	NITRATE	-26%	TOT P	106%		
LAB 9	FLAGS :	COPPER	-23%	MOLYBDENUM	-23%	POTASSIUM	-21%
	HDL :	ALKALINITY					
LAB 10	FLAGS :	COPPER	39%	SILICA	-12%	ALUMINUM	50%
		COBALT	-36%	LEAD	40%	T N DIS	55%
	HDL :	AMMONIA		TOT P		AMMONIA	
LAB 11	FLAGS :	NICKEL	16%	D O C	368% R	D I C	-22%
		TOT P	106%	IRON	38%	AMMONIA	126%
		HARDNESS	-19% R	POTASSIUM	13%	CALCIUM	-25% R
	HDL :	D O C		NITRATE			
LAB 12	FLAGS :	TKN	22%	ALUMINUM	25%		
	HDL :	D O C		D I C		SILICA	
LAB 13	FLAGS :	IRON	-20%	CONDUCT	-14%	ALKALINITY	19% R
		CONDUCT	-12%	SULPHATE	13%	POTASSIUM	31%
	HDL :	CHROMIUM		AMMONIA			
LAB 14	FLAGS :	MANGANESE	54% R	COPPER	108% R	CADMIUM	144% R
		LEAD	48% R	SODIUM	-35% R	MAGNESIUM	-15% R
		ALUMINUM	23%	IRON	-31%	ZINC	-38%
		SODIUM	-38% R				
LAB 15	FLAGS :	MANGANESE	-15%	ZINC	-16%	NITRATE	41%
		FLUORIDE	25% R	ALUMINUM	358% R	VANADIUM	82% R
		MANGANESE	-50%	IRON	-83% R	COBALT	55%
		ZINC	-63% R	MOLYBDENUM	-23% L	CADMIUM	-55% R
		LEAD	-95% L	SULPHATE	-11%		
	HDL :	NICKEL		MOLYBDENUM			
LAB 16	FLAGS :	MANGANESE	30%	IRON	19%	STRONTIUM	35% R
		LEAD	17%	TURBIDITY	203% R	D I C	25%
		TKN	143% R	SODIUM	-23%	TOT P	3075% R
		SULPHATE	55% R	CHLORIDE	18% R	CALCIUM	-12% R
		ALUMINUM	85% R	VANADIUM	118% R	MANGANESE	67%
		IRON	24%	COPPER	38%	STRONTIUM	74% R
		MOLYBDENUM	-62% L	LEAD	50%	ALKALINITY	552% R
		SODIUM	-21% R	TOT P	2281% R	SULPHATE	67% R
		CALCIUM	-11%				
	HDL :	CADMIUM		AMMONIA		FLUORIDE	
		MOLYBDENUM		NITRATE		AMMONIA	

NOTE: A very high frequency of flagged results (or a high %) is indicative of poor performance. On the other hand, labs with few if any flags, are judged to have very good performance.

Also, an "R" flag indicates a noncomparable result, that is, one produced with nonrandom factors. An "L" flag indicates a "less than" result lower than the comparator.

Also, "HDL" means that the method used has a high detection limit, and, consequently, data in this range may not be appropriate relative to the other laboratories that use more sensitive methods.

In the FP and PPWB QA programs, individual results are flagged if evidence suggests that they deviate significantly from design or target values. The concentration range in the FP and PPWB programs often covers two or three orders of magnitude (e.g., 0.010 to 10 ppm). For this reason, the evaluation of data (Caulcutt and Boddy 1983; Hunt and Wilson 1988) has required two general techniques to determine whether or not a result will be flagged. One approach is the statistical outlier test of Grubbs (Grubbs 1969; Mazu 1981), and the other approach is the 10% or one standard deviation rule. These two approaches are described below.

9.3.1 The 10% or One Standard Deviation Rule

The general approach in discerning a flag by the 10% process is given in Figure 11. The concept of having a 10% rule was introduced and accepted by the FP and PPWB QA programs, because adequate precision (10%) was very achievable for most constituents (metals, nutrients, and physical parameters) in hard surface waters at high concentrations. Unfortunately, at low concentrations, 10% of the design values (e.g., 10% of 0.010 μg Al/L) would be a small percentage of the interlaboratory standard deviation. For this reason, the criterion to flag at low concentrations was assigned in reference to the interlaboratory standard deviation. A result at low concentrations is assigned a flag "*" if the result reported deviates from the target (or design value) by more than one standard deviation. The interlaboratory standard deviation used is that calculated for each sample within each study. Results for aluminum in Table 21 illustrate the above flagging process.

9.3.2 Grubbs' Outliers (FP and PPWB QA Programs)

The second method for flagging data in the FP and PPWB QA programs is the method of Grubbs (Grubbs 1969; Mazu 1981; Taylor 1987). This statistical test discerns an outlier when a result deviates from the population mean by more than two times the standard deviation. Mathematically, this requires calculating the Grubbs' statistic for the suspect values (highest or lowest); if this statistic exceeds a critical value, the result reported is declared an outlier. Details of the procedure are found in the literature (see also Part V).

An example of a Grubbs' outlier is given for manganese in Table 22. The outlier is noted by the letter "R." Refer to Figure 12 for a graphic example of an outlier.

9.4 Laboratory Performance

For each bimonthly study, laboratory heads and their managers are shown their laboratory's comparability of data and the laboratory's performance in terms of the number of flagged results. A sample listing of flagged data in a typical study is given in Table 22.

To enhance the value of this table, computer programs are currently being written to reveal the percentage of flags, to generate appraisals via a computer program, and to have graphics reveal how performance varies over time.

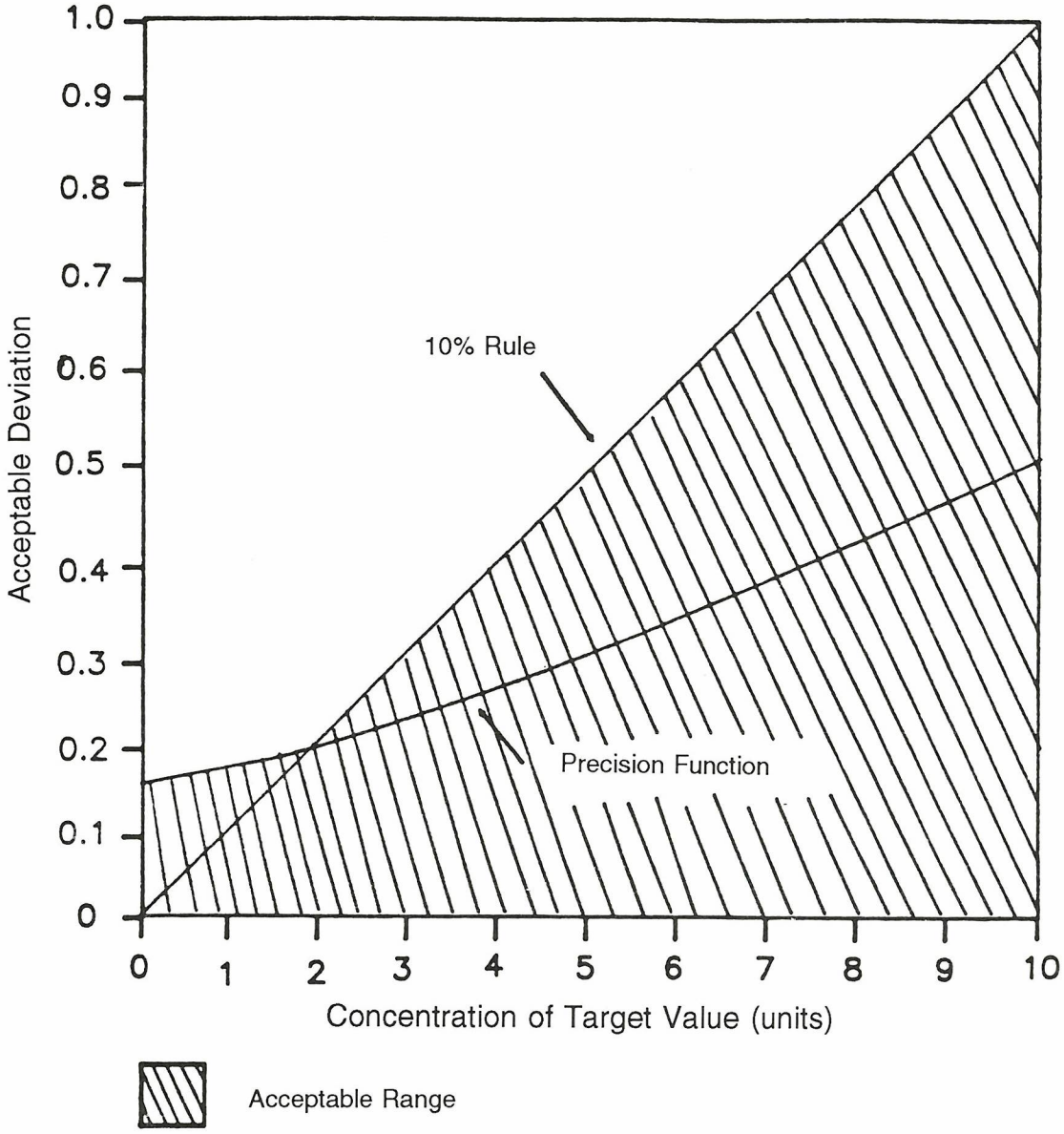


Figure 11. Evaluation of data based on the 10% and one standard deviation rule for FP and PPWB QA programs.

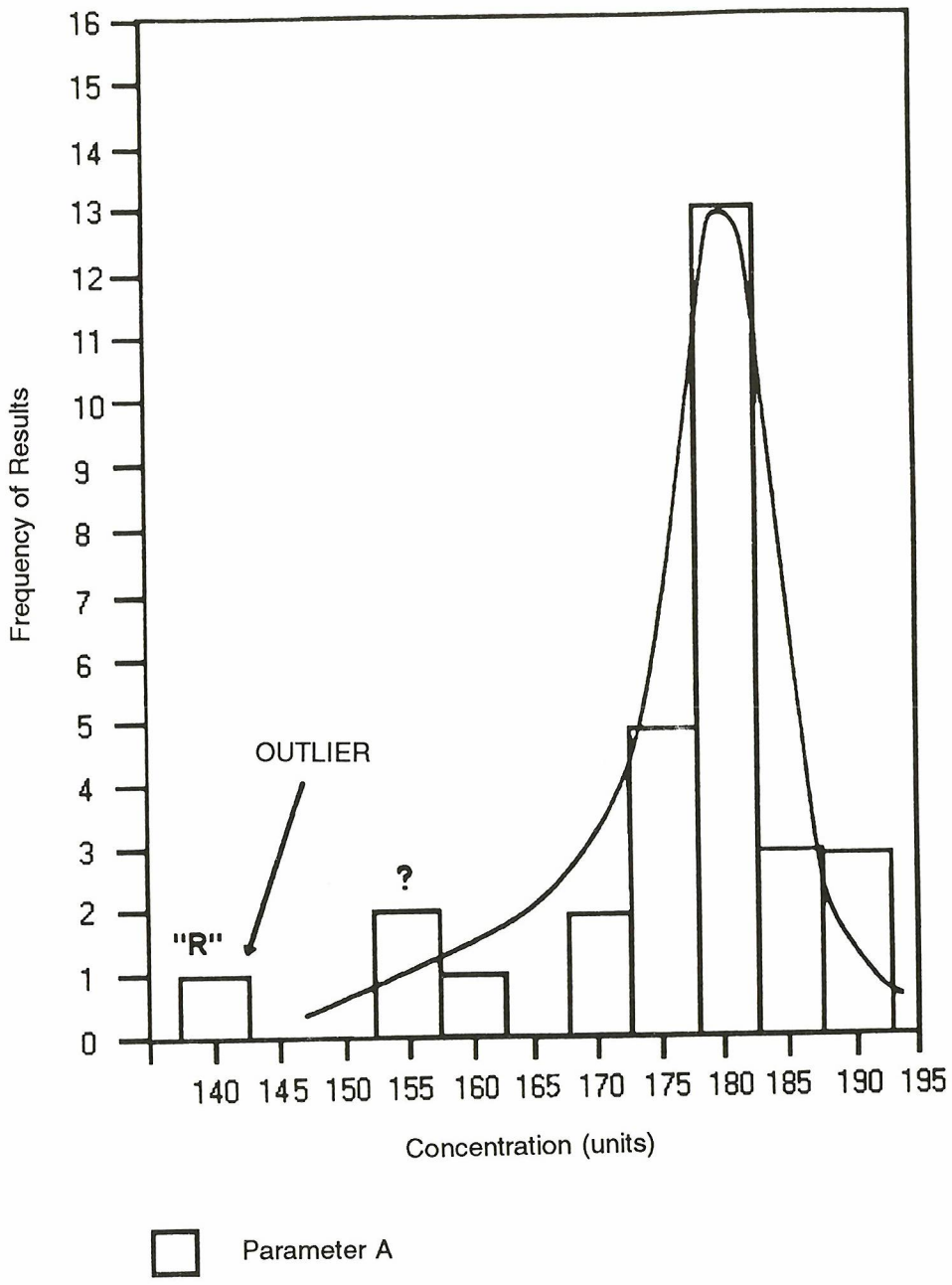


Figure 12. An example of a statistical outlier (Grubbs).

9.5 Impact of the FP and PPWB QA Studies

One of the key features of the original FP QA studies (now referred to as the IRQC: the interregional quality control) was the swift comparison of information on data and methods for a long list of parameters. The data and information program implemented in the early 1970s is currently very active and continues to provide rapid service. A typical report form is given in Table 21. When each laboratory reports data, it includes a NAQUADAT code for each methodology. A new methodology receives a particular code assigned to the method after application to and approval from NAQUADAT officials (Environment Canada, Ottawa). The assigned codes are essential in tabulating data by method. Such tabulations of data by method are useful to laboratory heads when they are evaluating the performance of their own laboratories. A typical summary given to each laboratory is shown in Table 22. This summary is circulated to all participants, laboratory managers, and program managers. On the issue of inadequate methodology, the QA chemist often refers a laboratory manager to another laboratory that has more suitable methods or has satisfactory performance. This is viewed as a constructive transfer of information.

The FP and PPWB QA programs' most important feature is the quick response evaluation. Reports indicating results that are suspect are often returned within four weeks; by noting that results are flagged, the laboratory manager can take corrective action to discover the possible error sources within the measurement systems. Anomalous results may relate to a random error, a blunder, an incorrect calibration, poor precision, an unsuitable method, or simply inadequate QC procedures within the laboratory. Laboratory managers agree that the overall effort within the quick response evaluation is constructive.

9.6 References

Caulcutt, R., and R. Boddy. 1983. Statistics for Analytical Chemistry. Chapman and Hall, London, U.K.

Cheam, V., and A.S.Y. Chau. 1982. Manual for the Bi-Monthly Inter-Regional Quality Control Studies. Manuscript No. 48-AMD-82-VC, National Water Research Institute, Burlington, Ont.

Dux, J.P. 1986. Handbook of Quality Assurance for the Analytical Chemistry Laboratory. Van Nostrand Reinhold Co., New York, N.Y.

Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. Technometrics 11: 1–21.

Hunt, D.T.E., and A.L. Wilson. 1988. Chemical Analysis of Water. 2nd ed. The Royal Society of Chemistry, London, U.K.

Mazu, M.J. 1981. Detecting outlying observations. *In* Transactions of the 35th Annual ASQC Quality Congress and Exposition, May 26–29, San Francisco, Calif.

Prairie Provinces Water Board. 1981. Task Two Report, April. Water Quality Analytical Methods Task Force, Regina, Sask.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers Inc., Chelsea, Mich.

10.0 OTHER EVALUATION TECHNIQUES

Chapters 8 and 9 described how interlaboratory data are evaluated for the major external QA studies that are conducted routinely by the NWRI Quality Assurance Group. These methods represent the more usual evaluation techniques, but they are often supplemented by other techniques, as described below.

10.1 Outliers

No measurement system is absolutely free from error or uncertainty. Hence, data from interlaboratory studies will fall into some form of distribution pattern (e.g., normal, skewed). Some results may be extreme and may not fall within a normal and expected range. Such results are referred to as outliers (Grubbs 1969; Mazu 1981). An example of such an outlier is given in Figure 12 in Chapter 9.

Inclusion of outliers into simple statistics such as calculated means (or averages) or variances will taint or bias the calculated estimates. If confidence intervals are required, the inclusion of outliers means broader intervals that may cause analysts to make erroneous decisions about the performance of a method or a laboratory.

How one addresses outliers can be summarized into several categories. One can arbitrarily throw out any results (high or low) that appear suspicious. One can be firm and use all results and exclude nothing. On the other hand, one can be cautious, analyze all methods and results, seek out all possible reasons for a deviant result, and then accept or reject the result or results. A fourth and even more cautious approach is to use traditional statistical techniques to identify outlying values such as described by Grubbs, Dixon, and Ferguson. Many techniques exist, and the associated literature is found in Section 10.3 and in Part V. The Grubbs technique is now routinely applied in the FP and PPWB QA programs outlined in Chapter 9.

Whichever outlier detection method is adopted, it is critical to be cautious in rejecting or removing data. This is especially valid for large interlaboratry data files, because the distribution of results may be bimodal or multimodal, a situation that is influenced by the differences in methods. It may sometimes be necessary to separate data and analyze the individual data groups.

The two principal external programs described earlier (Chapters 8 and 9) differ in their approach to outliers. In Chapter 8, the concept of an outlier is not even considered, because no data are rejected. All extreme results (high or low) are ranked and used to estimate bias for the whole data set. Very extreme results merely accentuate the bias. If many results are extreme, the decision is either severely biased or simply erratic. Although not used in Chapter 8, outliers and their detection certainly are issues implied when flagging data. The flags assigned are not legitimate outliers but simply mild warnings to each laboratory that its results are deviating. A warning implies that some internal investigation should be considered. An interesting and perhaps legitimate outlier appears in Chapter 8, where the program APPRAIS yields an out-of-control statement. This decision occurs when

the laboratory is very competent on all but one of 10 measurements (refer to Figure 2, Chapter 1). In this case (out of control), the isolated result is severely different.

10.2 The Youden Paired-sample Approach

Data are also graphically evaluated by the paired-sample plot technique originated by Youden (Youden and Steiner 1975). The technique requires that the two samples be of similar composition and analyte concentration. It has been successfully used in several national interlaboratory studies. The two common merits that the plot offers are the visual display of the data quality of each laboratory (Cheam and Aspila 1980) and of methods performance (Cheam and Chau 1987). For example, Figure 13, with the circles as acceptability limits, reveals the data quality of arsenic determinations by the various laboratories, whereas Figure 14 effectively compares the performance of various methodologies used for analyzing sulphate ion in coloured waters.

10.3 References

Cheam, V., and K.I. Aspila. 1980. Interlaboratory Quality Control Study No. 26—Arsenic and Selenium in Water. Inland Waters Directorate Report Series No. 68, Burlington, Ont.

Cheam, V., and A.S.Y. Chau. 1987. National Interlaboratory Quality Control Study No. 23—Sulphate in Coloured Waters. Inland Waters Directorate Report Series No. 75, Burlington, Ont.

Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. Technometrics 11: 1–21.

Mazu, M.J. 1981. Detecting outlying observations. *In* Transactions of the 35th Annual ASQC Quality Congress and Exposition, May 26–29, San Francisco, Calif.

Youden, W.J., and E.H. Steiner. 1975. Statistical Manual of the Association of Official Analytical Chemists AOAC, Washington, D.C.

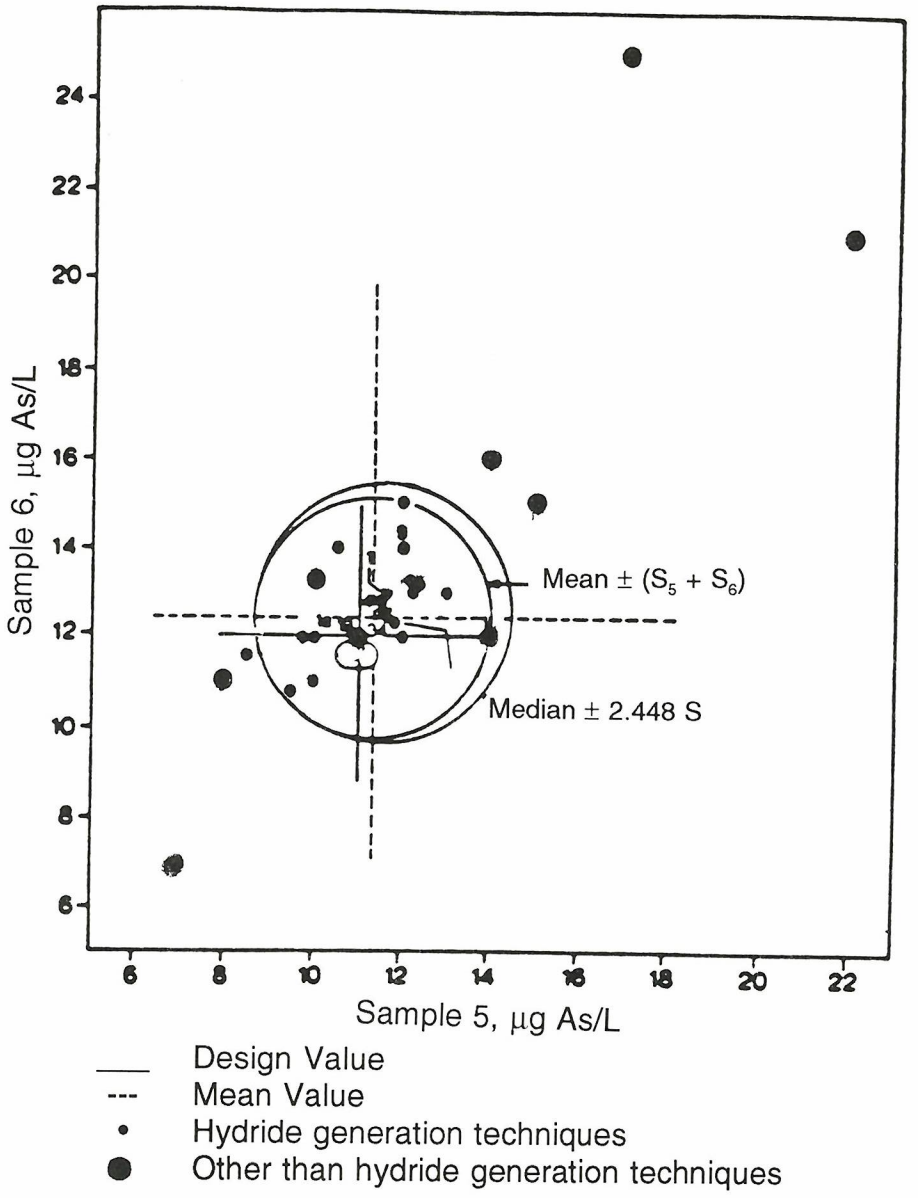


Figure 13. Youden paired-sample plot for samples 5 and 6.

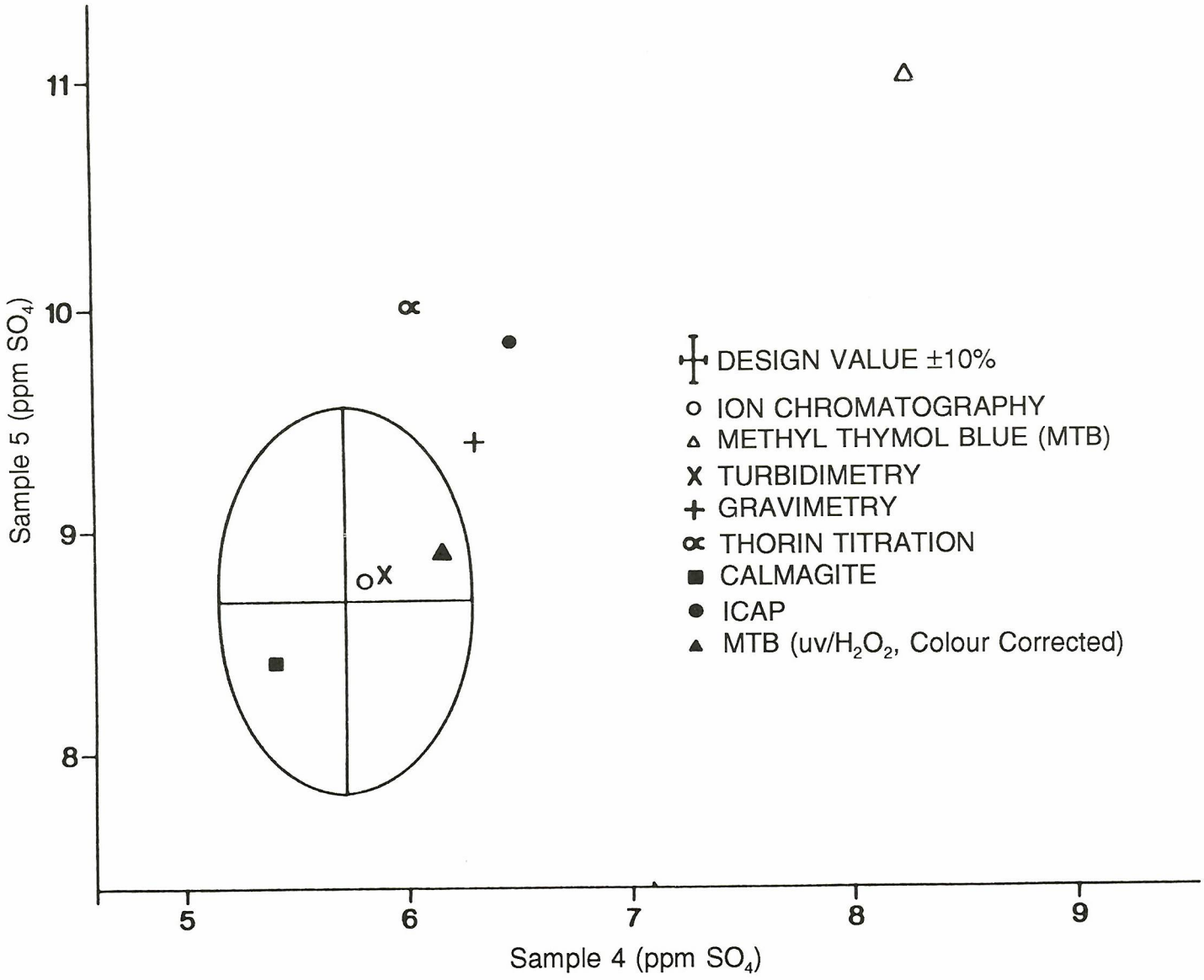


Figure 14. Comparison of the design value with median values in paired samples 4 and 5 determined by various methodologies.

11.0 THE REPORTING PROCESS (FEEDBACK)

When a large, effective interlaboratory study is nearing completion, it is necessary to prepare two reports in quick succession.

The first report to the analyst is often a preliminary computer report highlighting only the data that are specific to that laboratory. This simple report allows the analyst to verify that the transcription of raw data into the computer data base system was correct and clear of errors.

The second report, whether for a large or a small study, is the more important one. It is a report to the analyst, laboratory manager, and any other program person connected to the laboratory (e.g., QC person or program manager). This report is a one- or two-page letter and includes a laboratory-specific performance appraisal that is pertinent to the laboratory. Attached to the letter are all relevant supporting data, such as:

- a description of test samples;
- previous history of test samples;
- summary of historical data on test samples;
- a list of all participants;
- comparison of the performance of all laboratories; and
- a copy of all reported results.

The primary intent of this second report, normally mailed six to eight weeks after a participant has first received the test samples, is to inform him or her of any problems in the measurement system of the participating laboratory.

The third report is a formality and consists of a final document published and circulated for general reference. Effort should be made to complete this report within the year of the study.

The above reporting process relates to how the initiator of an interlaboratory study should relay information to the client laboratories. The specific protocols can be quite flexible depending on the particular program. It is important that there be rapid feedback on any problems perceived, because many measurement problems left unaddressed can, for some large, high-production laboratories, quickly taint an important environmental data base.

The reporting process within a laboratory is also critical. In most cases, the QC specialist responsible for the laboratory should be quickly notified, in addition to the analyst. The process to

initiate change is referred to as remedial action and requires management involvement. This issue is discussed in Chapter 12.

12.0 REMEDIAL ACTION TO QA REPORT

Interlaboratory comparison studies are normally produced not for the academic benefits of the producer or sponsor but for the mutual benefits of the client laboratories and the environmental program. As such, there should always be some reaction to the final results of interlaboratory comparison studies. The reaction will vary at the various levels of the management structure.

12.1 Intralaboratory Action

The bench analyst and laboratory manager are on the front lines and are the individuals who should receive first notice of preliminary reports to each study. If the study had been designed well and the supporting data are clear, concise, and informative, then the bench analyst and the manager involved should be able to use the evidence and react to the appraisal constructively. The usual and most common problem is the calibration (e.g., bias), although in other cases it may be lack of precision (e.g., erratic data).

For large laboratories that have an active interchange of information between analysts, managers, and QC officers, the first-line reaction should be a review of intralaboratory control data obtained at the date when the interlaboratory study test samples were analyzed. This review should be constructive, with analysts comparing their data against those of their peers and examining method issues (if pertinent), the possibility of a calibration failure (e.g., standards), and perhaps simply poor application of the method (e.g., erratic recovery or poor precision). If the study was elaborate in design, then the study report and associated laboratory-specific statements might cover some of these QC issues. The responsibility for much of the data interpretation remains with the analyst and the pertinent manager. Feedback to the originator of the study, although not essential, is sometimes of value.

12.2 Management Action

Large facilities very often have project or program managers who have a responsibility to manage several environmental projects for which analyses are carried out in one or more laboratories. These individuals should be kept informed on the performance of their laboratories. It is a line of authority that is often overlooked, but it is an important area, especially where performance in the laboratory has been judged as poor and requires internal review. How management reacts is an internal matter, but react it must, if the data base it has, or will produce, is to be protected and retained as credible for program requirements.

The nature and process of reaction to a study can vary. The appraisal that states satisfactory, well done, or average performance may be met with satisfaction. An average appraisal may result in an internal review or audit by laboratory staff. A poor performance, with severe bias, out-of-control statements, or the use of the term "erratic," may impel management and program managers to request the laboratory to cease data production until proof of adequate performance can be provided. This

extreme reaction would be an internal decision reached by management after careful review of the study report and the internal data quality objectives of the program.

On matters of performance, it remains the responsibility of the interlaboratory program and the originator of studies to retain spare test samples on hand to support the follow-up needs of the client laboratory if assistance is required to investigate identified problems. As a summary, the remedial actions to any study should always be viewed as a constructive process and indeed one with all parties involved.

12.3 Users of Data

Very often, the users of environmental data are the least informed on issues of quality. Many agencies tacitly assume that data are acceptable, because they have been continually told that the laboratory producing the data has internal QC. This can become an increasingly dangerous position if the agency data files originate from several laboratories. Experience has shown that different data bases can be disjointed. Data users should maintain files defining internal and external performance. This performance must match or exceed the objectives of their data needs. If they do not, acceptance limits must be defined.

Glossary of NWRI External QA Programs

CAPCO — Canadian Association of Pesticide Control Officials.

Eulerian — An external QA program that supports the "Eulerian Model Field Study" and involves external monitoring of the four primary laboratories that report precipitation data to Environment Canada, Ontario Ministry of Environment, U.S. Environmental Protection Agency, and the U.S.-based Electric Power Research Institute. Twenty-four studies (one per month) involving eight laboratories are in progress. The program parallels LRTAP but is specific to rain water.

FICP — Federal Interdepartmental Committee on Pesticides. This QA program is interdepartmental and involves toxic organics in a wide variety of substrates. The NWRI involvement (one study per year) involves water and sediments for a variety of toxic organics.

FP — The Federal–Provincial QA program. This QA program was called the Inter-Regional Quality Control Program (IRQC) when implemented in 1974. The studies are monthly and involve trace metals and major ions. Over 158 studies have been completed.

IJC — International Joint Commission. This QA program is in reference to the Canada–U.S. Great Lakes Water Quality Agreement, Great Lakes International Surveillance Program (GLISP). The external QA studies are provided two to four times per year to about 30 to 100 laboratories. Studies include:

- phosphorus in water or sewage plant effluents;
- toxic organics and inorganics in fish homogenates and sediments; and
- major ions, nutrients, physical properties, and trace metals in water.

Thirty studies have been completed (1976 to 1988).

LRTAP — Long Range Transport of Airborne Pollutants. This QA program involves almost 100 laboratories. Each laboratory is provided with 10 different unpreserved soft waters three times per year and is requested to analyze up to 23 constituents (major ions, nutrients, and physical parameters). Twenty studies have been completed.

LRTAPP — The same program as the LRTAP, but LRTAPP refers to the studies involving plant materials for nutrients and metals. These studies are provided through the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario (Dr. Ian Morrison).

National The NWRI QA program presented to all federal, provincial, university, and private sector laboratories. This large program initiated in 1970 involves a diverse series of substrates and constituents. Thirty-seven studies have been completed.

PPWB Prairie Provinces Water Board. This QA program involves laboratories in Alberta, Saskatchewan, and Manitoba. The program runs concurrent with the FP and is similar in design (see Chapter 8). Thirty-six studies have been completed.

UGLCCS Upper Great Lakes Connecting Channels Studies. This QA program, recently completed, was a binational program (Canada–U.S.).

Part V

QUALITY ASSURANCE GUIDELINES AND PRINCIPLES FOR THE HANDLING AND MANAGEMENT OF WATER QUALITY DATA

James E. Gaskin

Contents

Page

PREFACE viii

ACKNOWLEDGMENTS x

1.0 INTRODUCTION V-1
- 1.1 The need for a water quality data handling and data management QA program V-1
- 1.2 The objective of the data handling and data management QA program V-2
- 1.3 The QA manual for water quality data management V-3
- 1.4 The interrelationship between the QA activities for sampling, laboratory analysis, and data management V-5
- 1.5 References V-6

2.0 WATER QUALITY DATA COLLECTION, RECORDING, DOCUMENTATION, AND HANDLING V-8
- 2.1 Data collection V-8
- 2.2 Data recording and documentation V-8
- 2.3 Auxiliary data V-10
- 2.4 Data handling V-11
- 2.5 References V-12

3.0 DATA SCREENING AND VALIDATION V-13
- 3.1 Data screening and validation procedures V-13
  - 3.1.1 Data identification checks V-13
  - 3.1.2 Unusual events review V-14
  - 3.1.3 Checks for transmittal errors V-14
  - 3.1.4 Checks for temporal continuity V-14
  - 3.1.5 Flagged or rejected data V-14
  - 3.1.6 Checks for errors in automatic data processing V-14
- 3.2 Tests for internal consistency V-16
  - 3.2.1 Data plots V-16
  - 3.2.2 Dixon Ratio Test V-17
    - 3.2.2.1 Testing a single high value V-17
    - 3.2.2.2 Testing a pair of high values V-18
  - 3.2.3 Grubbs Test V-19
  - 3.2.4 The Gap Test V-20
  - 3.2.5 Regression analysis V-22
    - 3.2.5.1 Simple linear regression analysis V-22
    - 3.2.5.2 Multiple regression V-27

Contents (Continued)

Page

3.2.5.3 Curvilinear regression V-29
3.2.6 Control charting V-30
3.2.6.1 Application of control charts to data validation V-30
3.2.7 Sample consistency checks V-31
3.2.7.1 Ion balances V-31
3.2.7.2 Anion/cation ratio V-32
3.2.7.3 Measured versus theoretical conductance V-33
3.2.7.4 Ion exchange method V-33
3.2.7.5 pH checks V-34
3.2.7.6 Ratios of specific parameters V-34
3.2.7.7 Mass balances V-34
3.3 Tests for consistency of parallel data sets V-35
3.3.1 The Sign Test V-35
3.3.2 Wilcoxon Signed-Rank Test V-37
3.3.3 The Rank Sum Test V-39
3.4 Detection limits V-40
3.5 References V-40

4.0 DATA VERIFICATION V-44

5.0 DATA QUALITY ASSESSMENT V-47
5.1 Precision V-47
5.1.1 Intralaboratory and interlaboratory precision V-50
5.1.2 Total precision V-51
5.1.3 Precision and concentration V-52
5.1.4 Interpretation of precision data V-52
5.2 Accuracy and bias V-52
5.2.1 Significance of the analytical error V-55
5.2.2 Accuracy and concentration V-55
5.3 Method detection limit V-55
5.3.1 Limits for practical detection, quantitation, and validation V-59
5.4 Other data quality assessment criteria V-60
5.4.1 Completeness and representativeness V-60
5.4.2 Data comparability and compatibility V-61
5.5 Data review and evaluation V-61
5.6 References V-62

6.0 STATISTICAL TECHNIQUES FOR WATER QUALITY DATA ANALYSIS V-63
6.1 Descriptive statistics V-64

Contents (Continued)

Page

6.2 Theoretical probability density functions for discontinuous (discrete) data . . V-64
6.3 Theoretical probability density functions for continuous or measured data . . V-64
6.4 Confidence limits V-65
6.5 Inferential statistics V-65
6.6 Testing for normality of data measurement distribution V-66
6.7 Testing for homogeneity of variances V-66
6.8 Comparison of the means from two populations V-66
6.9 Simple linear regression V-66
6.10 Analysis of covariance V-67
6.11 Multiple regression V-67
6.12 Curvilinear regression V-67
6.13 Nonparametric statistics V-67
6.14 Data modification V-68
6.15 Sophisticated data handling techniques V-70
6.16 References V-70

7.0 DATA PRESENTATION AND DATA REPORTING V-72
7.1 Data presentation V-72
7.1.1 Presentation of statistical parameters V-72
7.1.2 Supporting data V-72
7.1.3 Tabular and graphical presentation V-73
7.1.4 Confidence limits V-73
7.1.5 Rounding-off procedures for numerical values V-73
7.1.6 Control chart method for data presentation V-74
7.2 Data reporting V-75
7.2.1 Internal data reporting V-75
7.2.2 External reporting to clients and data users V-76
7.2.3 Reporting of analytical values V-76
7.2.4 Quality assurance reporting V-77
7.2.4.1 Data quality reports to management V-77
7.2.4.2 Data quality reports to users V-78
7.3 References V-79

8.0 DATA CUSTODY AND DATA TRANSFER V-80
8.1 Data custody V-80
8.2 Data transfer V-80

9.0 MANAGEMENT OF WATER QUALITY DATA BASES V-82
9.1 Data processing V-82

Contents (Continued)

Page

9.2 Laboratory data base management V-82
9.3 Data management information system V-83
9.4 Data base structure V-84
9.5 Data storage and retrieval V-85
9.6 References V-87

10.0 TRAINING IN DATA HANDLING V-88

11.0 DATA AUDITS V-89
11.1 Objectives of the data audit program V-89
11.2 Audit team V-89
11.3 Audit visit V-90
11.4 Audit evaluation and reporting V-90
11.5 References V-91

APPENDIX A: Statistical tables V-92

APPENDIX B: Factors for estimating control limits of Shewart chart V-97

Tables

Page

1. Frequency distribution of a sample of specific conductance data V-22
2. Data for suspended sediment concentrations (mg/L) in samples taken by two sampling devices at an aquatic site in Alberta, Canada V-26
3. Statistical parameters related to the regression line for suspended sediment concentrations in samples collected by two water quality samplers (A & B) V-27
4. Application of the Wilcoxon Signed-Rank Test to water quality data for samples collected by two sampling devices at the Red Deer River at Bindloss, Alberta, Canada (1980) V-38
5. Application of the Rank Sum Test to two sets of total nitrogen data collected on two different dates by the same sampling device at the South Saskatchewan River, Highway 41, Alberta, Canada (1980) V-41
6. Suggested format for data validity and verification comment codes V-46
7. Duplicate and triplicate determinations for dissolved chloride in a number of water quality samples V-49
8. Measurements of various analyte concentrations to compare detection limit determinations by two different mathematical approaches V-59

Figures

Page

1. Schematic of data flow in a water quality monitoring program V-7
2. Linear relationship between D-77 and WQS values for total sediment concentrations V-28

Preface

The Water Quality Branch has for a long time recognized the importance of having a quality assurance (QA) program that would govern the collection of data during the sampling and laboratory analytical measurement stages of the monitoring program. Thus, the data QA activity has been entrenched as one of the basic components of the monitoring program in support of the QA and quality control (QC) activities in the sampling and analytical measurement operations.

This text is principally intended as a QA manual for the handling and management of water quality data, and it can be used concurrently with the QA manuals on field sampling (Part II) and intralaboratory/interlaboratory analytical work (Parts III and IV, respectively). The subject material is divided into several parts:

- data collection;
- data recording;
- data screening;
- data validation;
- data verification;
- data quality assessment;
- statistical techniques for water quality data analysis;
- data presentation;
- data reporting;
- data custody;
- data transfer;
- management of water quality data bases;
- training in data handling; and
- data audits.

The basic areas associated with each of these topics are discussed in separate sections with some occasional overlap.

Particular emphasis is placed on data validation as a systematic process for reviewing and ensuring the integrity of water quality data. Demonstration of the use of statistical techniques is supported by worked examples. Summaries of statistical procedures are clearly set out, and explanations on the reading of statistical tables in the appendices are also included in the appropriate sections.

Acknowledgments

We are indebted to the following Water Quality Branch and National Water Research Institute personnel for reviewing the text: P. Brooksbank, A.S. Chau, T. Dafoe, A.R. Davis, A. Demayo, A. Germain, J. Lawrence, J. Merriman, F. Philbert, T. Pollock, C. Thorpe, E. Watt, and J.-G. Zakrevsky.

A special word of thanks is due to Mrs. Francine Goulet, who patiently typed the manuscript.

Quality Assurance Guidelines and Principles for the Handling and Management of Water Quality Data

James E. Gaskin

1.0 INTRODUCTION

1.1 The Need for a Water Quality Data Handling and Data Management QA Program

The need for an effective data quality assurance (QA) program attached to the data collection process has been demonstrated over and over again in water quality monitoring programs that have data-reporting commitments to clients and data users. Good-quality data that have been produced under the guidance of accepted and proven QA principles find ready acceptance and use by management personnel and other intended audiences.

QA practices in water quality data handling and data management involve a number of systematic processes and protocols that are designed to provide a framework for equipping water quality measurements with a high degree of credibility.

A major step in ensuring good-quality data involves collecting representative samples and maintaining the integrity of these samples. When this step has been successfully carried out, a number of essential QA and quality control (QC) guidelines can be imposed on the analytical measurement and data collection systems in order to bring the data generating processes into a state of statistical control and maintain them there. A measurement process in a state of statistical control is fundamental to the generation of precise and accurate data. Data originating from systems that are not operating under a statistical control influence will be open to varying degrees of errors and biases, depending on the magnitude of the systems' departure from optimum operating conditions.

The usefulness of any data set is directly proportional to the quality of the data; the data in turn must reflect an acceptable degree of precision, accuracy, completeness, and confidence attributes. Data characteristics such as these will vary from one water quality parameter to another, or from one project to another within a monitoring program, depending on the projects' objectives and other related factors. However, emphasis must still be placed on the need for obtaining the most precise and accurate values that the measuring systems will deliver within the confines of any existing limitations.

From the beginning of the sampling operation to the stage where the collected data undergo analysis, interpretation, and evaluation, there must be clear and precise documentation encompassing QA guidelines and principles that cover every aspect of data collection.

An appropriately designed and comprehensive data QA program can assist management not only in the evaluation of project-related data, but also in the evaluation of the projects themselves and the overall monitoring program.

1.2 The Objective of the Data Handling and Data Management QA Program

The data handling and data management QA program focuses on the production of data that have the characteristics of accuracy, precision, completeness, and representativeness.

The quality of data is to a large extent evaluated on the basis of their uncertainty when compared with end-use requirements. When data are accompanied with consistency and their uncertainty is small when compared with the requirements, the data are considered to be of adequate quality. When excessively variable or when the level of uncertainty exceeds the needs, the data may be said to be of low or inadequate quality. The assessment of data quality can thus be a relative determination. What is high quality for one situation could be unacceptable in another (Taylor 1985).

The virtue of water quality data may be assessed on the basis of two aspects: the accuracy of identification of the parameter or variable measured, and the numerical accuracy. The qualitative identification must be made without reasonable doubt, and the quantitative measurements must be conducted precisely and accurately, in such a manner that any error or uncertainty in the measurements can be tagged with a stated probability. To this end, measurements must be made in such a way as to provide statistical predictability.

The objective of the data management QA program can be attained through the achievement of a number of specified goals. Foremost among these goals are the following:

- the operation of a good QA/QC sampling and field program;
- an effective and efficient laboratory analytical program that incorporates good laboratory practices; and
- the commitment of management to an effective and viable QA program that spans all levels of involvement, beginning at the monitoring design stage and concluding with the issuing of verified data reports.

In short, the data handling and management QA program is an ongoing process that seeks to ensure that the best possible data are produced under the most prudent and statistically controlled measurement conditions, together with appropriate, effective, and sound data handling techniques.

1.3 QA Manual for Water Quality Data Management

The QA manual for data handling and data management is essentially an outline of systematic principles and protocols that have been established for equipping water quality data with characteristics of good quality and reliability.

For the data QA program to be effective, the sampling and laboratory QA programs must also be working effectively and efficiently. An impressive set of standard guidelines established to control the manipulation and use of data will serve no valuable purpose if the data are collected under less than proper and less than experimentally sound conditions.

The steps and directions in the data QA manual must deal with the following aspects of data handling and data management:

(1) Data recording and documentation
- data acquisition
- transcription errors
- data processing
- computerized recorded data
- hand-written data
- data storage
- data retrieval

(2) Data transmission
- appropriate data format
- clarity and comprehensiveness
- adherence to turnaround times
- confidentiality
- interlaboratory and interagency correspondence

(3) Data validation
- data inspection
- statistical testing
- internal consistency checks
- control limits, QC charts
- detection limits
- data reduction
- data corrections
- transmittal errors
- data acceptance/rejection criteria
- data plots
- test for suitability of data
- regression analysis

- test for goodness of fit and outliers
- inter-method comparisons
- inter-instrument comparisons
- support data records
- sample consistency
- data comparability and compatibility

(4) Data verification

- data checking
- correct detection limits
- calculation errors
- decimalization errors
- appropriate measurement units
- transmission errors
- relationship between data results and program or project objectives
- water quality objectives (appropriate)
- sampling records
- transportation and preservation records
- analytical records
- support data records

(5) Data analysis

- use of appropriate statistical techniques
- data interpretation
- data integrity
- accuracy
- precision
- completeness
- method detection limit
- data representativeness
- data comparison and compatibility with specified external data sets
- QC charts
- data quality review and assessment
- significant QA problems
- corrective action procedures
- audit checks of data

(6) Data reporting

- performance evaluation reports
- data reports
- interpretive reports
- data presentation
- turnaround times

- report control procedures
- responsibility centres
- confidentiality
- provisional reports
- final reports
- feedback from clients and data users re: data reports.

The use of the preceding QA steps in conjunction with good sampling and good laboratory analytical practices can ensure that data are accurate and reliable.

1.4 The Interrelationship Between the QA Activities for Sampling, Laboratory Analysis, and Data Management

Each QA program, although specific to its own operational activity, must at the same time be complementary to the other two QA programs if the goal of good-quality data is to be achieved.

The three QA programs are bound together by a number of common threads. These take the form of:

- the monitoring system design, which can be viewed as a step-by-step process beginning at the evaluation of information expectations and concluding with the monitoring program evaluations (Sanders *et al.* 1983);

- program documentation consisting of
  - the overall QA program plan
  - the QA project plan
  - the program implementation plan;

- the QA/QC officer;

- the QA/QC Technical and Advisory Committee;

- the quality and representativeness of the collected samples, the laboratory analyses, and the collection and handling of the experimental data;

- management of the three interrelated QA programs; and

- the end use of the experimental data.

Shortcomings in any of the three QA programs will invariably and possibly proportionately lower the quality of the final data results.

Figure 1 illustrates the sequential flow of data and information between the operational activities that make up the monitoring program.

1.5 References

Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian, and V. Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, Colo.

Taylor, J.K. 1985. Principles of Quality Assurance of Chemical Measurements. Center for Analytical Chemistry, National Bureau of Standards, U.S. Department of Commerce, Gaithersburg, Md.

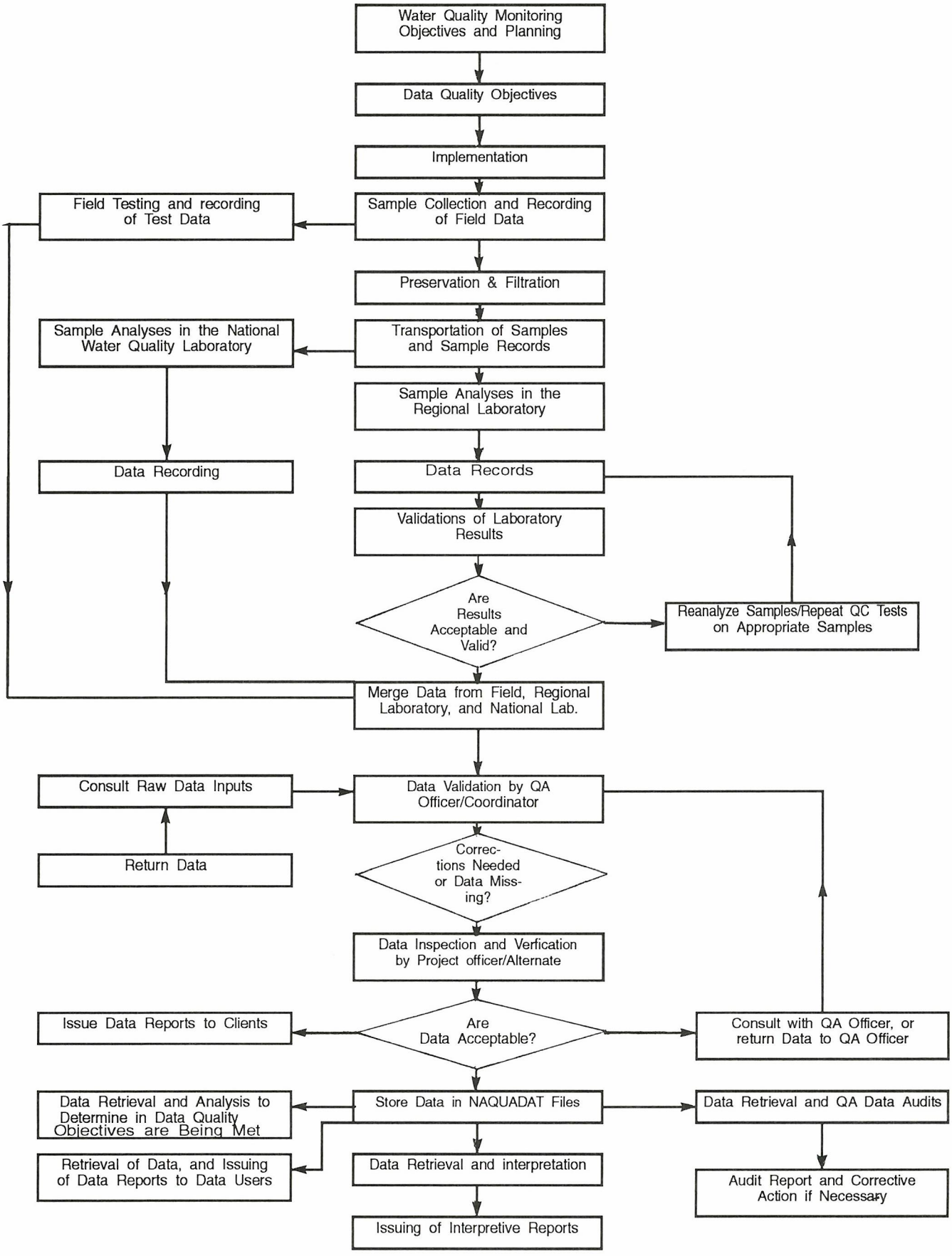


Figure 1. Schematic of data flow in a water quality monitoring program.

2.0 WATER QUALITY DATA COLLECTION, RECORDING, DOCUMENTATION, AND HANDLING

2.1 Data Collection

Data acquisition consists of the collection of data not only from the sample collection and laboratory analysis operations, but also from the monitoring design phase.

The goal of the data QA program as it relates to data acquisition is to ensure that all data and data records for any given sample are properly documented, complete, appropriately labelled and filed (stored), accessible, and easily retrievable.

Collected data should be synthesized into records that are accurate, legible, and comprehensible.

The acquired data and documentation should include chain-of-custody records, sample tracking records, information recorded in the analysts' notebooks and in instrumentation logbooks, hard-copy output from instruments (e.g., chart recordings, analog and digital printouts, chromatographs), computer printouts, raw data summaries, reduced data, field reports, analytical reports, methodology reports, and QA and QC reports.

2.2 Data Recording and Documentation

Data documentation should include the processes used in the calculations and computations of the data, corrections required, adjustment to standard conditions, normalization of data, computer programs, statistical procedures used to report data, method for evaluating limits of uncertainty, corrections for systematic errors, and source of all constants used in calculations.

Data in the form of charts, instrument recordings, and printouts should be given suitable identification numbers and maintained in a manner consistent with good record-keeping practices. Laboratory record books must refer to the location and identification of such records.

All calibrations and standardizations should be fully documented, and the data should provide clear traceability to the calibrations/standardizations to which they are related.

In order to avoid transcription errors, recording and documentation of data must be done accurately, and care and attention to detail must be exercised.

Systematic inspection and periodic review of notebooks and similar primary records are recommended to ensure the general quality of their contents. Changes or revisions of notebook entries must be justified and documented. Any changes should be made by crossing out the original entry and

substituting the new value. The person making the change should initial the entry and state why the change was made. No erasures of records or data should be permitted.

Maintenance of equipment must be suitably recorded. Routine maintenance may be indicated by labels on the equipment. Maintenance that results in modification of equipment must be described in sufficient detail and recorded in the operation manual for the particular equipment.

Field and laboratory records should be retained in permanent files for a length of time set by government, other legal requirements, or the employing institution, whichever is longer. Bound notebooks are preferred to looseleaf-type notebooks.

In summary, the QA objective of network documentation is to ensure that documents are produced and distributed in a manner suitable for the proper operation of the network and the fulfillment of its objectives. To support the QA documentation objective, QA procedures should, according to Vet and Onlock (1983), include:

- determining the type and contents of documents that should be produced; and
- implementing a document control system.

Suggested documents for the Water Quality Branch (WQB) QA/QC programs are:

- field operator's manual;
- field and laboratory instrumentation manual;
- laboratory analysis manual;
- data handling and management manual;
- QA/QC manuals;
- detailed job responsibility descriptions (e.g., for field and laboratory technicians and managers);
- sampling design and network description;
- WQB QA plan;
- site description document;
- QA plan;
- interim data reports;

- annual (final) data reports;
- data transmittal documentation;
- data analysis reports;
- QA reports;
- annual calibration reports;
- annual preventive maintenance reports;
- suppliers, apparatus, and instrument inventory reports; and
- field operations summary reports.

Suggested document control system components include:

- implementation of a document index system to record the production, revision, and distribution of documents;
- implementation of a document revision and updating control system (using a paging index system); and
- implementation of an operations manual distribution and recording system to ensure that all staff receive updates and revisions to manuals.

Finally, periodic checks must be made to ensure that the document control system is being used properly and effectively, and quality assessment forms should be used as the main tool for documentation of QC data. The use of quality assessment forms aids in both the evaluation of QC data and the appropriate feedback to laboratory managers, supervisors, and analysts.

2.3 Auxiliary Data

The practice of collecting supporting information during all phases of the measurement programs must be encouraged, and these auxiliary data must be easily available via EDP facilities or field and laboratory records. The use of auxiliary data may become necessary during the data interpretation process.

The following information may be considered as auxiliary data:

- data charts and printouts;

- equipment performance records;
- calibration records;
- operation logs;
- environmental conditions prior to and during sampling;
- measurement comparison records;
- QC and system audit records; and
- records of corrective actions.

Auxiliary data should be collected throughout the measurement program and reviewed periodically. Auxiliary data are important in determining the validity of the measurement program data. For example, auxiliary or support data could be used in deciding whether or not an outlier is a valid value or an artifact. Ideally, unusual conditions should be recorded on the field data reports. Failure to report such occurrences may place the acceptance of important data in jeopardy or may lead to the rejection of those data.

2.4 Data Handling

Electronic data handling, data reduction, and data storage systems are important parts of many analytical systems. They greatly facilitate data handling, and control errors due to misreading, faulty transcription, or miscalculations. However, the performance of the data system must be tested regularly to ensure that it is working properly and correctly. This should be done periodically by using known data that have already been calculated. These tests must have sufficient accuracy and precision to provide a reliable examination of the data handling system.

In data handling processes, the following subjects should be addressed as appropriate:

- procedures used in the calculation of data; corrections required; adjustment to standard conditions; normalization of data; computer programs;
- statistical procedures used to report data; method for evaluating limits of uncertainty;
- corrections for systematic errors;
- source of all constants used in calculations, such as Faraday constant, atomic weights, molecular weights, ideal gas constant, and constants derived from calibrations (e.g., slopes of graphs); and

- checks on accuracy of calculations; use of significant figures; propagation of errors; analysis of variance.

2.5 References

Vet, R.J., and S.G. Onlock. 1983. The Canadian Air and Precipitation Monitoring Network Quality Assurance Plan for Precipitation Monitoring Systems. CSC 110. 194-3-1, Air Quality and Inter-Environmental Research Branch, Atmospheric Environment Service, Environment Canada, Downsview, Ont.

3.0 DATA SCREENING AND VALIDATION

Data screening and validation are essential elements of data QA and are effective processes for reviewing a body of data against a set of criteria, so that assurance can be provided that the data of interest are adequate for their intended use. Data screening and validation are two of the primary activities for ensuring the reporting and use of good data, whereas QC activities are designed for the purpose of acquiring good data.

The screening and validation processes are most effectively performed at the location of the data collection, because questionable data can be most easily checked at this level.

The procedures included in the two processes vary with the method of obtaining data. In one application, the data may be recorded on a field data form and returned to the laboratory for further handling prior to storage in a data bank. In another application, the data may be obtained directly from an automatic monitoring device.

The screening and validation processes include not only the identification or flagging of questionable data, but also the investigation of apparent anomalies.

3.1 Data Screening and Validation Procedures

The following outline describes some of the routine procedures used in the screening and validation of water quality data. A number of techniques that are described in the literature (Rhodes 1977; Nelson *et al.* 1980) are also incorporated.

3.1.1 Data Identification Checks

Data with improper identification codes are of little use. Identification fields that must be correct are:

- time (start and stop time and date);
- location;
- sampling/analytical method code;
- parameter;
- parameter code (NAQUADAT No.);
- submitter I.D. (identification of sampling personnel);

- analytical value; and

- project and agency.

Because most of the identification problems are the result of human error, an individual other than the original person preparing the forms should verify the data coding forms prior to using the data for computer entry or manual summary. The data listings should also be checked after entry into a computer system or data bank.

3.1.2 Unusual Events Review

A log should be kept by the water quality agency in order to record unusual events (e.g., inclement weather, dead fish, algal growth, oil slicks or grease on the water surface, irregularities in morphology, construction activity, dust storms, vibration of analytical instruments due to the passage of motor vehicles), even if these events seem trivial at the time of their occurrence. Depending on the purpose of data collection, this information could also be used to explain why no data were reported for a specified time interval, or it could be the basis for deleting data from a file for specific analytical purposes.

3.1.3 Checks for Transmittal Errors

For paperwork systems, simple checks should be made to assure that the data have not been incorrectly transferred from one paper (medium) to another. With more sophisticated electronic and computer data handling and with telemetry of data, checks should be made to assure that the data have not been changed in the transmission or transfer process (see Section 3.1.6).

3.1.4 Checks for Temporal Continuity

Data validation should include checks for continuity with respect to time, as might be evaluated by having a chronological plot of the data to look for breaks, discontinuities or gaps, etc.

3.1.5 Flagged or Rejected Data

A data validation system should include a scheme for flagging questionable data and should make provision for complete rejection of the invalidated data for any type of intended use. The basis for data rejection should be centred around a number of criteria, which include such elements as larger-than-maximum values, non-historical method detection limits (MDLs), numerical errors, round-off errors, anomalous values, ionic imbalances, and mass imbalances.

3.1.6 Checks for Errors in Automatic Data Processing

There are currently a multitude of procedures in general use for detecting and, when possible, correcting errors as they occur in computer systems. Several of the procedures fall within the

categories of internal, historical, and parallel data consistency checks, whereas others are peculiar to data processing (U.S. Department of Commerce 1978). Some of the latter techniques are:

- checks on data sequence (e.g., input data are checked for correct time sequence);
- editing by classification of category, class limits, normal limits, and trend limits (e.g., the behaviour of an individual data item can be compared with its previous behaviour or the behaviour of the aggregate of individuals in its group);
- parallel check calculations, useful when the same results can be obtained by two independent calculation procedures;
- built-in test data, verification tests, and diagnostics to provide a test environment without the risk of allowing an unchecked program access to real files;
- on-line testing to exercise the fault detection logic;
- removal of redundancy in batches, files, and inputs in order to improve reliability; and
- addition of quality flags to items in a data base in order to avoid a mismatch between the data quality and the intended use.

Each data handling organization (e.g., laboratory, region) should develop a data flow diagram that indicates the steps in data handling at which an error can occur. In general, these steps can be classified as:

- data input;
- data transmission;
- data processing; and
- data output.

Principal sources of error in the data input stage include keypunch errors and the use of mislabelled computer files. A review should always be conducted for errors of this kind. This can be done conveniently if the input data are included in the computer output in a format for easy review.

The principal result of transmission errors is the loss or alteration of data. A simple way to check for transmission errors is to transmit the data a second time and then pair the two data streams. Gaps and alterations will immediately become apparent, unless the transmission error is systematic.

Processing errors are more difficult to characterize and detect. They are usually caused by deficiencies in the computer programs that manipulate the data files, perform mathematical calculations and computations, and format the output results. A standard method of checking for processing errors is to make up a small but typical data set, perform the appropriate data manipulations and calculations by hand, and compare these results with the results from the data processing system. This procedure should provide a good check if the data processing errors are not related to the size of the data set. This procedure is also appropriate for checking the part of the data processing system that outputs the data.

3.2 Tests for Internal Consistency

Internal consistency tests can be used to check for values in a data set that appear atypical when compared with values of the whole data set. Common anomalies of this type include unusually high or low values (outliers) and large differences in adjacent values.

There are a number of potential sources from which outlier values can originate. These include faulty instrument or system components, inaccurate instrument readings, errors in transcribing data, and calculation errors. Any occurrence of frequent outliers underlies the need for closer control of the data collection process. The presence of outliers may also lead to incorrect conclusions during data analysis.

Data outliers should not be eliminated without a review of the data and supporting information acquired during the sample collection and measurement programs. Outliers attributable to unusual conditions of measurement, such as abnormal local environmental conditions during the time of sample collection or during analytical measurement, not only must be recorded, but must be identified and dealt with on a case-by-case basis. Taylor (1987) suggested that in some cases it may be possible to take corrective actions and salvage what would otherwise be a bad measurement. Records should be kept of the problems and corrective actions used for guidance in possibly modifying QC procedures. Pollock (1988) also suggested that when bad data or outliers have been discovered, flags should be attached to the data if the data are currently held in a data information and retrieval system. Alternatively, the data should be archived or discarded depending on whether the data will be re-examined in light of corresponding new data.

The following statistical tests are applicable to the measurement of data errors but will not detect errors that can alter all values of the data set by either an additive or multiplicative factor (e.g., an error in the use of a meter or recorder).

3.2.1 Data Plots

Data plotting (including strip chart records) is one of the most effective means of identifying possible data anomalies. However, plotting all data points may require considerable manual effort or computer time. Nevertheless, data plots will often identify unusual data that would not ordinarily be identified by other internal consistency tests.

Although data plots (e.g., regression analysis) are particularly appropriate for checking the internal consistency of data, they can also be used for checking historical consistency (e.g., the Shewart control chart) and parallel consistency (e.g., intersite or intersampler correlation tests).

3.2.2 Dixon Ratio Test

The Dixon Ratio Test (Barnett and Lewis 1978; U.S. Environmental Protection Agency 1978) is the simplest of the statistical tests recommended for evaluating the internal consistency of data.

3.2.2.1 Testing a Single High Value

The only data preparation that is required for testing a single high value is the identification of the lowest and the highest two or three values (depending on N, the total number of values), where $X_1, X_2, X_3, \ldots, X_{N-2}, X_{N-1}, X_N$ is the arrangement of the lowest three and the highest three values in ascending order of magnitude. The statistic of interest for $N \leq 7$ is the ratio (R) of the difference between the highest and second highest values to the difference between the highest and lowest values in the data set, the latter difference corresponding to the range. Thus,

$$R = \frac{X_N - X_{N-1}}{X_N - X_1} \qquad (1)$$

The ratio R is a fraction between 0 and 1. As this ratio increases, the probability that the highest value is consistent with the rest of the data—that is, $p(H_0$ is true)— decreases. Table A-1 (Appendix A) lists critical values of R associated with $p(H_0$ is true) = 1, 5, and 10% for normally distributed data sets. Critical values for $N > 25$ are not listed in the table, and the test procedure changes with changes in N (e.g., Equation 1 is used for $N \leq 7$).

The Dixon Ratio Test is not recommended for large data sets, which possess the potential of having fluctuating and wider variabilities among individual data elements than would be the case in smaller data sets.

The above procedure assumes that the data are normally and independently distributed. Because normal distributions are symmetrical in shape, the mean and median values are equivalent. However, a sizeable portion of water quality data is not usually normally distributed. Data that are nonnormal are generally positively skewed (mean > mode > median). In these cases, a Dixon Ratio Test based on the lognormal distribution may be more appropriate; thus, R can be calculated by using

$$R = \frac{\ln X_N - \ln X_{N-1}}{\ln X_N - \ln X_1} \qquad (2)$$

Equation 2 is less sensitive to outliers of a normal distribution than Equation 1 and should be used only for testing data that are known to be adequately fitted by the lognormal distribution. If the

data are not normally distributed and if there is uncertainty regarding the adequacy of the lognormal distribution, the Gap Test described in Section 3.2.4 should be considered.

The following example illustrates the use of the Dixon Ratio Test in evaluating 16 turbidity results for replicated samples taken from the South Saskatchewan River at Highway 41, Alberta, Canada (May 6, 1980).

* * *

EXAMPLE:

Turbidity measurements (NTUs) obtained are 92; 77; 82; 85; 37; 81; 82; 71; 60; 51; 82; 71; 72; 68; 73, 70.

Test for normality in the data distribution, and investigate the respective R_{22} value for the data at p = 0.01, 0.05, and 0.1.

Application of Chebyshev's Inequality Test or the Empirical Test for normality in data distribution shows that more than 90% of the data lie within ±2 standard deviations, as required. The Dixon Ratio Test for normal distribution may then be applied to test if the first data value of 37 is an outlier among the other values (the measurements are rearranged so that $X_1 \leq X_2 \leq X_3 \ldots \leq X_{N-2} \leq X_{N-1} \leq X_N$). Using the formula for γ_{22} and N = 16 in Table A-1 (Appendix A) for the rearranged values of 37; 51; 60; 68; 70; 71; 71; 72; 73; 77; 81; 82; 82; 82; 85; 92,

$$\gamma_{22} = \frac{X_3 - X_1}{X_{N-2} - X_1} = \frac{60 - 37}{82 - 37} = \frac{23}{45} = 0.511$$

The calculated γ_{22} value is larger than the respective 10 and 5% critical values in Table A-1 but less than the 1% critical value; consequently, it can be stated that the value of 37 is an outlier at significance levels of 0.05 and 0.1, but not at the 0.01 level.

In a similar way, the data value of 92 can be tested as a single high value (i.e., as an outlier).

$$\gamma_{22} = \frac{X_N - X_{N-2}}{X_N - X_3} = \frac{92 - 82}{92 - 60} = \frac{10}{32} = 0.313$$

Comparing the value of 0.313 with the 10, 5, and 1% critical values in Table A-1 shows that 92 is not an outlier in the given data set at either of the respective significance levels.

* * *

3.2.2.2 Testing a Pair of High Values

The Dixon Ratio Test described in Section 3.2.2.1 will not identify data sets that have two or more outliers of similar magnitude. To test for a pair of high values, calculate

$$R = \frac{X_N - X_{N-2}}{X_N - X_1} \quad (3)$$

for normal data sets and

$$R = \frac{\ln X_N - \ln X_{N-2}}{\ln X_N - \ln X_1} \quad (4)$$

for lognormal data sets.

Decisions concerning acceptance or rejection of the assumption H_o, that the two highest values are consistent with the rest of the data set, are made according to the procedure described in Section 3.2.2.1 using Table A-2 (Appendix A). Barnett and Lewis (1978) described related procedures appropriate for testing three or more outliers.

The Dixon Ratio Test and similar statistical procedures for identifying outliers in a data set are not recommended for repeated use on the same data set. The error risks specified in the tables were theoretically derived on the assumption that no extreme values have been removed from the data prior to the test. In practice, however, the tests are often applied successively; caution must be exercised in doing this, and the user must never discard data until a just cause has been determined. The user might also limit the rejected data to a small percentage of the data set, say 5 to 10%.

In all applications, the treatment of the outliers should be properly documented in order that a subsequent user of the results will know the impact of the outliers. For example, data analyses can be given with and without the discarded data.

3.2.3 Grubbs Test

Like the Dixon Ratio Test, the Grubbs Test (Grubbs and Beck 1972) can be used to determine whether the largest observation X_N in a sample from a normal distribution is too large with respect to the internal consistency of the data set. The Grubbs Test differs from the Dixon Ratio Test in that its test statistic,

$$T = \frac{X_N - \bar{X}}{S_x} \quad (5)$$

is calculated using all of the values in the data set. In addition to X_N, the data analyst must determine the arithmetic mean ($\bar{X}$),

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad (6)$$

and the standard deviation (S_x),

$$S_X = [\Sigma(X_i - \bar{X})^2/(N-1)]^{1/2} \quad (7)$$

The following equations can be used for positively skewed data that approximate the lognormal distribution:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} \ln Y_i \quad (8)$$

$$S_Y = [\Sigma(\ln Y_i - \bar{Y})^2/(N-1)]^{1/2} \quad (9)$$

Table A-3 (Appendix A) gives upper probability levels for T, where $3 \leq N \leq 50$.

The water quality data used in the illustration of the Dixon Ratio Test will be used again in the following example of the Grubbs Test in order to compare the results derived from using both tests.

* * *

EXAMPLE:

The T statistic relating to the largest observation for the data set 37; 51; 60; 68; 70; 71; 71; 72; 73; 77; 81; 82; 82; 82; 85; 92 is

$$T = \frac{92 - 72.13}{13.68} = \frac{19.87}{13.68} = 1.452$$

Examination of row 16 (N = 16) in Table A-3 indicates that the largest value of the data set is not an outlier at significance levels ranging from 0.1 to 10%. The Dixon Ratio Test provided the same conclusion at all of the recorded significance levels in Table A-2.

* * *

3.2.4 The Gap Test

The Gap Test (Curran *et al.* 1977; U.S. Environmental Protection Agency 1978) identifies spurious outliers by examining the frequency distribution for large gaps. The length of the gap between the largest X_N and the next largest X_{N-1} is $X_N - X_{N-1}$, between the second and third largest values is $X_{N-1} - X_{N-2}$, and similarly for other gaps. The occurrence of a gap length larger than a predetermined critical value indicates a possible data anomaly.

The test assumes that the upper percentiles of the frequency distribution can be closely fit by a two-parameter exponential curve defined as:

$$F(x) = 1 - \exp[-b(x-l)] \quad (10)$$

where: F(x) is the fraction of the total observations less than or equal to x;
b is the slope parameter;
and l is the location parameter.

Only the b value of the fitted curve is used in the Gap Test. Values of b can be estimated from the expression:

$$b = \frac{\ln[1-F(x_i)] - \ln[1-F(x_2)]}{x_2 - x_1} \qquad (11)$$

where $F(x_1)$ and $F(x_2)$ are the quantile values corresponding to the concentration values x_1 and x_2. Equation 11 is derived from Equation 10 for the same location parameter by taking logarithms of the $F(x_1)$ and $F(x_2)$ functions in Equation 10 and then subtracting one result from the other.

The probability of a gap of at least k units occurring in a data set free from erroneous values is given as:

$$p[\text{gap} \geq k] = \exp(-bk) \qquad (12)$$

A small p-value indicates that the gap size is in question and that the corresponding data value(s) should be flagged.

The following example demonstrates the use of Equation 12.

* * *

EXAMPLE:

A frequency distribution for a sample of data of specific conductances follows the pattern shown in Table 1; in addition, there is a suspect value of 114 μS/cm, which is at a distance of 30 μS/cm from its nearest neighbour ("gap" = 30 units). Determine if the value of 114 μS/cm should be flagged.

Substituting the respective values of 69.5 and 79.5 μS/cm for the 80th and 95th percentiles into Equation 11 yields

$$b = \frac{\ln[1-0.80] - \ln[1-0.95]}{79.5-69.5} = \frac{\ln 0.2 - \ln 0.05}{10} = 0.1386$$

The "gap" is 30 units in length; consequently, $p[\text{gap} \geq 30] = \exp[-(0.1386)(30)] = 0.016$.

There is less than a 2% probability that a gap of this length or larger would occur in the data set under the assumption stated. The 114 μS/cm value should be flagged for further investigation.

Table 1. Frequency Distribution of a Sample of Specific Conductance Data

Interval (μS/cm)	Frequency	Cumulative frequency
35 – 39	6	6
40 – 44	9	15
45 – 49	10	25
50 – 54	14	39
55 – 59	16	55
60 – 64	13	68
65 – 69	12	80
70 – 74	8	88
75 – 79	7	95
80 – 84	5	100

* * *

3.2.5 Regression Analysis

Bivariate and multivariate analysis regression techniques can be applied successfully to water quality data in order to check the internal consistency of a given data set and in some cases to check the level of correlation and functional relationship between the independent and dependent variables.

3.2.5.1 Simple Linear Regression Analysis

The simplest of all the regression methods is the simple linear regression technique. In applying the simple linear regression procedure to water quality data, a number of assumptions are made. Stated briefly, they are as follows:

(1) The relationship between the two variables under investigation is linear; that is, the relationship is of the functional form: $Y = mX + c$.

(2) The departures (errors or residuals) of the sample observations from the regression line are independent (demonstrated by evenly scattered residual points around the zero residual line).

(3) The variance of the observed Y values around the regression line is constant over the range of X values.

(4) The values of X are assumed to be measured without error.

(5) The variation of the observed Y values about the regression line should follow a normal distribution for the application of the F-test (for significance of the regression line) and the t-test (for computation of the confidence limits about the regression line).

To perform a simple test for data consistency and the existence of outliers, the method of least squares is used to establish the regression line, which has the form:

$$Y_c = a + bX \tag{13}$$

where: a is the value of the Y intercept when X = 0;
b is the coefficient establishing the slope of the line; and
Y_c is the calculated or estimated value of Y from the regression line.

The objective of the least squares method is to establish the values for the coefficients a and b which will yield a line for which the sum of the squared deviations from the observed Y values to the straight line is the least possible.

By the use of two normal equations (Equations 14 and 15), which are solved simultaneously, b and a can be written in functional terms of X and Y:

$$\Sigma Y = an + b\Sigma X \tag{14}$$

$$\Sigma XY = a\Sigma X + \Sigma X^2 \tag{15}$$

$$b = \frac{\Sigma(XY) - n(\bar{X})(\bar{Y})}{\Sigma X^2 - n(\bar{X})^2} \tag{16}$$

$$a = \bar{Y} - b\bar{X} \tag{17}$$

The values of a and b can now be used in Equation 13 to predict a Y value corresponding to any X value. This process assists the investigator in pinpointing potential outliers.

Data points presumed to be outliers should be included in the simple regression analysis, unless there is unambiguous substantiation that the data in question are anomalous.

There are several sources that can give rise to the creation of outlier values. Some of the responsible sources are:

- faulty instrument or system components;
- inaccurate instrument readings;
- errors in transcribing data;

- calculation errors; and

- abnormal local environmental conditions during the time of sample collection or field measurements.

There are a number of approaches available for determining if the suspected data are outliers. Some of these approaches have already been discussed in Sections 3.2.2 to 3.2.4.

The examination of auxiliary or support data can help in ascertaining some of the possible causes for the occurrence of the anomalous or skewed data. When there is complete surety that the questionable data are indeed anomalous, the erring data can then be omitted from the regression analysis.

The significance of the simple linear regression line can be tested after the regression line is fitted to the data. The ANOVA technique can be used as one of the tests in relating the "explained" and "unexplained" regression errors to the "total error" (Yamane 1967).

$$Y - \bar{Y} = Y_c - \bar{Y} + Y - Y_c \qquad (18)$$

(total error) (explained error) (unexplained error)

However, a simpler test is provided by using the test statistic t to determine the variation of the slope of the line from the value of 1.0. The t-statistic for testing the slope parameter is given as:

$$t = \frac{b - b_0}{S_E/\sqrt{S_{xx}}} \qquad (19)$$

where: b is the calculated slope from the regression line;

b_0 is equal to 1.0 ($H_0 : b_0 = 1$ for a 1:1 correspondence between the two variables);

$$S_E = \left[\frac{S_{YY} - S_{XY}^2/S_{XX}}{n-2}\right]^{1/2} \qquad (20)$$

$$S_{XX} = \Sigma X^2 - n\bar{X}^2 \qquad (21)$$

$$S_{YY} = \Sigma Y^2 - n\bar{Y}^2 \qquad (22)$$

$$S_{XY} = \Sigma(X - \bar{X})(Y - \bar{Y}) \qquad (23)$$

The H_0 hypothesis is rejected if $|t| \geq t_{\alpha/2,n-2}$; i.e., if $t \geq t_{\alpha/2,n-2}$ or $t \geq -t_{\alpha/2,n-2}$. The number of degrees of freedom (DF = n-k for n observations [values] and k estimates) is equal to n-2 because $\bar{X}$ and $\bar{Y}$ have been estimated.

A similar test exists for the intercept parameter a, where the t-statistic is formulated as:

$$t = \frac{a - a_0}{S_E\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}}} \qquad (24)$$

where n is the number of data points, with $a_0 = 0$ for the null hypothesis. Note that a_0 does not have to be zero, and other values may be used in formulating the null hypothesis, if this is reasonable for the experiment.

The intercept test is not as critical as the slope test, and in most instances the slope test will suffice.

The correlation coefficient (γ) also provides a useful measure of the degree of association between the X and Y variables and can vary between -1.0 and 1.0. (γ) is computed from Equation 25:

$$\gamma = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}} \qquad (25)$$

where S_{XY}, S_{XX}, and S_{YY} have the same meanings as in Equations 21, 22, and 23.

The following example illustrates the application of simple linear regression analysis to water quality data.

* * *

EXAMPLE:

Suspended sediment samples were simultaneously collected by two different water quality sampling devices (D-77 and WQS) at an aquatic site in Alberta, Canada over a four-month period. From the tabulated results shown in Table 2, determine or comment on:

- the simple linear regression line that can be fitted to the data;
- the significance of the regression slope at $p = 0.05$;
- the 95% confidence limit around the regression slope;
- the correlation coefficient for the comparison between the two data sets;
- the scatter diagram linking the two data sets; and

Table 2. Data for Suspended Sediment Concentrations (mg/L) in Samples Taken by Two Sampling Devices at an Aquatic Site in Alberta, Canada*

No.	Sediment concentrations associated with D-77 sampler	Sediment concentrations associated with WQB sampler
1	278	265
2	269	278
3	672	641
4	704	701
5	265	165
6	249	192
7	102	88
8	106	106
9	3924	4136
10	3952	3973
11	550	558
12	594	555
13	2548	2504
14	2535	2434
15	157	192
16	174	138

* Each pair of data represents a different sampling date during the four-month survey period.

- the claim (H_0 hypothesis) that the performance of the two sampling devices was similar in the collection of suspended sediment water samples.

Table 3 gives the values for the slope (b), the intercept (a), the correlation coefficient (γ), the t-test statistic for the slope, the t-test statistic for the comparison of the two data sets (arising from the two samplers), and the expression Y_c for the regression line. These values and expressions have been determined from the use of Equations 16 to 25 and Equation 26:

$$t_d = \frac{\bar{d} - d_0}{S_d/\sqrt{n}} \qquad (26)$$

where: $\bar{d}$ is the mean of the differences for the paired data;
d_0 is the value for the H_0 hypothesis (in many cases d_0 is assigned the value of zero);
S_d is the sample standard deviation; and
n is the number of paired data values.

Table 3. Statistical Parameters Related to the Regression Line for Suspended Sediment Concentrations in Samples Collected by Two Water Quality Samplers (A&B)

Parameter/expression	Value	Comments
b	0.98	Indicates an excellent linear relationship between the two data sets
a	35.09	
$Y_c = a + bX$	$S_A = 0.98S_B + 35.09$	
r	0.9989	Infers a very good association between the two samplers' related data.
$t=\frac{\bar{d} - d_0}{S_d/n}$	0.5409 (DF = 15)	The calculated t value at p = 0.05 is smaller than the t-value from the t-tables. H_0 is accepted—the two data sets are equivalent, indicating that the two samplers were similar in performance.
$t_{slope} = \frac{b - b_0}{S_{XY}/\sqrt{S_{XX}}}$	$t_{slope} = -0.0267$ $0.0267 << 1.771$ (DF = 14)	The calculated t value at p = 0.05 is smaller than the t-value from the t-tables, and there is therefore a significant relationship (approximately 1:1) between the two variables.

The plotted scatter diagram in Figure 2 does not reveal any data point that can be regarded as a significant outlier.

* * *

3.2.5.2 Multiple Regression

The assumptions underlying the methods of fitting a multiple regression (surface) to a given set of data are the same as those for a linear regression.

In some situations, the dependent variable (Y) is related to more than one independent variable (X_1, X_2, X_3, ..., X_n).

As was previously shown, a simple linear regression can be fitted to the data using only one independent variable. However, the amount of variation explained by a regression would probably be much better if all the independent variables were used together. The statistical method that enables this procedure to be carried out is called multiple regression.

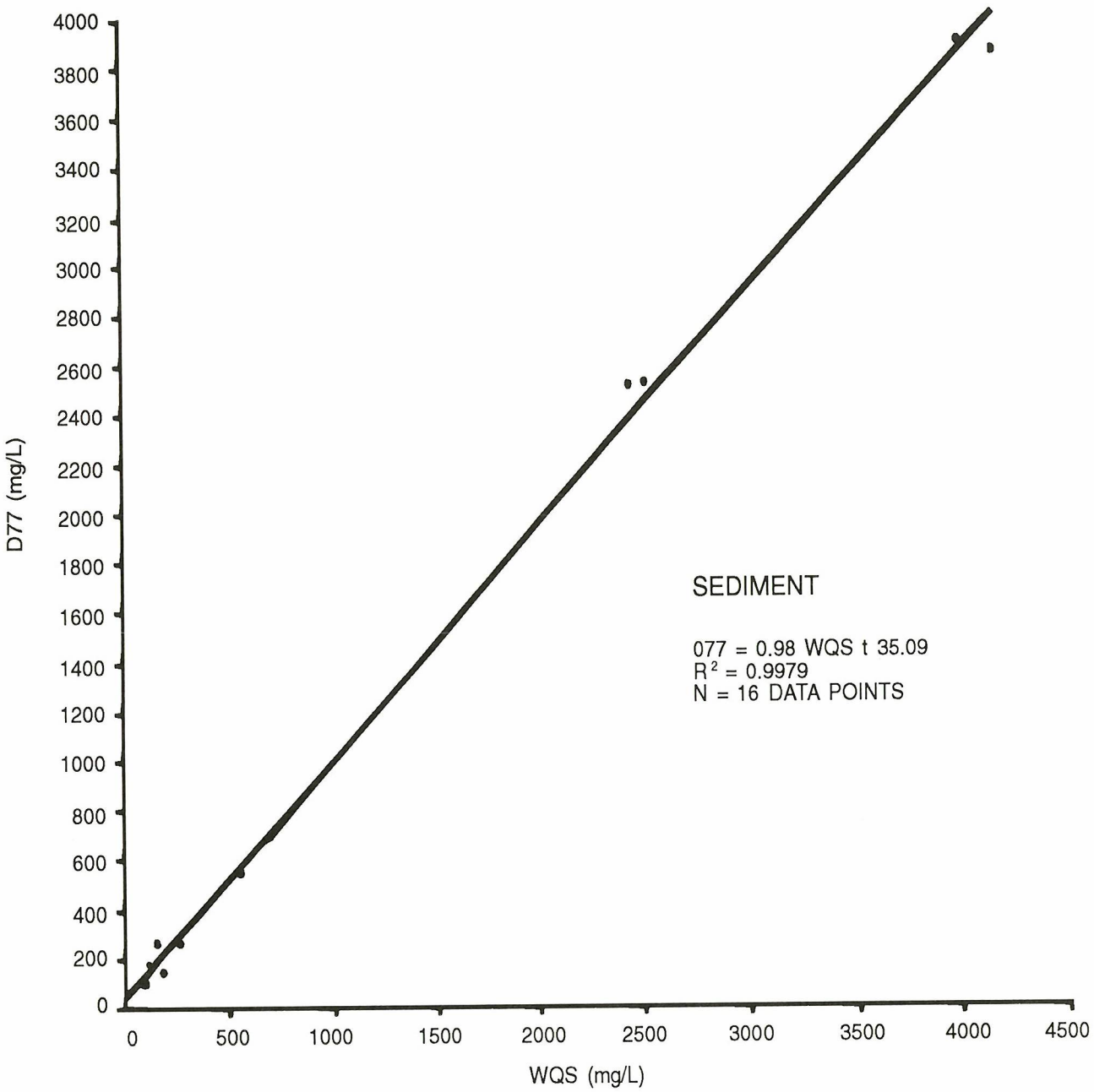


Figure 2. Linear relationship between D-77 and WQS values for total sediment concentrations.

As was the case with simple linear regression, the best regression line can be determined using the method of least squares. Again, a series of normal equations is established and the equations solved simultaneously. For the general linear model with a constant term as shown in Equation 27, where

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n \tag{27}$$

it is easy to develop the normal equations once a general pattern is recognized. Equation 28 shows the general model for the ith row:

$$(\Sigma X_1X_i)b_1 + (\Sigma X_2X_i)b_2 + (\Sigma X_3X_i)b_3 + \ldots + (\Sigma X_i^2)_{b_i} + \ldots$$
$$(\Sigma X_iX_n)b_n = \Sigma X_iY_i \tag{28}$$

The full development of the multiple regression model as it applies to the internal consistency of water quality data is beyond the scope of this text. For a detailed account of the mechanics of the least squares determination of the multiple regression line, the work by Freese (1967) should be consulted. The text gives worked examples that are applicable to water quality data; in addition, it discusses methods that are used in testing the significance of the multiple regression line and provides procedures for tests on the partial regression and partial correlation coefficients.

3.2.5.3 Curvilinear Regression

Only a very brief outline of curvilinear regression is given here. Generally, many of the curvilinear relationships can be handled using the regression methods already discussed.

Consider the simple power function:

$$Y = aX^b \tag{29}$$

Water quality data collected from streams will regularly follow this relationship. This function can be made linear using a logarithm transformation as in Equation 30:

$$\log Y = \log a + b \log X \tag{30}$$

Many other curvilinear relationships can be linearized (Dowdy and Wearden 1983). Some curvilinear functions, such as

$$Y = a + b^X \tag{31}$$

or

$$Y = a (X - b)^2 \tag{32}$$

cannot be fitted by the methods already discussed. The procedures required to regress these functions can be found in a number of texts (e.g., Furnival 1971; Draper and Smith 1981).

Whatever the form of the regression model, the computations and calculations can be simplified with the use of well-known statistical packages such as RS/1, SPSS, and SAS. The RS/1 package is available on the VAX system, and the SPSS and SAS packages can be accessed through the respective RS/1-SPSS and RS/1-SAS interface procedures (RS/1 User's Guide 1984).

3.2.6 Control Charting

The development and application of control charts to experimentally derived data provide an appropriate technique for monitoring a measurement process. The most common type of control charts involves the scheduled repetitive measurement (precision statistics) of a typical stable test sample and plotting of the results. Control limits are established within which the results are expected to lie as a result of random fluctuations of the measurement process. As long as the random fluctuations lie within the appropriate established control limits, the system is said to be in a state of statistical control.

3.2.6.1 Application of Control Charts to Data Validation

Control charts must be maintained in a real-time mode to the greatest extent possible if they are to be most effective in data validation. Adherence to a real-time mode agenda will allow responsive corrective actions to be taken as soon as problems are detected and will also provide the possibility for the minimization of anomalous data arising from out-of-control operations.

In the data validation program, the control chart changes from being a systematic historical record to being an active monitor of the measurement process; at the same time, it provides an opportunity for data checking and feedback process control.

To interpret control charts in relation to any specific data base, the following are among a number of rules used to indicate if the measurement process is out of control:

- one or more points outside three standard deviations;
- two or more consecutive points outside two standard deviations;
- a run of four points outside one standard deviation; and
- a run of seven or more points (this may be either seven consecutive points above or below the mean or seven consecutive points increasing or decreasing).

Westgard *et al.* (1977) suggested that it is better to use a combination of rules, such as the four above, than a single rule. These authors also discussed the performance characteristics of such rules.

There are many factors that could lead to problems in an analysis that would show up on the control chart. The basic symptoms seen on control charts are either a shift or trend. Some of the more common causes of these symptoms are as follows (Garfield 1984):

Symptom	Common causes of symptom
Shift in mean	Incorrect preparation of standard. Incorrect preparation of reagents. Contamination of sample. Incorrect instrument calibration. Analyst error.
Trend of mean upward	Deterioration of standard. Deterioration of reagents.
Trend of mean downward	Increased concentration of standard because of evaporation of solvent. Deterioration of reagents.
Increase in variability	Analyst performance, e.g., poor technique, lack of training, deviation from procedure.

Overall, the control chart not only provides a basis for action by the analyst, but also provides simple techniques and criteria that have been found useful in analyzing and interpreting data.

3.2.7 Sample Consistency Checks

Sample consistency checks should be undertaken to determine the validity of a given sample by investigating the relationships between the individual chemical species in the sample. These internal sample checks make use of the relationships between measured and calculated parameters associated with solution chemistry. However, in all of these types of tests, the results become difficult to interpret when any of the major chemical constituents are not reported.

A number of sample consistency checks that are used in the data validation process are described below.

3.2.7.1 Ion Balances

In any given sample, the theoretical sum of the anions must equal that of the cations, when both are similarly expressed in milliequivalents per litre. In practice, however, the sums are seldom equal because of variations in analysis. This inequality increases as the ionic concentration increases. Another possible source of error in the ionic balance equation may arise because only the traditional major ions (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, F^-, SO_4^{2-}, NO_3^-, NO_2^-, CO_3^{2-}, HCO_3^-) and SiO_2 are considered,

whereas organic ions and ions such as Fe^{3+}, Fe^{2+}, Cu^{2+}, and PO_4^{3-}, which can be ubiquitous, are omitted from the balance calculations. The size of this error is hard to predict, but it might be of importance in some cases, depending on the source(s) from which the water samples are taken.

Two approaches are currently favoured for the investigation of ionic balances: a Control Chart Method (APHA/AWWA/WPCF 1985; Gaskin 1985a), and the Percent Error Method (Campbell and Scott 1985).

In the Control Chart Method, the absolute difference between the sum of the cations and the sum of the anions (or the ratio of the two sums) is plotted against the sum of the anions; lines indicating ± 1 standard deviation are established as acceptable limits, whereas lines indicating ± 2 standard deviations are set as action limits.

The mechanics of setting up the control chart from the regression line and the standard error of estimate (SEE) are laid out in the paper by Gaskin (1985a).

Sometimes an accidental combination of erroneous analyses resulting in the balancing of errors (compensating errors) may produce agreement between the sum of the anions and the sum of the cations even though two or more individual analytical results are seriously incorrect.

The Percent Error Method involves an integrated chemoelectrical view of cations and anions leading to a commonly used equation for evaluating the overall accuracy of cation–anion analyses. One form of the electroneutrality equation is:

$$\text{Error} = \left|\frac{\Sigma A - \Sigma C}{\Sigma A + \Sigma C}\right| \times 100 \qquad (33)$$

where: ΣA is the sum of all anions in milliequivalents per litre;
ΣC is the sum of all cations in milliequivalents per litre; and
$|\ |$ represents the absolute value of the expression.

Milliequivalents are obtained by multiplying the number of millimoles per litre of each ion by the absolute value of each species' ionic charge. The "error" in Equation 33 implies the extent to which the water sample is theoretically electroneutral or "charge balanced." Hydrogen ions (H^+) are included in the cationic sum in acid precipitation and wet deposition studies.

Peden *et al.* (1979) suggested that samples having an error greater than about 10% should be reanalyzed. If the analytical results for some ions (e.g., PO_4^{3-}) have not been considered or if large amounts of weak acids are present, large errors should be anticipated.

3.2.7.2 Anion/Cation Ratio

The anion/cation ratio utilizes the fact that the net charge on an acid precipitation sample must be zero, similar to the Percent Error Method. The ratio of anions to cations should be 1.0. The ratio

is calculated by comparing the sum of anion equivalents (SO_4^{2-}, NO_3^-, NO_2^-, Cl^-) with the sum of cation equivalents (H^+, NH_4^+, K^+, Ca^{2+}, Mg^{2+}, Na^+).

If the ratio is greater than 1.0, a cation deficiency is indicated, whereas a ratio less than 1.0 indicates an anion deficiency. In both cases, an error in analysis, sample contamination, or a missing major constituent is indicated. A correction of the anion equivalence for the contribution of bicarbonate may be applied (Stensland 1978) to account for the effects of atmospheric CO_2 on the precipitation sample.

The anion/cation ratio should be calculated and reported by the laboratory as part of its QC procedures. It has been recommended that the ratio be retained in the data base and reported in all data reports (Vet and Onlock 1983).

Typically, it is found that the measured cation equivalence is greater than the anion equivalence, and the flagging limits should be skewed to reflect this. To accommodate the imbalance, some investigators (Kirk 1982) use a range of -30 to +10%, and others (Topol *et al.* 1981) use a range of -50 to +20% for the flagging limits. The latter investigators omit the carbonate correction.

3.2.7.3 Measured Versus Theoretical Conductance

A comparison of the measured conductance with that predicted by the concentrations of the major ionic species can be used to investigate the sample composition. The theoretical conductance may be computed from the molar concentrations and theoretical conductivities for the individual ionic species. The comparison between measured and theoretical conductances can be particularly sensitive to errors when hydrogen ions (H^+) are involved. This is due to the relatively high conductance of H^+ (350 μS/cm·L).

3.2.7.4 Ion Exchange Method

The ion exchange method is based on the replacement of cations in the original sample with hydrogen ions supplied by a strongly acidic cation exchange resin. Following the ion exchange, the cationic strength (meq/L) can be determined by titration.

The total cation concentration can also be found by the summation of the determined cationic constituents (e.g., by using atomic absorption spectrophotometry, inductively coupled argon plasma determinations). The total alkalinity of the original water sample must also be ascertained in order to complete the calculation.

Any large difference between the two cationic concentrations (which have been independently determined) may indicate the occurrence of a gross error (or errors) in the analysis of the water sample.

3.2.7.5 pH Checks

A theoretical pH value may be determined by assuming that all the hydrogen atoms originate from the major acidic species (H_2SO_4, HNO_3, and HCl) and by using the measured SO_4^{2-}, NO_3^-, and Cl^- concentrations. This comparison can prove particularly useful in identifying errors in cases where the acidity of the sample is being influenced by moieties such as H^+, SO_4^{2-}, NO_3^-, or Cl^-.

3.2.7.6 Ratios of Specific Parameters

Ratios of specific parameters may also be considered as a sample consistency check to investigate specific types of sample contamination. These checks may prove to be particularly useful for any site at which a specific type of sample contamination may be expected (e.g., acid rain precipitation [SO_2, NO_x], acid effluent from a metal mining industry). The assignment of typical or acceptable values for this type of comparison would, however, demand a detailed special study to determine the composition of the suspected contaminants at the selected site(s).

3.2.7.7 Mass Balances

When appropriately representative water samples have been collected and the corresponding laboratory analyses have been accurately carried out, the determination of mass balances on some water quality parameters can sometimes serve as a useful validation tool. As an example, the mass balance between total nitrogen (TN), total dissolved nitrogen (TDN), and particulate nitrogen (PN) can be investigated from the formula TN = TDN + PN, where all three species are measured independently.

At present, the partial validation of a data set through the wide application of the mass balance process is severely limited, either because of the number of chemical species that need to be accurately and independently measured in order to give a satisfactory mass balance, or because of the semiqualitative basis on which a number of the involved component species are determined. Two cases in which these difficulties can be experienced are reflected in the following examples:

(a)	TP (total phosphorus)	=	TDP (total dissolved phosphorus)	+	PP (particulate phosphorus)
(b)	TC (total carbon)	=	TDC (total dissolved carbon)	+	PC (particulate carbon)

Caution should therefore be exercised in the use of mass balancing as an aspect of data validation, especially when there is doubt concerning the number of species making up the mass

balance equation and when the accuracy involved in the measurements of those species is questionable.

3.3 Tests for Consistency of Parallel Data Sets

The tests for internal consistency described in Section 3.2 implicitly assume that most of the values in a data set are correct. Consequently, if all of the values in a data set incorporate a small positive bias, tests such as the Dixon Ratio Test would not identify the data set as being inconsistent. One method of identifying a systematic bias of this type is to compare the data set with other data sets that presumably have been sampled from the same population (i.e., same aquatic site and same time period) and to check for differences in the average value or the overall distribution of values.

This section describes three procedures that are recommended for comparisons involving two parallel data sets. These procedures are presented in order of increasing sensitivity to differences between the data sets and increasing computational complexity. These tests are nonparametric: i.e., they do not assume that the data have a particular distribution. Therefore, these tests can be used for those nonnormal data sets that may occur in water quality analysis.

3.3.1 The Sign Test

The Sign Test is a fairly simple procedure for testing the hypothesis, H_0, that two related paired samples, such as data sets originating from similar sampling devices, similar methodologies, or "before" and "after" effects with the same instrument or methodology (all other things being equal), have the same median value. The sign (+ or -) of the algebraic difference between each corresponding pair of data points is determined, and the total number of positive signs (N_+) and the total number of negative signs (N_-) are counted; differences of zero are ignored. The probability that both samples have the same median is given as:

$$p[H_0 \text{ is true}] = 2N!(1/2)^N \sum_{j=0}^{n} \frac{1}{j!(N-j)!} \qquad (34)$$

where: n is the smaller of the two numbers n_+ and n_-; and
$N = n_+ + n_-$.

For large N (say > 30), one can use the normal approximation to the above probability by calculating:

$$Z = \frac{2n - N}{N} \qquad (35)$$

The Sign Test can also be applied in determining the validity of the null hypothesis $\mu_1 = \mu_2$ for the interdependent species 1 and 2. Again, the individual differences in the corresponding experimental values form the basis for the nonparametric statistical analysis. The total number (N) of signs (both + and -) and the respective proportions of positive and negative signs are determined. For $N \geq 30$, the test statistic (Siegel 1956) can be written as:

$$Z = \frac{p - \rho}{\sqrt{\frac{\rho(1-\rho)}{N}}}$$

where: p is the proportion of either positive or negative signs; and
ρ is the population proportion ($\rho = 0.50$).

The following example will demonstrate the use of Equations 35 and 36.

* * *

EXAMPLE:

In a comparison of "Total Metals" analytical results using two digestion techniques (Gaskin 1985b), the respective numbers of positive and negative differences were 23 and 32, and the total number of the two signs (+ and -) was 55. Would one digestion technique instead of two have been sufficient so that resources could be saved? In other words, could it be proven that both digestion techniques gave equivalent analytical results?

The H_0 assumption is $C_L = C_R$ at $p = 0.05$, where C_L is the mean of the analytical values associated with the National Water Quality Laboratory digestion method, and C_R is the mean analytical value associated with the Western and Northern Region's digestion method.

Using Equation 35 for the smaller number of positive signs,

$$Z_1 = \frac{(2)(23) - 55}{55} = -1.214$$

This corresponds to $p_1 = 0.113 = \alpha/2$, i.e., $\alpha = 0.226$.

Using Equation 36 for the proportion of positive signs (23/55 = 0.43),

$$Z_2 = \frac{0.42 - 0.50}{\sqrt{\frac{(0.5)(0.5)}{55}}} = \frac{-0.08(7.416)}{(0.50)} = -1.187 \qquad (36)$$

This corresponds to $p_2 = 0.117 = \alpha/2$, or $\alpha = 0.234$.

The two results Z_1 and Z_2 are very close for the two approximations, and the H_0 assumption is correct, because p_1 and p_2 not only are close but are both greater than 0.05.

Note that, statistically, the data analyst will not reject H_0 if p[H_0 is true] > 0.05.

* * *

Unfortunately, the Sign Test is a gross measure of the significance of difference. It does not take into account the magnitude of change in scores (values) under different conditions. For this problem, the Wilcoxon Signed-Rank Test may be used.

3.3.2 Wilcoxon Signed-Rank Test

Like the Sign Test, the Wilcoxon Signed-Rank Test can be used to test the assumption (H_0) that two samples come from populations having the same medians. The Wilcoxon Signed-Rank Test is generally more powerful than the Sign Test, as it considers both the sign and the magnitude of the difference between paired data. The steps in the procedure are as follows:

(1) Determine the sign and magnitude of the algebraic difference between each of the pairs of data points; assume that N scores remain after throwing out all "zeros."

(2) Assign ranks to the "absolute values" of these N differences; use the average rank in case of ties. The ranks must increase as the absolute values increase.

(3) Assign to each rank the sign that it represents.

(4) Determine the sum (T_1) of the ranks associated with a minus sign and the sum (T_2) of the ranks associated with a plus sign.

(5) To test H_0, either use published tables (such as Table A-4) based on T_2 and N or use a large sample normal approximation with

$$Z = \frac{T_1 - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \qquad (37)$$

The probability that H_0 is true—i.e., p[H_0 is true]—is equal to twice the area between Z and Z_0 (Z = 0) under the standard normal curve at a significance level of $\alpha/2$.

The following example illustrates the application of the Wilcoxon Signed-Rank Test to water quality data.

* * *

EXAMPLE:

Water quality data were gathered at the Red River at Bindloss, Alberta, Canada in 1980; the data relate to nutrients and turbidity parameters. Two different sampling devices (D-77 and WQS) were used in the sample collection process. Sites and sampling time periods (dates) were the same for each matched pair of data. Table 4 shows the results and the arrangements for the Wilcoxon Signed-Rank Test.

Table 4. Application of the Wilcoxon Signed-Rank Test to Water Quality Data for Samples Collected by Two Sampling Devices at the Red Deer River at Bindloss, Alberta, Canada (1980)

Parameter	No.	Sampler D-77	Sampler WQS	Absolute difference	(Signed) Rank
Turbidity	1	26.6	26.6	0.0	
	2	26.3	26.4	0.1	-4.5
	3	27.3	27.4	0.1	-4.5
	4	26.8	26.9	0.1	-4.5
	5	26.8	26.8	0.0	
	6	27.0	27.0	0.0	
	7	27.3	27.4	0.1	-4.5
	8	27.1	27.1	0.0	
Total phosphorus	1	1.20	1.20	0.0	
	2	1.21	1.21	0.0	
	3	1.25	1.24	0.1	+4.5
	4	1.18	1.18	0.0	
	5	1.13	1.13	0.0	
	6	1.50	1.51	0.1	-4.5
	7	1.51	1.52	0.1	-4.5
	8	1.62	1.63	0.1	-4.5
Total nitrogen	1	0.62	0.65	0.03	-10.5
	2	0.54	0.57	0.03	-10.5
	3	0.55	0.58	0.03	-10.5
	4	0.55	0.58	0.03	-10.5
	5	0.58	0.63	0.05	-13

The sum (T_1) of the ranks associated with the minus sign is 86.5, and the sum (T_2) of the ranks associated with the plus sign is 4.5. Table A-4 (Appendix A) indicates that when N = 21, the probability of H_o being true would be less than 10% if $T_2 \leq 67$ or $T_2 \geq 164$. Because $4.5 < 67$, then the probability that the D-77 data are equivalent to the WQS data can be accepted with at least a 90% degree of confidence (α = 10%).

The Z-value for T_1 is found from Equation 37:

$$Z = \frac{\frac{86.5 - (21)(22)}{4}}{\sqrt{\frac{(21)(22)(43)}{24}}} = -1.008$$

$Z = -1.008$ corresponds to a value of $\alpha = 0.316$, which is much greater than 0.05 or 0.10. The hypothesis that the two data sets associated with the two samplers are equal is therefore acceptable.

* * *

The Wilcoxon Signed-Rank Test has a definite advantage over the Sign Test. Not only is the direction of the difference in scores (values) taken into account, but the magnitude of the change is also observed. This test is very popular, in part because of this characteristic. A table exists that makes the interpretation of test results easy, and the test is also easy to apply and to understand.

When sample sizes exceed 25, a modification of the Z test (Equation 37) can be used to identify significant changes in scores. This characteristic also enhances the popularity of the test.

3.3.3 The Rank Sum Test

The Rank Sum procedure is useful when testing the hypothesis (H_0) that two samples represent populations with the same distribution. Unlike the Sign Test and the Signed-Rank Test, the Rank Sum Test is applicable to independent (unrelated) samples. The steps involved in the Rank Sum procedure, and which are appropriate for tests of 10 or more data pairs, are as follows:

(1) Combine the n_1 observations from population 1 and the n_2 observations from population 2. Arrange the observations in ascending order of size, and then assign ranks from 1 to $(n_1 + n_2)$ where $n_1 \leq n_2$. In case of ties, the average rank is used.

(2) Calculate T_1, the sum of the ranks assigned to observations from population 1.

(3) Compare T_1 with the critical value in Table A-5, $\alpha = 0.10$ (Remington and Schork 1970). If $T_l \leq T_1 \leq T_\gamma$ for the values of T_l and T_γ listed for samples n_1 and n_2 in the table, then the $p[H_0$ is true$] \leq 0.10$. If T_1 is outside the indicated range, go to the $\alpha = 0.05$ table and repeat the test. If $T_l \leq T_1 \leq T_\gamma$ then $0.05 \leq p[H_0$ is true$] \leq 0.10$. If not, continue the search until the α value is found for which $T_l \leq T_1 \leq T_\gamma$. If $\alpha < 0.05$, there is good reason to reject H_0.

(4) To test H_0 for $n_1 \geq 20$, use the large-sample normal approximation as represented by Equation 38.

$$Z = \frac{T_1 - n_1(n_1+n_2+1)/2}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} \qquad (38)$$

The probability that H_0 is true—i.e., p[H_0 is true]—is equal to twice the area between Z and Z_0 (Z = 0) under the standard normal curve at a significance level of $\alpha/2$.

* * *

EXAMPLE:

Table 5 illustrates the application of the Rank Sum Test to two sets of total nitrogen (mg/L) data collected weeks apart by the same sampling device at the South Saskatchewan River at Highway 41, Alberta, Canada (1980). The tied values are assigned an average rank value.

The sum T_1 of the ranks from Data Set 1 is equal to 77. In Table A-5, for $\alpha = 0.10$, $T_l = 82$, and $T_\gamma = 128$ for $n_1 = n_2 = 10$. Consequently, $T_l \leq T_1 \leq T_\gamma$, and p[$H_0$ is true] ≥ 0.10. Therefore, the assumption that the two sets of total nitrogen data came from a population that did not change very much over a five-week period (from July 22 to August 28, 1980) should not be rejected.

* * *

Whereas the Sign and Wilcoxon Signed-Rank tests are particularly sensitive to differences between related observations, the Rank Sum Test is relatively insensitive to small data shifts, mainly because it compares sample distributions without actually pairing the data. The Rank Sum Test is more appropriate for identifying censored data.

3.4 Detection Limits

In concluding this section on screening and data validation procedures for water quality data, a case can be made for the strict observation of detection limits so that no data are reported below the detection capabilities of the analytical systems. This check should consist of a simple comparison of reported values for each parameter against the tabulated detection limit criterion. The program should also check for flags for those data reported below the detection limit. Ideally, this check should be performed before the data leave the laboratory.

3.5 References

APHA/AWWA/WPCF. 1985. Standard Methods for the Examination of Water and Wastewater. 16th ed. American Public Health Association/American Water Works Association/Water Pollution Control Federation,Washington, D.C.

Table 5. Application of the Rank Sum Test to Two Sets of Total Nitrogen Data Collected on Two Different Dates by the Same Sampling Device at the South Saskatchewan River, Highway 41, Alberta, Canada (1980)

Data Set 1	Data Set 2	Data values (in numerical order)	Data set	Rank
0.30	0.36	0.21	2	1.5
0.30	0.37	0.21	2	1.5
0.28	0.33	0.27	1	4
0.27	0.34	0.27	1	4
0.28	0.30	0.27	1	4
0.27	0.30	0.28	1	8
0.28	0.36	0.28	1	8
0.27	0.36	0.28	1	8
0.28	0.21	0.28	1	8
0.28	0.21	0.28	1	8
		0.30	1	12.5
		0.30	1	12.5
		0.30	2	12.5
		0.30	2	12.5
		0.33	2	15
		0.34	2	16
		0.36	2	18
		0.36	2	18
		0.36	2	18
		0.37	2	19

Barnett, V., and T. Lewis. 1978. Outliers in Statistical Data. John Wiley & Sons, New York, N.Y.

Campbell, S., and H. Scott. 1985. Quality assurance in acid precipitation measurements. Am. Soc. Test. Mater. Spec. Tech. Publ. 687: 272–283.

Curran, T.C., W.F. Hunt, Jr., and R.B. Faoro. 1977. Quality Control for Hourly Air Pollution Data. Presented at the 31st Annual Technical Conference of the American Society for Quality Control, May 16-18, Philadelphia, Pa.

Dowdy, S., and S. Wearden. 1983. Statistics for Research. John Wiley & Sons, Toronto, Ont.

Draper, N., and H. Smith. 1981. Applied Regression Analysis. 2nd ed. Wiley, New York, N.Y.

Freese, F. 1967. Elementary Statistical Methods for Foresters. U.S. For. Serv. Agric. Handb. 317, 87 pp.

Furnival, G.M. 1971. All possible regressions with less computation. Technometrics 13: 403–408.

Garfield, F.M. 1984. Quality Assurance Principles for Analytical Laboratories. Association of Official Analytical Chemists, Arlington, Va., pp. 98–108, 185–187.

Gaskin, J.E. 1985a. An Alternative Approach to the Question of Ionic Balances. Water Quality Branch, Inland Waters Directorate, Environment Canada, Regina, Sask.

Gaskin, J.E. 1985b. Comparison of Total Metals Analytical Results Using Two Digestion Techniques. Water Quality Branch, Inland Waters Directorate, Environment Canada, Regina, Sask.

Grubbs, F.E., and G. Beck. 1972. Extension of sample sizes and percentage points for significance test of outlying observations. Technometrics 14(4).

Kirk, R.W. 1982. Data Validation Procedures—Daily Precipitation Monitoring Network 1980/81. Acid Precipitation in Ontario Study, Ontario Ministry of Environment.

Nelson, C.A., Jr., D.W. Armentrout, and T.R. Johnson. 1980. Validation of Air Monitoring Data. EPA-600/4-80-030, June, U.S. Environmental Protection Agency.

Peden, M.E., L.M. Skowrom, and F.F. McGurk. 1979. Precipitation Sample Handling, Analysis, and Storage Procedures. Research Reprint 4, Contract EY-76-5-02-1199, prepared for U.S. Department of Energy, Washington, D.C.

Pollock, T.L. 1988. Personal communication. Water Quality Branch, Inland Waters Directorate, Environment Canada, Moncton, N.B.

Remington, R.D., and M.A. Schork. 1970. Statistics with Applications to the Biological and Health Sciences. Prentice-Hall, Englewood Cliffs, N.J.

Rhodes, R.C. 1977. What is data validation? *In* Data Validation Conference Proceedings. PB82-114364, Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, N.C.

RS/1 User's Guide. 1984. Book 2: Graphics & Statistics. BBN Software Products Corporation, Cambridge, Mass.

Siegel, S. 1956. Nonparametric Statistics for the Behavioural Sciences. McGraw-Hill, New York, N.Y.

Stensland, G.J. 1978. A Comparison of Precipitation Chemistry Data at a Central Illinois Site in 1954 and 1977. Paper 78-48.7, 71st Annual Meeting of the Air Pollution Control Association, June 25–29, Houston, Tex.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Chelsea, Mich.

Topol, L.E., P. Flanagan, P. Chen, M. Lev-on, R. Schwall, and L.S. Shepard. 1981. Quality Assurance Manual for Precipitation Measurement Systems. U.S. Environmental Protection Agency Contract No. 68-02-3262.

U.S. Department of Commerce. 1978. Computer Science and Technology: Performance Assurance and Data Integrity Practices. National Bureau of Standards, Washington, D.C.

U.S. Environmental Protection Agency. 1978. Screening Procedures for Ambient Air Quality Data. EPA-450/2-78-037, July.

Vet, R.J., and S.G. Onlock. 1983. The Canadian Air and Precipitation Monitoring Network Quality Assurance Plan for Precipitation Monitoring Systems. CSC 110. 194-3-1. Air Quality and Inter-Environmental Research Branch, Atmospheric Environment Service, Environment Canada, Downsview, Ont.

Westgard, J.O., T. Groth, T. Aronsson, and C.H. de Verdier. 1977. Clin. Chem. 23: 1857–1887.

Yamane, T. 1967. Statistics, An Introductory Analysis. 2nd ed. Harper & Row, New York, N.Y.

4.0 DATA VERIFICATION

Verification controls are required for data originating from sampling, field testing, and laboratory analysis. Data must not be released before checks are made for quality and correct documentation.

Data verification checks must be conducted by field and laboratory personnel before the data are sent out for storage and, ultimately, to data users. Any discrepancies or errors found must be corrected prior to storage, and the data must also be checked again on their merger into the existing data base during the storage process.

The merging of field data with laboratory data provides an additional QA check. Any mismatch will indicate a sample loss or data loss as a result of shipping or data transmittal inefficiencies.

All data at the laboratory and field operational levels should be verified by one or more of the following methods:

- simultaneous measurement of a standard reference material (SRM) or other certified reference material (CRM);
- measurement by an independent technique;
- measurement by an independent operator/apparatus;
- spiking/recovery technique; and
- replicate measurements by a recognized technique.

The laboratory data system must be designed to incorporate comment codes into the results reported for individual analyses. For example, codes must be available to indicate reasons for missing parameters (e.g., insufficient sample volume), results invalidated at the laboratory (e.g., calibration problem), missing samples, nonpreserved samples, and data reported at the detection limit of the analytical system. It is recommended that numerical codes (e.g., -100) be inserted into the individual parameter fields to represent the various comments that will be used. In particular, it is important that the numerical value of the detection limit be reported (with a flag) for samples at the detection limit rather than simply "less than the detection limit" in order to facilitate data manipulation routines. A common procedure is to flag the detection limit values with a minus sign.

When all the laboratory analyses (physical, chemical, biological, computational) have been completed and the laboratory data have been subjected to data validation procedures, further data verification checks should be made before the data leave the laboratory. These checks consist of:

- verifying that a result is reported for each sample;

- checking for transcription errors by reviewing all transcribed data;
- spot-checking the laboratory data printout against original field sheets;
- ensuring that all laboratory checks of field QA/QC are reported (e.g., shipping temperature, sample volumes, preservation); and
- ensuring that any missing or invalid results are explained (i.e., with comment codes).

The laboratory should be responsible for performing a number of QC checks on the field portion of the monitoring program. These checks consist of:

- verifying sample labelling and matching field sheets with samples;
- ensuring that samples were submitted as recorded in the sample logbook (or computer sample registry); monitoring sample shipping (time, temperature, mode of shipping), sample condition (leaks, contaminants), and sample volume (independent measurement of sample volume); and
- reporting any other comments that may be important on the sample submission form.

In the final verification step, the program manager or the project officer should be responsible for reviewing (and editing if necessary) all data (connected to a specific project or program) before those data are entered into the final data base for subsequent data reporting and analysis. This review should consist of an investigation of all data points that were flagged as a result of the gross sample checks, data screening, and data validation checks, as well as an overall evaluation of the data set.

It should be the responsibility of the individual(s) carrying out this evaluation to decide whether the suspect data points should be accepted or rejected. All suspect data points should be examined with the aid of field data sheets (comments), logbooks, and QA and QC information (field and laboratory).

In general, data should be rejected only in clear cases of nonrepresentative or contaminated samples. Comparison of the suspect data points with related historical data could aid in the acceptance or rejection decision process.

Any data that are rejected should be referenced in the data base through the use of an appropriate data flag (e.g., -1000.00) for individual parameters; the flag should indicate that the measurement was made but the result was not reliable.

Rejected data should be deleted from the data base, but a record of values should be kept on file in the event that the point(s) may be of some future interest. In addition, comments made concerning the reasons for rejecting data points should be documented and kept for future reference, in case the data users want to examine why the data were rejected.

Personnel involved in the verification (evaluation or validation) of data should also have the opportunity to enter a set of comment codes (related to the sample validity) into the data base to reflect the results of both the data validation and data verification processes. Table 6 shows a suggested format for listing data validity and verification comment codes.

Table 6. Suggested Format for Data Validity and Verification Comment Codes

Code	Comments
001	Data acceptable
002	One or more parameters rejected
003	Evidence of sample contamination
004	Potential sample contamination
005	Sample lost in shipment
006	Part of event missed
007	Composite sample
008	Split sample
009	Abnormal collection efficiency
010	Poor ionic balance/poor mass balance
011	Outliers identified and explained
012	One or more parameters exceed acceptable (allowable) limits — indicate which one with parameter flag

5.0 DATA QUALITY ASSESSMENT

One of the objectives of the WQB QA/QC program is to ensure that environmentally related data used by the WQB are of known and documented quality. A most important feature of data quality is the variability of the measurement system, including the operations of sample acquisition and sample analysis.

Another objective of the WQB QA/QC program is the establishment of uniform procedures for calculating and reporting data quality indicators (e.g., accuracy, precision, bias, method detection limit [MDL]).

Application of data quality assessment procedures to the measurements on any given sample should be based on the assumption that all measurements, including the QC measurements, are made with the measurement system in "statistical control." A measurement system is deemed to be in statistical control when its variability is due only to chance causes. Data quality assessment results used with control charts or statistical techniques such as the construction of frequency distribution plots may assure that the measurement system is in a state of statistical control.

Statistical measures of data quality should be used in conjunction with archived information on the operational capability of the respective measurement systems.

The principal indicators of data quality are accuracy, bias, precision, and MDL.

5.1 Precision

Precision is generally defined as the agreement among repeated observations made under the same conditions. The sample standard deviation S and the sample coefficient of variation (CV), where

$$CV = \frac{(S)}{\bar{x}} 100 \qquad (39)$$

are used as indices of precision. The smaller the values of S and CV, the better the precision. The use of S in the place of the CV becomes important when the mean ($\bar{x}$) of the observations is a very small quantity in the vicinity of zero. Such an instance can occur when $\bar{x}$ is the mean of a set of measurements made in the neighbourhood of the detection limit, and S is of the same order of $\bar{x}$ or larger.

In instances where precision estimates are obtained from analyses of replicate pairs from the same sample, the relative range RR is sometimes used as an index of precision:

$$RR = \frac{100R}{\bar{x}} \qquad (40)$$

where R is the maximum value minus the minimum value within the observations.

For replicate pairs, the relationship between the range and standard deviation is given as:

$$S = R/\sqrt{2} \tag{41}$$

Another index of precision is derived from individual differences (d_i) between duplicate measurements. The precision (standard deviation) as calculated from k differences is given as:

$$S = \sqrt{\frac{\Sigma(d_i)^2}{2k}} \tag{42}$$

When there is a combination of duplicate and triplicate results on different samples relating to the same parameter, the precision S can be determined from Equation 43:

$$S = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N - K}} \tag{43}$$

where: X_i is the individual value in each duplicate or triplicate set;
$\bar{X}$ is the average of results for the same duplicate or triplicate set;
N is the total number of repeats; and,
K is the sum of the number of duplicates and triplicates combined (i.e., the total number of repeats).

An example illustrates the uses of Equations 42 and 43 in determining standard deviations (precisions).

* * *

EXAMPLE:

Table 7 contains a set of dissolved chloride results obtained in duplicate and triplicate measurements on a number of water quality samples taken from the same population. Calculate the standard deviation (precision) using Equations 42 and 43.

For the duplicate determinations, $\Sigma d_i^2 = 0.24$ and $K = 8$. Therefore, S_1 is computed as:

$$S_1 = \sqrt{\frac{0.24}{(2)(8)}} = 0.1225$$

For the duplicate and triplicate determinations, $\Sigma(X_i-\bar{X})^2 = 0.22$, and $N-K = 25-11 = 14$. Therefore, S_2 is computed as:

$$S_2 = \sqrt{\frac{0.22}{14}} = 0.1253$$

Table 7. Duplicate and Triplicate Determinations for Dissolved Chloride in a Number of Water Quality Samples

Sample	Chloride concentration X_i (mg/L)	$\bar{X}$	$\lvert (X_i - \bar{X}) \rvert$	$(X_i - \bar{X}_2)$	$\lvert d_i \rvert$	d_i^2
A	2.1		0.1	0.01		
	2.1	2.0	0.1	0.01	–	–
	1.8		0.2	0.04		
B	2.2		0.1	0.01		
	2.0	2.1	0.1	0.01	0.2	0.04
C	1.9		0.1	0.01		
	2.1	2.0	0.1	0.01	0.2	0.04
D	2.0		0	0		
	2.0	2.0	0	0	0	0
E	1.8		0.1	0.01		
	2.0	1.9	0.1	0.01	0.2	0.04
F	2.0		0	0		
	2.1	2.0	0.1	0.01	-	-
	1.9		0.1	0.01		
G	2.2		0.1	0.01		
	2.0	2.1	0.1	0.01	0.2	0.04
H	1.9		0.1	0.01		
	2.1	2.0	0.1	0.01	0.2	0.04
I	2.3		0.1	0.01		
	2.2	2.2	0	0	-	-
	2.1		0.1	0.1		
J	2.1		0.1	0.01		
	2.3	2.2	0.1	0.01	0.2	0.04
K	2.1		0	0		
	2.1	2.1	0	0	0	0

The difference between S_1 and S_2 is small (2.2%), which suggests that the measurement system was in a state of statistical control.

* * *

5.1.1 Intralaboratory and Interlaboratory Precision

For analytical procedures, precision may be specified as either intralaboratory (within-laboratory) or interlaboratory (between-laboratory) precision. Intralaboratory precision estimates represent the agreement expected when the same laboratory (or operator) uses the same method to make repeated measurements on the same sample. Interlaboratory precision refers to the agreement expected when two or more laboratories (or operators) analyze the same or identical samples with the same or different methods. Intralaboratory precision is more commonly reported; however, where available, both intralaboratory and interlaboratory precision should be listed in the data compilation.

At present, there is no completely valid method available to directly calculate the inter-laboratory precision from the intralaboratory precision or vice versa. This is unfortunate, because frequently only one type of precision estimate is available for a method. However, in general, the following comments are applicable. First, the intralaboratory precision should be smaller than the interlaboratory precision because of additional variables in the latter. Secondly, if the interlaboratory precision is much larger than the intralaboratory precision, this indicates that the method is very technique-sensitive. Such information can be very important when considering a project that might involve several laboratories analyzing the same sample.

A semiempirical formula that is useful in relating the intralaboratory precision to the interlaboratory precision is represented by the following relationship:

$$S_B = [S_{\bar{X}} - \frac{S_W}{N}]^{1/2} \qquad (44)$$

where: S_W represents the within-laboratory precision;
S_B represents the between-laboratory precision;
$S_{\bar{x}}$ denotes the standard error or the overall standard deviation; and
N is the number of replicate observations or measurements used for finding the average (mean).

The following example illustrates the relationship between S_W and S_B.

* * *

EXAMPLE:

The data for a water quality sample containing cadmium as tested at three laboratories were as follows:

Laboratory	Average concentration (μg/L)	Standard deviation (μg/L)
1	15.5	0.31
2	16.1	0.20
3	15.8	0.25

Each average was obtained from four replications.

Assuming that the three standard deviations are estimates of one and the same population standard deviation, it is quite proper to pool the variances and take the square root of the pooled variance. Using this procedure, the best estimate of the within-laboratory standard deviation (S_W) is obtained as:

$$S_W = \frac{[(0.31)^2 + (0.20)^2 + (0.25)^2]^{1/2}}{3} = 0.257$$

If the standard error of the mean ($S_{\bar{x}}$) is found for the three averages, 15.5, 16.1, and 15.8, we obtain:

$$S_{\bar{X}} = \frac{[(0.30)^2 + (0.30)^2 + (0.0)^2]^{1/2}}{2} = 0.300$$

If the laboratories displayed no systematic differences, the standard deviation that has been calculated from the averages (of four replicates) should be equal to $S_W/\sqrt{4}$ (cf. standard error of the mean: $\sigma_{\bar{X}} = \sigma_X/\sqrt{N}$); i.e.,

$$S_W / \sqrt{4} = \frac{0.257}{2} = 0.129$$

The fact that the calculated value of $S_{\bar{x}}$ is larger than 0.129 can be explained only through the presence of an additional between-laboratory component of variability. This component, expressed as S_B (between-laboratory precision), is calculated by subtracting the anticipated variance S_W^2/N from the observed variance $S_{\bar{x}}^2$ and taking the square root:

$$S_B = \sqrt{(0.300)^2 - (0.129)^2} = 0.271$$

In general, if $S_{\bar{x}}$ and S_W are known, then S_B can be easily calculated using Equation 44.

* * *

5.1.2 Total Precision

In the analysis of water samples, cyclical replication (e.g., triplication) at sampling sites is used in an attempt to identify contamination problems (unless they are systematic) as well as sampling errors. To distinguish between the errors occurring in the field and those occurring in the laboratory, repetitive testing is conducted in the laboratory on replicate samples and single samples. It is then

possible to extract the field component error from the total error, using the following approximate equation:

$$\text{Field precision} \approx \begin{matrix}\text{Total precision} \\ \text{(replicate samples)}\end{matrix} - \begin{matrix}\text{Laboratory precision} \\ \text{(repeats per sample)}\end{matrix}$$

A sample subdivided in the field and preserved separately should be used whenever possible to assess the variability of sample handling, preservation, and storage along with the variability of the analysis process. If the nature of the matrix, the sample acquisition procedure, or the analytical technique prevents the assessment of the entire measurement system, the replicated samples used to assess precision should be selected to incorporate as much of the measurement system as possible.

5.1.3 Precision and Concentration

Frequently, the precision of a method is not constant over the concentration range of interest. When the precision varies significantly with concentration, a plot of precision versus mean concentrations should be made in order to describe or determine the concentration dependence. If the precision varies curvilinearly with the concentration level, a graphical plot should be presented, with the equation for the curve included.

5.1.4 Interpretation of Precision Data

Interpretation of precision data must always be based on a clear knowledge of how the data were created. For example, the precision of the entire measurement system, including sample acquisition, can only be assessed by analyses of "collocated" samples. Precision data generated from multiple analyses of a standard will describe only the stability of the measurement device or instrument and represent only the ultimate precision that could be achieved for a field sample if the sampling activity, subsequent sample preparation steps, and the sample matrix had no impact on final results.

5.2 Accuracy and Bias

Accuracy is a measure of the closeness of an individual measurement (or the average of a number of measurements) to the true (or accepted) value. Accuracy is determined by analyzing a reference material of known concentration (or value) or by reanalyzing a sample to which a material of known concentration has been added (recovery method). For example, if T is the total concentration of a water quality parameter in a spiked sample, U is the concentration of the same parameter in the unspiked sample, and A is the amount of the parameter added to the hitherto unspiked sample, then the percent recovery R is calculated as:

$$R = \frac{(T - U)}{A} \times 100 \qquad (45)$$

Besides being expressed as a percent recovery (R), accuracy can also be stated in terms of a percent bias (R-100).

Determination of accuracy always includes the effects of precision, so the accuracy expression must also describe the variability. Thus, accuracy can be reported as a 95% probability, which is the mean bias or percent recovery ± 2 standard deviations. Alternatively, accuracy can be expressed in terms of the bias associated with a given measurement. The term "bias" can be expressed as an estimate (B) of the difference between the average value ($\bar{X}$) of a set of measurements on a standard and the reference value of the standard (T), as shown in Equation 46:

$$B = \bar{X} - T \tag{46}$$

Other estimates of bias are percent bias:

$$\%B = 100\ (\bar{X} - T)/T \tag{47}$$

and the average percent recovery (R):

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i \tag{48}$$

where:

$$R_i = 100(C_i - U_i)/T \tag{49}$$

C_i is the analytical result from the spiked sample;
U_i is the analytical result from the separate analysis of the unspiked sample; and
T is the reference value of the standard used as a spike.

The relationship between percent bias and percent recovery is:

$$\%B = R - 100 \tag{50}$$

* * *

EXAMPLE:

A reference chemical sample at an assumed concentration, T = 65 concentration units (CU), was analyzed 10 times to assess the precision and bias of the analysis phase of the measurement system. The values (x_i) observed were 63; 62; 65; 64; 65; 63; 64; 62; 66; 64 CU.

Values of precision (S) and bias (B) may be calculated as shown below:

Measurements	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
i = 1	63	0.9	0.81
2	62	1.9	3.61
3	65	1.1	1.21
4	64	0.1	0.01
5	65	1.1	1.21
6	63	0.9	0.81
7	64	0.1	0.01
8	62	1.9	3.61
9	66	2.1	4.41
10	64	0.1	0.01

Average:

$$\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_{10})/10$$
$$= (63 + 62 + 65 + \ldots + 64)/10 = 63.9$$

Bias:

$$B = \bar{x} - T$$
$$= 63.9 - 65.0 = -1.1 \text{ CU}$$

Variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{10} (x_i - \bar{x})^2 \qquad (53)$$
$$= (0.81 + 3.61 + \ldots + 0.01)/9 = \frac{15.70}{9} = 1.744$$

Precision:

$$S = \sqrt{1.744} = 1.32$$

Based on the derived results, the estimated bias of the measurement system, under the above conditions, is -1.1 CU, and the estimated precision, expressed as the standard deviation, is 1.32 CU. On a practical basis, the bias is not considered to be significant, because the assumed value T is contained within the range of the 10 measured values.

* * *

5.2.1 Significance of the Analytical Error

In testing the significance of the analytical error, which is assumed to be systematic, the t-test can be applied to determine whether a bias exists or not. The t-statistic can be calculated as follows:

$$t = \frac{\bar{x}_M - x_R}{\sqrt{\frac{S^2}{N}}} \qquad (51)$$

where: $\bar{x}_M$ is the mean measured value of replicate analyses of a reference sample spiked into blank solutions;
x_R is the true or reference value;
S^2 is the sample variance; and
N is the number of replicate determinations.

If the value of t is less than or equal to the tabulated value (using the Student's t-tables for a two-tailed test) for N-1 degrees of freedom for the chosen confidence level, the random error and not the systematic error can be considered as contributing to the difference between $\bar{x}_M$ and x_R at the identified confidence level.

5.2.2 Accuracy and Concentration

As with precision, the accuracy of a method may not be constant over the concentration range of interest. When a significant concentration dependence is observed, it can be reported by a linear slope and intercept, by an alternative functional relationship, or by a graphical plot or table.

The accuracy estimate may apply to a portion of the measurement system, in which case appropriate adjustment must be made in order to estimate the accuracy for the entire measurement system.

5.3 Method Detection Limit

The method detection limit (MDL) is the minimum concentration that can be measured with 95% confidence that the true concentration is above zero (some texts use the 99% confidence limit).

The MDL is determined by measuring the variability of replicate measurements at zero or near-zero sample concentration. Depending on the method, this may be accomplished by measuring:

- a zero concentration reference material;
- the sample carrier (in gas chromatographic analysis);
- unexposed sample reagent or extract; or

- similar zero concentration matrices.

For methods that preclude actual zero measurements, nonzero concentration measurements as close to zero as practical can be used.

More appropriately, the MDL can be regarded as the smallest concentration of the analyte in distilled water (for water quality samples) that a method can detect with a stated level of accuracy. The MDL is statistically different from the response obtained from a blank carried through the complete analytical method, including chemical extraction and pretreatment of the sample. With this definition in focus, the MDL is expressed as:

$$MDL = \overline{C}_b + tS_b \quad (52)$$

where: C_b is the average concentration value for the blank replicates;
S_b is the standard deviation of the replicate determinations; and
t corresponds to the t-statistic for n-1 degrees of freedom at a 95% confidence level (5% significance level).

In the past, there have been many terms used to designate detection limits, and these terms have been defined in different ways (see Chapter 3 of Part III). Lower limits of detection (LLD), minimum detection amount (MDA), method detection limit (MDL), practical detection limit (PDL), and limit of detection (LOD) are some of the terms that find common usage.

Most authorities in the QA/QC field agree that the smallest detectable quantity, by whatever name, is related to the standard deviation of sample analyses at or near zero analyte concentrations or values (Gaskin 1990).

Recently, an alternative approach to the calculation of detection limits has been developed. This approach utilizes the theory of "propagation of errors" and is illustrated in the following outline.

The IUPAC definition of limit of detection states that the limit of detection, expressed as a concentration C_L, is derived from the smallest measure Y_L that can be detected with reasonable certainty for a given analytical procedure. In algebraic terms, this is written as:

$$Y_L = \overline{Y}_B + k\,S_B \quad (53)$$

where: Y_L is the smallest discernible analytical signal;
$\overline{Y}_B$ is the average response to the blank;
S_B is the standard deviation of the responses; and
k is a factor chosen according to the desired confidence level (usually k = 2 or 3).

From the value of Y_L, the detection limit C_L may be calculated from the equation:

$$C_L = (Y_L - \overline{Y}_B)/m \tag{54}$$

where: C_L is the detection limit in concentration units; and
m is the analytical sensitivity, i.e., the slope of the line describing the calibration curve.

Combining Equations 53 and 54 leads to Equation 55 for the IUPAC detection limit:

$$C_L = k\, S_B/m \tag{55}$$

The IUPAC detection limit tacitly assumes that there are no errors involved in the parameters that are determined from the regression analysis of the system responses to the standards. In other words, it is assumed that the slope and intercept of the line are determined exactly. Knowing this not to be the case, a more rigorous attempt has been made in the determination of the detection limit (Long and Winefordner 1983; Doull 1987).

The concentration of the analyte in a sample is determined from a rewritten version of the regression equation describing the straight line through the data points for the standards. The regression equation

$$Y_I = a + bx_I \tag{56}$$

can be rewritten in terms of x_i to give Equation 57:

$$x_I = (Y_I - a)/b \tag{57}$$

where x_i has been substituted for C_L, and b for m, in Equation 54. The variance S_x^2 in x_i can be determined directly by propagation of the errors in Y, a, and b, as shown in Equation 58:

$$S_x^2 = \left(\frac{\delta x}{\delta Y}\right)^2 S_Y^2 + \left(\frac{\delta x}{\delta a}\right)^2 S_a^2 + \left(\frac{\delta x}{\delta b}\right)^2 S_b^2 \tag{58}$$

For the partial derivatives, the following results are obtained:

$$\frac{\delta x}{\delta Y} = \frac{\delta}{\delta Y}\,\frac{(Y-a)}{b} = \frac{1}{b} \tag{59}$$

$$\frac{\delta x}{\delta a} = \frac{\delta}{\delta a}\,\frac{(Y-a)}{b} = -\frac{1}{b} \tag{60}$$

$$\frac{\delta x}{\delta b} = \frac{\delta}{\delta b}\,\frac{(Y-a)}{b} = \frac{a-Y}{b^2} \tag{61}$$

and C_L can be rewritten as:

$$C_L = kS_x = k\{(\frac{1}{b})^2 S_Y^2 + (\frac{-1}{b})^2 S_a^2 + \frac{(a-Y)^2}{b^4} S_b^2\}^{1/2}$$
$$= \frac{k}{b}\{S_Y^2 + S_a^2 + \frac{(a-Y)^2}{b} S_b^2\}^{1/2} \quad (62)$$

If it is assumed that $S_a = S_b = 0$, then Equation 62 reverts to Equation 55 (around a blank reading of 0, for $Y = 0$).

The difference in results from the use of the two computational methods (Equations 55 and 62) is best illustrated with a worked example.

* * *

EXAMPLE:

Utilizing the data in Table 8, compare the results for the determination of the limit of detection using the IUPAC and propagation of error models.

The results of a regression analysis fitting the data given in columns 1 and 2 of Table 8 to a linear model are:

a =-3.55
b = 7.54
r^2 = 0.9952
S_a = 1.222
S_b = 0.0596

Using these data and assuming that the average signal for the blank is zero (with background correction) and has a standard deviation the same as that of the lowest standard (see Table 8), the detection limit values are calculated as follows:

IUPAC:

$$C_L = \frac{3(0.5)}{7.54} = 0.20$$

Propagation of errors:

$$C_L = 3S_x = (3)\frac{1}{7.54}\{0.5^2 + 1.22^2 + (\frac{-3.55}{7.54})^2(0.060)^2\}^{1/2}$$
$$= 0.52$$

Table 8. Measurements of Various Analyte Concentrations to Compare Detection Limit Determinations by Two Different Mathematical Approaches

Concentration (x)	Responses (Y)	Mean value ($\bar{Y}$)	$(Y-\bar{Y})^2$	Degrees of freedom (DF)
0.5	3,3,4,3,3,3, 3,3,4,4,4,4 4,3,3,3	3.375	3.75	15
2.5	17,17,17,16 17,19,19,19 15	17.333	16.00	8
5.0	31,32,32,32 32,30,30,32 32,32,36,35 37,36,29,32	32.500	80.00	15
10.0	67,60,69,65 66,67,69,76 75,77,75,75, 75,64,74	71.214	268.36	14
25.0	179,182,176, 171,173,177, 195,197,195, 195,188,168, 161,172	180.286	1820.86	13
50.0	379,375,359 361,373,403 400,389,359	377.556	2274.22	8
Totals			4463.19	73

The results of the two calculations show that the value of C_L determined via Equation 62 is larger than that determined via Equation 55. This is not unexpected when one considers the fact that the standard deviations in the slope and the intercept may not be negligible. Any determination of C_L that includes the standard deviations for the slope and intercept will be larger, if one wishes to maintain the same confidence level as used in the more simplistic equation (Equation 55).

* * *

5.3.1 Limits for Practical Detection, Quantitation, and Validation

Various analytical and computational means have been used to determine the smallest quantity of a given analyte that can be measured with an expected certainty using a particular analytical

methodology. Such quantities as the practical detection limit (PDL), the limit of quantitation (LOQ), and the limit of validation (LOV) have found acceptance in the determination of analytes near the blank level.

The PDL is the smallest concentration of the analyte in a real sample matrix that an analytical method can detect at a stated level of accuracy and which is statistically different from the response obtained from a blank carried through the complete method. For water quality samples, the PDL is different in definition from the MDL by virtue of the type of sample matrix involved. Whereas the MDL relates to distilled or deionized water as the matrix, the PDL is more general and relates to other matrices besides distilled or deionized water.

The PDL is calculated in a similar manner to the MDL.

The LOQ is defined as:

$$LOQ = \overline{S}_b + 10\ SD \qquad (63)$$

where: $\overline{S}_b$ is the average signal (or level) for the blank;
SD is the standard deviation of the replicate determinations.

Equation 63 defines the level above which quantitation is reliable and at the same time indicates a region between MDL and LOQ where detection is reliable but quantitation is not. The LOQ is specifically the level above which quantitative results may be obtained with a chosen degree of confidence.

LOVs are the lower and upper levels at which a method has been experimentally validated. Results that do not fall between these limits become questionable and in some cases are not valid. The LOV effectively establishes the quantitation range of a method; if a specific application requires quantitation below the lower LOV, validation should be conducted at or below the required level.

5.4 Other Data Quality Assessment Criteria

5.4.1 Completeness and Representativeness

Completeness is a measure of the number of valid data obtained from a measurement system and is expressed as a percentage of the number of valid measurements that should have been (i.e., were planned to be) collected (Mitchell *et al.* 1985). As used in this context, "data completeness" is not necessarily the same as "data representativeness," which represents how closely the measured results reflect the actual concentration or pollutant (analyte) distribution in the media sampled. Thus, a study could have 100% data completeness (i.e, all samples planned to be collected were actually collected and found to be valid), but the results do not accurately reflect the pollutant (or analyte) concentration actually present. For example, the method might be biased, or the sampling points might not be representative of the average distribution in the media.

When a low level of data completeness is reported for a method, the method should be scrutinized carefully for possible inherent causes.

5.4.2 Data Comparability and Compatibility

Data comparability is based on the measure of confidence with which one data set can be compared with another.

Data compatibility among data sets relies on the likeness and consistency with which the data sets being compared are acquired (e.g., similar sampling, analyses, data procedures, and treatments; similar QA/QC protocols; similar reporting data units).

The achievement of comparability and compatibility of data requires that:

- the objectives of the data collection program(s) are clearly defined;
- data requirements are specified unambiguously;
- appropriate, documented sampling and measurement procedures are selected and correctly used;
- common data formats are used for the presentation of intra- and interregional data sets;
- sound QC procedures are in operation; and
- results are correctly archived with their associated uncertainties.

The establishment of data comparability and compatibility between the data sets of groups and agencies engaged in similar water quality work provides a basis for the enhancement of data quality and credibility.

Interlaboratory and collaborative testing programs can serve as aids for ensuring a closeness in data inter-relationship.

5.5 Data Review and Evaluation

The review and evaluation of data should be centred around a number of performance indicators, such as the accuracy and precision with which the data were gathered and the representativeness, ruggedness, and completeness of the assembled data base.

The regional branches of the WQB and provincial agencies should participate in interlaboratory and interagency data comparison exercises as one of the means of evaluating the accuracy, ruggedness, and completeness of a number of their data sets. The provinces and the WQB should also

singly and jointly review and evaluate those data bases that are of particular relevance to the QA/QC programs within the Federal–Provincial Water Quality Monitoring Agreement and the QA/QC sequential sample triplication programs. For example, the data review and evaluation process for the sequential sample triplication program could aid in the assessment of the effectiveness of the present level of sample replication and could help to determine if there is a need for increased replication or if the current level is adequate.

Data review and evaluation exercises not only should be structured to address aspects of accuracy, precision, etc., but could also link the performance indicators to the stated objectives of the given measurement program.

5.6 References

Doull, J.A. 1987. Determination of Detection Limits using the IUPAC and Propagation of Errors Approach, and a Discussion on Calibration Curve Modelling. Water Quality Branch (Atlantic Region), Environment Canada, Moncton, N.B.

Gaskin, J.E., T. Dafoe, and P. Brooksbank. Analyst, 1990, Vol. 115, pp. 507-510.

Long, G.L., and J.D. Winefordner. 1983. Limit of detection, a closer look at the IUPAC definition. Anal. Chem. 55: 712A.

Mitchell, W.J., R.C. Rhodes, and F.F. McElroy. 1985. Determination of measurement data quality and establishment of achievable goals for environmental measurements. Am. Soc. Test. Mater. Spec. Tech. Publ. 867: 41–52.

6.0 STATISTICAL TECHNIQUES FOR WATER QUALITY DATA ANALYSIS

The use of statistics in water quality sampling design and the use of the statistical approach to both analytical method evaluation and QC of routine environmental analyses are on the increase. This increase is the result of more planners and analysts becoming familiar with the benefits to be derived from the use of statistics and the need for closer agreement among agencies doing similar work. In addition, the proper use of statistical methods in the reduction and analysis of water quality data allows analysts to take an objective and comprehensive view of the data being analyzed.

Although a number of statistical packages designed for hand-held calculators or computers have made data processing and data analysis relatively easy, it is still important for the data analyst and the data user to clearly understand the theory behind the statistical methods that are employed.

Most data analysis that is currently being performed by WQB personnel employs parametric statistical methods. Some of the basic assumptions underlying parametric statistical tests are as follows:

(1) The observations or determinations made are independent.

(2) The distribution of the population is known.

(3) The variances of any set of populations being compared are equal or are of known ratio.

In addition, the parametric statistical tests require that the data:

- be collected in a random manner;
- have error variation independent of the means and be normally distributed and homogeneous; and
- have variance components that are additive.

The data commonly encountered in water quality studies rarely satisfy all the assumptions and requirements of parametric statistical methods. In some cases, nonparametric methods must be used in order to carry out the required analyses. The nonparametric methods are associated with distribution-free statistics (i.e., there are no precisely defined μ, σ, $\bar{x}$, or s values). However, the relevant concern is not whether the assumptions and requirements are met exactly, but rather whether plausible violations of the same assumptions and requirements have serious consequences on the validity of the probability statements based on them. There are procedures that can be applied to raw data to test if the requirements are met and to what extent (Glass *et al.*, 1972).

The following sections give a brief summary of the statistical techniques that find ready application to the analysis of water quality data.

6.1 Descriptive Statistics

(1) Mean, Median, and Mode: Used to locate central tendencies within a data set.

(2) Geometric Mean: Also an indicator of central tendency for data distributions that are greatly skewed to the right. This statistic is particularly useful for describing water quality random variables that vary over several orders of magnitude, such as biological data (e.g., coliform counts).

(3) Range, Variance, and Standard Deviation: Used to determine level(s) of data dispersion (spread).

(4) Bar Charts (Graphs), Histograms, Frequency Polygons: Used to group and display data in methodical fashion, either as tables or as graphs.

6.2 Theoretical Probability Density Functions for Discontinuous (Discrete) Data

(1) Binomial Distribution: Used in assessing issues such as effects and noneffects, or occurrences and nonoccurrences, in water quality data, particularly biological data.

(2) Poisson Distribution: Used in addressing data where discrete independent measurements (numbers) are directly proportional to the size of the interval (as time, length, area, volume, etc.). This distribution can be applied effectively to physical and biological data.

6.3 Theoretical Probability Density Functions for Continuous or Measured Data

(1) Normal Distribution: Characterized by raw data acquired by random collection from a population having a normalized functional distribution of $\mu = 0$ and $\sigma = 1$. The distribution is ideally applicable to 30 or more data measurements acquired under statistically controlled conditions.

(2) Lognormal Distribution: If $Y = \log X$ and if Y is normally distributed, then X is lognormally distributed, where X is the quantity being measured and Y is the response variable. The range of values for X is 0 to $+\infty$, and the range for Y is $-\infty$ to $+\infty$. The lognormal distribution generally finds strong application among biological data in water quality measurements.

(3) Chi-Square Distribution (χ^2): Used to determine whether the results (data) of an experiment deviated from expected results.

(4) F-Distribution: Generally used as a comparative test for the variances of two data sets arising from two populations with equal population variances (e.g., comparing two water quality data sets from the same sampling site using two dissimilar sampling devices).

(5) Student's t-Distribution: Similar in application to the normal distribution but particularly applicable to data having a small sample size (<30). The t-distribution is dependent on f, the number of degrees of freedom (DF = n-1 for n measurements and one estimated parameter such as the mean; DF = n-k, where k is the number of estimated parameters and n is the number of data values).

6.4 Confidence Limits

(1) About a Single Observation from a Normal Distribution

If μ (the population mean) and σ (the population standard deviation) are known, then a single observation has a probability of 1-σ of falling within the following tolerance limits:

$$\mu - Z_{\alpha/2} \cdot \sigma \leq Y \leq \mu + Z_{\alpha/2} \cdot \sigma$$

(2) About a Population Mean

$$CI_{1-\alpha} : \overline{Y} - Z_{\alpha/2} \cdot \sigma/\sqrt{n} \leq \mu \leq \overline{Y} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

This expression holds when $n \leq 30$, the population is normally distributed, and σ is known; or when $n \geq 30$. If σ is unknown and $n \geq 30$, σ is estimated by s the sample variance.

(3) Inference about a Mean (sample mean) using a t-distribution

$$CI_{1-\alpha} : \overline{Y} - t_{\alpha/2} \cdot s/\sqrt{n} \leq \mu \leq \overline{Y} + t_{\alpha/2} \cdot s/\sqrt{n}$$

For this expression, it is assumed that the population variance is unknown, there are n-1 degrees of freedom, the distribution is normal, or at leat symmetry and unimodality exist.

Limits in (1) find particular application in making inferences about a single observation and those in (2) and (3) find application as rational criteria for selecting sampling frequencies, respectively.

Confidence limits can also be extended to geometric means.

6.5 Inferential Statistics

(1) Hypothesis Testing: Making predictions (related to experimental data, events, etc.) before statistical tests are conducted, with the possibility that the prediction could be in error. Rejecting a true hypothesis leads to an error called a type I error, and failure to reject a false hypothesis leads to a type II error.

(2) Estimation Theory: Making estimates or inferences on a population based on sample statistics. The population could be a group (suite) of water quality parameters or a body of data.

6.6 Testing for Normality of Data Measurement Distribution

Tests used for investigating normality include:

- Normal probability plots;
- Chi-square goodness-of-fit tests;
- Kolmogorov-Smirnov Test ($N > 50$);
- Chebyshev's Inequality Test;
- Empirical Test or Proportional Test ($N > 30$); and
- Wilk-Shapiro Test ($N \leq 50$).

6.7 Testing for Homogeneity of Variances

Tests for homogeneity of variances include:

- F-test;
- Bartlett's Test; and
- analysis of variance (ANOVA).

6.8 Comparison of the Means from Two Populations

This can be done for:

- unknown populations' variances and unpaired data (pooled variance statistics); and
- unknown populations' variances and paired data (matched or paired test).

6.9 Simple Linear Regression

(1) Perform a least squares determination of the simple linear line ($Y_c = a + bX$) for a bivariate data set.

(2) Determine the correlation coefficient (r), representing the degree of association between the two variables, and the coefficient of determination (r^2), representing the proportion of variability in Y explained by the linear relationship.

(3) Test the significance of the simple linear regression line (i.e., test the significance of the slope, intercept, and correlation coefficient).

6.10 Analysis of Covariance

This compares the simple linear regression lines developed from several groups of data:

- testing first for common slopes; and
- testing for a difference in the levels of regression (if the slopes tested for are common).

6.11 Multiple Regression

(1) Perform a least squares determination of the multiple regression line ($Y_1 = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n$).

(2) Determine the coefficient of multiple determination ($\gamma_{1\,2\,3\cdots n}$).

(3) Test the significance of the multiple regression line.

6.12 Curvilinear Regression

Water quality data collected from streams will very often follow the relationship: $Y = aX^b$. This function can be linearized by using a logarithm transformation ($\log Y = b \log X + \log a$).

Other curvilinear functions, such as $Y = a + b^x$ or $Y = a(x-b)^2$, require complex procedures for regression.

6.13 Nonparametric Statistics

(1) Sign Test: For related samples or related data sets (such as "before" and "after" sample collections or analytical measurements) (see Section 3.3.1).

(2) Wilcoxon Signed-Ranks Test: Applied to data from related samples, taking into account the ranked differences between two samples, where ranks have been assigned according to the magnitude of the difference between matched pairs of subjects (see Section 3.3.2).

(3) Mann-Whitney U-Test: A useful test for evaluating the difference between two means of independent samples of elements, without making any assumption of the distribution(s) involved.

(4) Chi-Square Test: Can be applied to a single sample, two independent samples, or k independent samples and is used in determining whether the observed values differ from the

expected values, due to chance, when the chance is defined according to a particular set of rules. No assumptions need to be made concerning the underlying distributions of the variables when the Chi-Square Test is being applied.

(5) Kolmogorov-Smirnov Test: Finds use in a single sample test to determine whether or not the distribution of observations is significantly different from a theoretical distribution (e.g., normal distribution, t-distribution). The method also finds use with small independent samples to determine whether two independent samples could have come from the same distribution.

(6) Other Nonparametric Tests: These tests (Siegel 1956; Downie and Heath 1965) include:

- Cochran Q-Test for investigating differences between related samples;
- Cox-Stuart Test for Trend;
- Friedman Two-way ANOVA Test;
- Kendall's Tau for Bivariate Random Samples (measuring association between original variables);
- Kruskal-Wallis One-way ANOVA Test;
- McNemar Test for significance of data changes;
- Spearman's Rho Value for measuring the degree of association between two sets of ranked data; and
- Wald-Wolfowitz Runs Test for determining whether or not two samples could have come from the same population of elements.

6.14 Data Modification

(1) Trimmed Means: While the mean is the most useful and convenient measure of central tendency in a data set, an outlier can pull the mean in its direction and cause the mean to appear atypical of the values in the data set. For this reason, the idea of the TRIMMED MEAN has been developed. To find a 20% 'Trimmed Mean' for instance, the highest 10% and lowest 10% of the values are dropped after the data values are arranged in ascending order of magnitude, and the mean is computed from the remaining values. In general, if a data set consists of N values and the number of values dropped from both ends is equal to L, then the number of remaining values is equivalent to N-2L, and,

$$\bar{X}_T = \frac{1}{(N - 2L)} \sum_{L+1}^{N-L} X_i = \text{Trimmed Mean}$$

(2) Winsorization Process: Consider the case where a data set of size n contains g observations below the detection limit. The observations are then arranged in ascending order of magnitude, and the 'g' non-detected values are individually replaced by the first value above the detection limit, while the 'g' largest values are individually replaced by the (n-g)th value. This procedure is called level 'g' Winsorization (Dixon and Massey 1969; Gaskin 1988). The Winsorized mean $\bar{x}_w$ is the mean of the observations in the modified data set, and is given as:

$$\bar{x}_w = \frac{1}{n} \sum_{i}^{n} x_i$$

When 'g' is the level of Winsorization, and n is the sample size, h = n-2g represents the number of unchanged observation.

EXAMPLE:

As an example of the use of (a) 'Trimmed Means' and (b) 'Winsorization', consider the following data obtained for the total phosphate parameter from 10 water quality samples:

ND, ND, 7.5, 6.0, 8.3, 2.2, 3.6, 4.5, 9.1, 5.8

The measurements are in μg/L and the detection limit is 2.0 μg/L. The data are arranged in ascending order of magnitude to give:

ND, ND, 2.2, 3.6, 4.5, 5,8, 6.0, 7.5, 8.3, 9.1

(a) TRIMMED MEAN:

The re-arranged data are modified by eliminating the two ND values and the two largest values to give:

2.2, 3.6, 4.5, 5,8, 6.0, 7.5

The Trimmed Mean is then calculated as:

$$\bar{X}_T = (2.2 + 3.6 + \ldots 7.5)/6 = 4.93$$

(b) WINSORIZATION:

For a Winsorization level of g = 2, the modified data become:

2.2, 2.2, 2.2, 3.6, 4.5, 5.8, 6.0, 7.5, 7.5, 7.5

From these values, $\bar{X}_w = (2.2 + 2.2 + \ldots + 7.5)/10 = 4.90$

(3) Data Reduction: Combining or grouping closely related data into subsets in order to reduce the number of computational exercises or to facilitate a better understanding of data values (or trend in data values). The process of data reduction can also involve the increase or diminution of the individual numerical values of the data set by arithmetical operations (e.g., the addition of a constant quantity to each data value, or the multiplication of each data value by a constant).

6.15 Sophisticated Data Handling Techniques

Sophisticated data handling techniques include the following:

- autocorrelation in trend analysis;
- Box-Jenkins plots;
- cumulative distribution functions;
- mathematical and statistical modelling;
- Monte Carlo simulation;
- numerical methods;
- random generation of numbers;
- spectral analysis; and
- time series analysis.

6.16 References

Downie, N.M., and R.W. Heath. 1965. Basic Statistical Methods. Harper, New York, N.Y.

Dixon, W.J. and F.J. Massey. 1969. Introduction to Statistical Analysis. McGraw-Hill.

Gaskin, J.E. 1988. Approximate Estimation of Analytically non-detected Values for Water Quality Parameters. Water Quality Branch, IWD, Environment Canada, Ottawa, Ontario, Canada.

Glass, C.V., P.D. Peckham, and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Rev. Educ. Res. 42: 237–288.

Siegel, S. 1956. Nonparametric Statistics for the Behavioural Sciences. McGraw-Hill, New York, N.Y.

7.0 DATA PRESENTATION AND DATA REPORTING

7.1 Data Presentation

7.1.1 Presentation of Statistical Parameters

The American Society for Testing and Materials (1986) has made the following recommendations for the presentation of experimental data that relate to the case where one is dealing with a sample of n observations of a single variable and data values have all been obtained under the same essential conditions:

(1) Present at least the average ($\bar{x}$), the standard deviation (s), and the number (n) of observations taken.

(2) If the number of observations is large, present also the value of the "skewness" (g_1) and the value of the "kurtosis" (g_2). An alternative procedure is to present a grouped frequency distribution.

(3) If the data were not obtained under statistically controlled conditions and it is desired to give information regarding the extreme observed effects of assignable causes, present the values of the maximum and minimum observations in addition to the average, the standard deviation, and the number of observations.

(4) Present as much evidence as possible to support statements that the data were obtained under statistically controlled conditions.

7.1.2 Supporting Data

Presented data should be accompanied and supported by available relevant information, particularly information on precisely the conditions under which the measurements were made and evidence that the data are good. Relevant information should relate to:

- the method of selecting the sample; steps taken to ensure its randomness or representativeness (the manner in which the sample is taken has an important bearing on the interpretability of data);
- the specific method of test (if an ASTM or other standard test is made, this should be stated, together with any modifications of procedure);
- the specific conditions of test, particularly the regulation of factors that are known to have an influence on the feature or system under investigation;

- the precautions or steps taken to eliminate systematic or constant errors of observation;
- the difficulties encountered and eliminated during the investigation;
- information regarding parallel but independent paths of approach towards the same end results; and
- evidence that the data were obtained under statistically controlled conditions; the results of statistical tests made to support belief in the constancy (statistical control) of conditions, with respect to the analytical work carried out on the sample(s).

7.1.3 Tabular and Graphical Presentation

The information contained in most bodies of data may be summarized by presenting a tabular grouped frequency distribution, if the number of observations is large. A graphical presentation of a distribution makes it possible to visualize the nature and extent of the observed variation (Demayo and Whitlow 1986).

Although some condensation is effected by presenting grouped frequency distributions, further reduction is necessary for most of the uses that are made of water quality data. This need can be fulfilled by means of a few simple functions of the observed distribution, notably the average and the standard deviation.

7.1.4 Confidence Limits

In presentation of data, it is desirable to affix limits to a number of computed statistical parameters (e.g., $\bar{x}$), and it is quite important that the significance of the limits be clearly indicated.

For an observed value of X that is subject to the uncertainties that arise from sampling and analytical fluctuations and which tends to vary from the true average more widely if the number of observations N is small, the overall uncertainty in the measurements should be included as:

$$\overline{X} \pm k \frac{s}{\sqrt{N}} \qquad (63)$$

where k is the value obtained from the Student's t-distribution table for N-1 degrees of freedom at a chosen significance level (e.g., $p = 0.05$).

7.1.5 Rounding-off Procedures for Numerical Values

The following working rule is recommended in carrying out computations incident to determining averages, standard deviations, and limits for averages:

(1) In all arithmetical operations (e.g., adding, subtracting, multiplying, dividing, squaring, extracting square root) on the sample of N observed values, the equivalent of two or more places of figures than in the single observed values should be retained. For example, if observed electrical conductance values are read or determined to the nearest microsiemen, the numbers should be carried to the nearest 0.01 μS in the computations.

(2) Deleting places of figures should be done after computations are completed, in order to keep the final results substantially free from computational errors. Deleting places of figures should be conducted according to the rounding-off procedure that has been suggested by the American Society for Testing and Materials (1986).

7.1.6 Control Chart Method for Data Presentation

The control chart method for the presentation of data provides a criterion for detecting lack of statistical control. The control chart method of analysis and data presentation is used for the following two distinct purposes:

- to discover when observed values of $\bar{X}$, S, R, etc., for several samples of N observations each, vary among themselves by an amount greater than what should be attributed to chance.

- to discover whether observed values of $\bar{X}$, S, R, etc., for samples of N observations each, differ from standard values $\bar{X}_0$, σ_0, R_0, etc. by an amount greater than what should be attributed to chance.

In both uses, the display of the control chart consists essentially of symmetrical limits (control limits) placed above and below a central line. The central line in each case indicates the expected or average value of $\bar{X}$, S, R, etc. for subgroups (samples of N observations each) (American Society for Testing and Materials 1986).

The control limits used are referred to as 3-sigma control limits and are placed at a distance of three standard deviations from the central line; here, the standard deviation refers to the standard deviation of the sampling distribution of the statistical measure in question (e.g., $\bar{X}$, S, R) for subgroups (samples) of size N. This standard deviation is not the computed standard deviation of the subgroup values (of $\bar{X}$, S, R, etc.) plotted on the chart but is computed from accepted formulas (American Society for Testing and Materials 1986). For example, for large samples of equal size, in the absence of standard values, the formulas for control chart lines become:

	Central line	Control limits
For averages, $\bar{X}$	$\bar{\bar{X}}$	$\bar{\bar{X}} \pm 3\frac{\bar{S}}{\sqrt{N}}$
For standard deviations, S	$\bar{S}$	$\bar{S} \pm \frac{3\bar{S}}{\sqrt{2(N-1)}}$

where: $\bar{\bar{X}}$ = the grand average of the observed values of X for all samples that make up the subgroups (1 to k)

= $(\bar{X}_1 + \bar{X}_2 + {}_3 + \ldots + \bar{X}_k)/k$; and

$\bar{S}$ = the average subgroup standard deviation

= $(S_1 + S_2 + S_3 + \ldots + S_k)/k$, where S_1 is the standard deviation for N samples or measurements in subgroup 1, S_2 is that for N measurements in subgroup 2, and so forth.

In the case of small samples of equal size, the control chart lines are as follows:

	Central line	Control limits
For averages, $\bar{X}$	$\bar{\bar{X}}$	$\bar{\bar{X}} \pm 3\frac{\bar{S}}{C_4\sqrt{N}}$
For standard deviations, S	$\bar{S}$	$\bar{S} \pm \frac{3\bar{S}}{C_4\sqrt{2(N-1)}}$

The value for C_4 is found in Table B-1 (Appendix B).

7.2 Data Reporting

7.2.1 Internal Data Reporting

The internal reporting process should provide sufficient information to support claims made for the results, with an accompaniment of estimates of uncertainty. Any new methodology used should be described in detail, including the results of exhaustive tests of the methodology. The use of existing methodology can be cited by reference to published literature, provided that the published description is complete and adequate and that any modifications have been fully tested and reported.

The field and laboratory sections should provide data reports to project officers (or resource officers) on a regularly scheduled basis in order to promote timely data processing (data verification, data transfer, and data reporting to clients and data users).

7.2.2 External Reporting to Clients and Data Users

All data should be subject to peer review before external release. All releases should be in writing (typed). Verbal releases (including telephone releases) should be permitted when appropriate, provided that they are subsequently (within approximately 48 hours) confirmed in writing, and provided the data have been previously reviewed and approved via established practices.

The lowest level of review for release of data must be by the project officer (or resource officer) or his/her designate. In the case of sensitive data, a higher level of review may be required.

Publication of data is predicated on editorial review. Data obtained for a sponsor are considered to be the property of the sponsor, and their dissemination is subject to the sponsor's release (within the Freedom of Information Act principles).

Reports should be prepared in a format suited to the particular end use. Each report should contain, as far as practical, the following information (Taylor 1985):

- a summary of the problem investigated;
- a complete description of samples;
- description of methodology (or reference);
- summary of data, and reference to original data set(s);
- interpretation/conclusions; and
- attestation.

The signatures of analyst(s), supervisors, laboratory head, or respective designates must be affixed to the report.

7.2.3 Reporting of Analytical Values

The appropriate expression "ND" (none detected) or "L" (less than) must be used in cases in which no analyte is detected. The term 0 (zero) must not be used to report an analytical result. If the result falls between the method detection limit (MDL) and the practical detection limit (PDL), the value may be reported in parentheses, with an explanation that the result is less than the PDL and the recovery and precision have not been evaluated at that level (see Chapter 3 of Part III).

The practice of using symbols such as T or 'tr ' to represent trace amounts and the use of similar terms or statements to denote nondefinite analyte concentrations should be avoided because of the relative nature of such terminology, the confusion surrounding it, and the danger of its misuse.

Data measured at or near the detection limit are associated with two problems. The uncertainty can approach and even equal the reported value. Furthermore, confirmation of the species reported is virtually impossible; hence, the identification must depend solely on the selectivity of the methodology and knowledge of the absence of possible interferents. These problems decrease when measurable amounts of analytes are present. Accordingly, quantitative interpretation, decision-making, and regulatory actions should be limited to data at or above the limit of quantitation (LOQ), where, as previously defined, the LOQ is the level above which quantitative results may be obtained with a specified degree of confidence. Confidence in the apparent analyte increases as the analyte signal increases above the limit of detection (LOD). The value for LOQ = 10σ is recommended and corresponds to an uncertainty of $\pm$ 30% in the measured value ($10\sigma \pm 3\sigma$) at the 99% confidence level (Taylor *et al.* 1983).

7.2.4 Quality Assurance Reporting

Two types of QA reporting should be employed by the WQB: reports of data quality to management, and QA reports to data users. QA reports to WQB management will report on the level of QA that has been applied to the water quality monitoring program, its effectiveness, and recommendations on QA-related problems. The QA reports to management should also provide a means of evaluating the cost of the QA activities in terms of the increased data quality.

7.2.4.1 Data Quality Reports to Management

The QA officer (or alternate) should prepare QA reports to management on a regular (quarterly or semiannual) basis. These reports should discuss the status of the field and laboratory QA systems and summarize the QA-related activities that have been carried out over the previous quarter. They should include the following:

- a summary of data status including data recovery summaries and reasons for lost or voided data;
- the level of QA that has been applied to the data;
- estimates of field and laboratory precision and accuracy where applicable, for the data produced over the last quarter or six-month period;
- a summary of new QA/QC-related activities that have been implemented over the last quarter, and a measure of their effectiveness;

- a summary of outstanding QA/QC-related problems and recommendations for action (with schedules);

- an appraisal of the effectiveness of the QA/QC system (over the last quarter or six months) relative to QA costs;

- a summary of the results of interlaboratory exercises and comparisons to assess data comparability between participating water quality laboratories and agencies;

- a summary of any audit results from the previous period; and

- a list of revisions or additions to QA/QC procedures.

These reports will provide the means of documenting QA-related activities and problems for management (Vet and Onlock 1983) and help to establish the priority areas to be addressed by the WQB QA/QC system.

7.2.4.2 Data Quality Reports to Users

A data quality summary should be provided with each data report produced by the WQB. This provides users with an indication of the confidence they may place in the parameters that are of interest to them. These data quality summaries should illustrate the WQB's concern for collecting and reporting quality data by briefly documenting the QA steps that have been applied to the monitoring program and the acquired data prior to publishing. The summaries should also document the specific data quality parameters associated with the particular data set. The following QA elements should be reported with each data set produced:

- a brief summary of the WQB's QA procedures (sampling, field testing, and laboratory analysis) as they relate to data quality;

- a summary of the analytical methods used for each parameter, the capabilities of the methods, and a summary of detection limit values (and how they may be interpreted) for each parameter of interest;

- a summary of data recovery and reasons for lost or invalid data;

- the level of QA/QC in the field and laboratory as applied to any given data set, including:
  - repeat analyses
  - blanks
  - duplicates
  - prepared samples (QC samples);

- estimates of accuracy and precision that may be applied to the reported values of a particular data set;

- specific QA-related problems that may affect the data reported at a particular site or with a given analytical method;

- a summary of data screening and validation criteria along with data verification procedures applied to the data;

- a description of all comment codes that may be reported in the data set; and

- summaries of interlaboratory comparisons and collaborative tests that may have been undertaken to evaluate the comparability and compatibility of data between participating water quality laboratories and agencies.

This type of summary provided with each data report will prove invaluable to data users in assessing the applicability of a given data set to a particular use. It will also help to promote feedback between data users and WQB management with respect to users' needs relative to the WQB's emphasis on data quality. This feedback will help to fine-tune the WQB QA program to meet data users' requirements.

7.3 References

American Society for Testing and Materials. 1986. Manual on Presentation of Data and Control Chart Analysis. 6th ed. Am. Soc. Test. Mater. Publ. 15 D. Part 3, Philadelphia, Pa.

Demayo, A., and S. Whitlow. 1986. The Use of Graphs in the Field of Water Quality. Water Quality Branch, Inland Waters Directorate, Environment Canada, Ottawa, Ont.

Taylor, J.K. 1985. Principles of Quality Assurance of Chemical Measurements. Center for Analytical Chemistry, National Bureau of Standards, U.S. Department of Commerce, Gaithersburg, Md.

Taylor, J.K., R.A. Libby, and G. Wentler. 1983. Principles of environmental analysis. Anal. Chem. 55: 2210–2218.

Vet, R.J., and S.G. Onlock. 1983. The Canadian Air and Precipitation Monitoring Network Quality Assurance Plan for Precipitation Monitoring Systems. CSC 110. 194-3-1, Air Quality and Inter-Environmental Research Branch, Atmospheric Environment Service, Environment Canada, Downsview, Ont.

8.0 DATA CUSTODY AND DATA TRANSFER

8.1 Data Custody

The WQB's objectives of data custody should be to ensure that water quality data are of known quality and, if need be, are legally defensible.

The QA objectives in the area of data custody must be to ensure that data handling and data management procedures follow well-organized chains of custody, and to ensure that every sample appears in the data base.

The QA procedures should include:

- development of a chain-of-custody system for acquired data;
- development of job responsibility descriptions for all individuals in the chain-of-custody system;
- use of field and laboratory tracking forms;
- confirming the existence of chain-of-custody systems in the field and laboratory;
- implementing a chain-of-custody system for the data centre;
- implementing a procedure for authorizing changes to data bases; and
- implementing checks by the QA officer or his/her alternate to ensure that data bases are complete.

8.2 Data Transfer

The mechanism and schedule for data transfer should be determined by the laboratory and field sampling data processing personnel in order to ensure that the respective formats for data reports and data tapes are suitable and that reasonable schedules (turnaround times) can be maintained.

The data transfer procedures and schedules should be documented in the respective field and laboratory operational manuals.

The data QA program must ensure that correct and complete data are recorded and transmitted through all stages of the measurement program. This will best be ensured by the use of:

- data recording forms;
- data entry procedures;

- data manipulation routines; and
- the use of trained personnel.

Complete and accurate transfer of data from and through the following stages of the monitoring program must also be assured:

- data recording in the field and laboratory;
- data transcribing and computer entries;
- data screening and validation;
- data verification and review;
- data storage; and
- data retrieval and reporting.

The QA officer (or alternate) should coordinate the procedures and protocols that are established to ensure and facilitate the transfer of data through and between the various work stations. In addition, the officer should institute a data traceability mechanism for pinpointing the location of any specific piece or block of data at any given time.

9.0 MANAGEMENT OF WATER QUALITY DATA BASES

The objective of the data management operation of the WQB must be to correctly and efficiently gather, process, verify, report, and archive the data associated with water quality samples collected in the water quality monitoring and survey programs. In order to meet this broad objective, the following operational and QA/QC objectives must be achieved:

(1) A well-conceived set of data recording forms, data entry procedures, and data manipulation routines must be designed and documented, and all operators and analysts must be well trained in their use.

(2) The quality of the data being generated and transferred through each stage of the monitoring or survey program must be assured and reported in a timely manner.

(3) Data must be reported and stored in a manner that facilitates access and interpretation by users and clients. Data reports should include relevant field, laboratory, and QA/QC information (e.g., measures of precision, accuracy, completeness, detection limits, method and instrument sensitivities), as well as analytical results and simple statistical information.

9.1 Data Processing

The data collection and entry systems must provide the means of transferring the sampling and analysis information from the field and laboratory into the preliminary (working) data base as quickly and as efficiently as possible. This system should function efficiently and effectively to ensure that all relevant information is reported at each stage of the process and that potential errors in data recording and transmission are minimized.

9.2 Laboratory Data Base Management

The principal goal of the laboratory data management system should be to provide the network with a complete, quality-controlled set of analytical results that can be merged with the field sampling information for subsequent data validation. The laboratory QA/QC system should provide the laboratory with the means of monitoring and reporting the quality of the analytical data produced.

The transfer of data from the laboratory to the data management centre should occur on a regular basis, in a timely, structured fashion. The frequency and methods of data transfer should be determined during the sampling and network design stage.

Laboratory support data should also be forwarded to the data management centre on a routine basis. This information should be used in conjunction with the sampling QA information in order to document the overall quality of the data produced in a given project or program.

9.3 Data Management Information System

The regional data management centre should be responsible for all aspects of data management, including:

- entering field data to the data base;
- merging and editing field and laboratory data;
- the creation of a working data base;
- data screening and validation;
- data evaluation;
- final data archiving; and
- data reporting.

The first data handling step at the data centre should consist of a review of all data sheets immediately upon reception, to:

- determine that a data sheet is received from each site that has been sampled;
- check for labelling errors and ensure that each data sheet is legible for data entry operations and key-punch operations; and
- ensure that complete information is recorded for all water quality samples, and note any field-related problems that may cause errors in the data base at a later date (e.g., samples partially spilled or samples possibly mislabelled).

The preliminary review at the data reception stage would be useful in that it provides the mechanism to correct any omissions or errors in the data forms while the sampling operator is still familiar with the samples. Immediate initiation of corrective action will also be facilitated by this system. The results of the preliminary review at the data reception stage should be summarized for future use in data validation and QA considerations.

Once the data reception and initial review have been completed for a block of field data, the field sampling information should be entered into the working base, and a computer listing of the total information should be generated. This computer listing should then be manually compared with the original data sheets in order to ensure that no data transcription errors or omissions of sample identification parameters have occurred. At this stage, the field data become ready for merging with the laboratory data once all analyses have been completed.

The merging of field and laboratory data should be carried out on the data management centre's computer system. A software package should be developed to match the field and laboratory data for all samples submitted and synthesize them into a single data set. This computer program should list samples that do not have matching field and laboratory data and should have the capability of correcting and editing the working data base.

9.4 Data Base Structure

The structure of the data base and data analysis system must also play an important role in the WQB data management system. From an operational perspective, it is important that the data base be structured in a manner that lends itself to applications, i.e., one that facilitates data entry, data manipulation, and statistical analysis. For QA purposes, the data archiving system must provide the means of ensuring that a complete correct data set is maintained, that it is secure (i.e., restricted access for editing), and that it has the capability of retrieving QA/QC information for all data collected.

Some of the important features that should comprise the data base structure are as follows:

(1) *The data format:* The data base must be structured to ensure that all relevant parameters and sampling information are included in a standardized format. The individual parameter fields must be capable of storing a sufficient number of characters (significant figures) corresponding to the format of the resulting data. A numbered character flag field (two or three characters) could be assigned to each parameter in order to permit data flagging for individual parameters. Each parameter field could also have a space for a negative sign to accommodate the requirements for indicating data at the detection limit or for indicating other comments. In order to provide a consistent data set that is easily manipulated, there should be some type of coded identifiers in use to indicate missing or invalid data. Numerical codes such as -999 or -888 could be used to facilitate the handling of such data.

(2) *Data handling routines:* A standardized package of data access and data handling routines must be developed early in the monitoring program to provide the WQB with the capability of efficiently extracting and sorting subsets of data. These routines should be used to retrieve and manipulate data for data analysis purposes as well as for generating data sets (tapes/records) for external users.

(3) *Data security:* Access to the WQB's data archives must be restricted to ensure that only authorized personnel have access to editing and deletion routines. A system of backup data tapes and files should also be employed to prevent accidental destruction of all or part of the data set.

(4) *QA/QC information:* A data filing system should be set up to ensure that the QA/QC information that has been generated is available for all samples. The QA/QC data do not necessarily need to be stored in a computerized data base but must be stored in a manner that makes them readily available for interpretation of data quality.

(5) *Future data requirements:* The data base should be structured such that it is capable of future expansion: i.e., expansion in the number of stations, in the number of parameters reported, and in inputs from QA/QC developments.

(6) *Data base parameters:* All relevant sampling and analysis information should be stored in appropriate data units in the data base. In addition to the parameters that are routinely reported, any information about the sample that may be useful, but not routinely reported, can be stored. During the phase that the data base is under development, careful consideration must be given to the parameters that will be included, the reporting units used, the parameter fields used to store them, and the flagging mechanism(s). The decisions on which parameters to report and the exact data base structure should be made once the network reporting objectives and data base resources are known.

9.5 Data Storage and Retrieval

The storage of data in hierarchical levels for data files in the NAQUADAT system (Environment Canada 1983) provided an exemplary system that facilitated retrieval of water quality data. The essential details that were recorded in the storage files were:

(1) Station (sampling point)
- station reference
- sample location
- descriptive narrative

(2) Sample
- date of sample
- project and agency
- comments about samples

(3) Parameter
- analytical method
- analytical measurement of value.

A gradual changeover is being made from the hierarchical system of storing data. The replacement is the "relational structure," in which logical relationships between information levels are established by means of pointers (addresses of storage locations), and the access and retrieval processes are driven by a sophisticated retrieval language. With the use of unique identifiers for individual samples and a change in the data base management system (DBMS) from System 2000 (see Chapter 8 of Part IV) to ORACLE, the new data storage and retrieval arrangement should prove superior to the former system.

Attention is also currently being directed towards the structuring of data bases to allow the inclusion of QA/QC values and statements both as records for the primary user(s) and to indicate to

secondary users the likely suitability of the data for their purposes. Recommendations being developed for the storage of QA/QC data relate to:

- the levels and amount of QA/QC data that should/could be stored in the NAQUADAT, AQUALABS, and similar data base management systems, and the separate storage of an extensive library of data relating to the various analytical measurement processes, accompanied by auxiliary sampling and field analysis measurements;

- the separation of QA/QC data to reflect laboratory, regional, external, and national requirements, not only in the area of storage, but also in the subsequent activities of accessing and retrieving specific blocks of data. This QA/QC data separation into hierarchical storage segments would be a function of the frequency of the need to re-examine the stored data from time to time. The periodic examination of stored data could facilitate the fine-tuning of the regional and national laboratories' analytical operations, especially in areas where corrective measures are needed; and

- the format and criteria that would be used in storing data segments and the accommodation of different levels of data to reflect, among other factors, the precision, accuracy, and confidence level.

The regions should participate in efforts to establish both a national and inter-regional storage and retrieval system for QA/QC data and for efficient information exchange. All data systems (regional and national) should be structured to provide data in a format that is useful to data users and to issue data in a timely fashion. A common format for gathering and reporting data will allow effective cross-linkages to be set up between regional field offices, regional laboratories, regional headquarters, the National Water Quality Laboratory (NWQL), provincial data inventory systems, and the NAQUADAT storage and retrieval system at the WQB national headquarters. The common format could also allow the development of compatible interpretive reports relating to issues of both regional and national concerns.

Each region must be able to review and capture its data from the NWQL data banks for those samples that are still incomplete in the Burlington laboratory. This automatic and ready data capture will eliminate the task of re-entering data from hard-copy printouts into the region's data system.

Regional branches should encourage the development of captive command procedures for regional personnel to access the Burlington data system (AQUALABS) and create files of data from any sample, whether or not the data compilation for the sample is complete.

At the intra-regional level, the WQB could set up an automated cross-reference subject index with a parameter-based vocabulary, controlled by a coded system for accurate storage and comprehensive information retrieval for both general and QA/QC data. This controlled arrangement of data records and retrieval would help to ensure information integrity and security. The automated

inventory system at the WQB regional headquarters could be integrated with the data sub-systems in regional (WQB) and provincial district offices and laboratories.

9.6 References

Environment Canada. 1983. NAQUADAT data files. Water Quality Branch, Inland Waters Directorate, Ottawa, Ontario.

10.0 TRAINING IN DATA HANDLING

All personnel involved in the collection of water quality data must be adequately trained for their appointed tasks and must be made aware of the importance of obtaining data of known quality.

All personnel participating in the handling of water quality data must become familiar with the QA program for data handling and data management and must be cognizant of the various ways of achieving and maintaining good-quality data.

Personnel training in data handling should start before operations begin and should continue during operations, with periodic reviews, upgrading, and retraining when indicated.

On-the-job training in the QA aspects of the data handling program should be mandatory and should involve direct training and instructions rather than serial training.

Data handling personnel must be trained in the proper use of report forms, record-keeping methods, and procedures for storing and retrieving data via the computer terminal.

Instructions and guidance in the application and interpretive use of statistical techniques should be part of the continuing training program given to selected personnel.

External training (via colleges, technology centres, universities, or specially offered courses) may be required for personnel desiring familiarity, upgrading, or competency in data management techniques and procedures.

Recommended training systems include the preparation of manuals, development of a training program, initial training, follow-up training, updating/upgrading, and training under the direction of an external agency.

11.0 DATA AUDITS

The data audit program could be a stand-alone program or could be incorporated into the total audit program, which not only addresses operational elements but also gives consideration to personnel practices, training, personnel needs, and workload.

The data audit program should consist of a review of the total data production process. This would include on-site reviews of the field and laboratory's operational systems and physical facilities for sampling, calibration measurement protocols, and measurement procedures.

The national QA/QC Technical and Advisory Committee should design and supervise the QA/QC data audit program for the integrated operational activities at the regional and national levels.

Guidelines in the audit program will specify who will conduct the audit, what protocols and procedures will be used, what the acceptance criteria will be, and to whom the audit reports will go.

11.1 Objectives of the Data Audit Program

The audit program will be used as an essential and effective tool for both project coordination and QA/QC activities in water quality monitoring and will help to identify the corrective measures that must be put in place in order to remedy or eliminate the errors, defects, and problems that have been discovered.

11.2 Audit Team

The audit team should be composed of two or three individuals. Its size should permit it to conduct the audit exercise in a timely manner and at the same time cause as little disruption as possible in the work (or work station) being audited. The team should be made up of persons who are separate and independent from the operation being audited.

The auditors should have enough scientific and technical knowledge, audit training, and experience to understand the work and the requirements of the QA/QC program.

The audit team should be well prepared before the audit is undertaken. It should decide or be told what elements of the data collection operations are to be evaluated. It should determine dates and inform the facility to be audited as to the date that the audit will be made, what will be covered, and who will make the inspection.

In some circumstances, the audit team may want to perform unannounced audits. However, such unannounced visits should be infrequent and should be conducted only in emergency or extraordinary conditions.

In preparation for the audit visit, the audit team should study relevant background information on the operational unit(s) to be visited. Included among this information should be documentation pertaining to organizational structure, staff size, expertise and experience of the various personnel, principal instruments and apparatus, operational procedures and methodologies, the QA/QC program currently in place, past and present problem areas, and previous audit reports (where these exist).

11.3 Audit Visit

The visit to the audit site should commence with a meeting between the audit team and the head or supervisor of the operational section to be audited. Other staff members of the section may be present at this meeting at the discretion of the section head or supervisor. An additional familiarization meeting may be held with the staff at large, in order to create a more receptive climate to the audit and to emphasize that the principal aim of the audit is to bring about improvements in the operational activities that contribute to the data collection processes.

Prior to the audit tour, arrangements should be made with the section head or supervisor to have the audit team accompanied by a staff person. This action enables the staff to observe the audit operation firsthand and gives them the opportunity to discuss any operational problems encountered.

During the audit, if suspected problems are encountered or if deviations from standard practices are uncovered, they should be brought to the attention of the staff representative, so that any misunderstanding can be clarified or a course of action decided on in order to alleviate the problem. All deviations should be recorded for later discussions with the section head or supervisor.

On completion of the audit tour, the audit team should meet with the section head, the staff representative, and any other staff members invited by the section head so that a verbal debriefing may be undergone in order that any misunderstanding and biases may be corrected or eliminated.

11.4 Audit Evaluation and Reporting

The personnel of the audit section should be encouraged to respond in writing to the audit findings, as disclosed at the debriefing session. This will permit the audited group to offer its views and to comment on the findings and proposed recommendations of the audit team.

A report of the audit's findings should be prepared for regional management or for senior management when requested or advised.

In preparing its report and recommendations, the audit team should be as objective as possible, and the audit's findings should be ranked in order of importance to indicate the degree of concern for each finding.

For the audit process to be truly successful, a follow-up mechanism is needed to ensure that the needed changes and corrections are made.

Freeberg (1980) and Garfield (1984) provided a number of general and useful suggestions on steps to be taken in planning and conducting an audit program for analytical laboratories.

11.5 References

Freeberg, F.E. 1980. Optimizing chemical laboratory performance through the application of quality assurance principles. *In* Meaningful Quality Assurance Program for the Chemical Laboratory. Edited by F.M. Garfield *et al.* Association of Official Analytical Chemists, Arlington, Va., pp. 13–23.

Garfield, F.M. 1984. Quality Assurance Principles for Analytical Laboratories. Association of Official Analytical Chemists, Arlington, Va., pp. 98–108, 185–187.

Appendix A: STATISTICAL TABLES

Table A-1. Dixon Criteria for Testing of Extreme Observation (Single Value)

n		Criterion	Significance level % 10	5	1
3	$Y_{10} = \frac{x_2-x_1}{x_n-x_1}$	if smallest value is suspected;	0.886	0.941	0.988
4			0.679	0.765	0.889
5			0.557	0.642	0.780
6	$= \frac{x_n-x_{n-1}}{x_n-x_1}$	if largest value is suspected.	0.482	0.560	0.698
7			0.434	0.507	0.637
8	$Y_{11} = \frac{x_2-x_1}{x_{n-1}-x_1}$	if smallest value is suspected;	0.479	0.554	0.683
9			0.441	0.512	0.635
10	$= \frac{x_n-x_{n-1}}{x_n-x_2}$	if largest value is suspected.	0.409	0.447	0.597
11	$Y_{21} = \frac{x_3-x_1}{x_{n-1}-x_1}$	if smallest value is suspected;	0.517	0.576	0.679
12			0.490	0.546	0.642
13	$= \frac{x_n-x_{n-1}}{x_n-x_2}$	if largest value is suspected.	0.467	0.521	0.615
14	$Y_{22} = \frac{x_3-x_1}{x_{n-2}-x_1}$	if smallest value is suspected;	0.492	0.546	0.641
15			0.472	0.525	0.616
16			0.454	0.507	0.595
17			0.438	0.490	0.577
18	$= \frac{x_n-x_{n-2}}{x_n-x_3}$	if largest value is suspected.	0.424	0.475	0.561
19			0.412	0.462	0.547
20			0.401	0.450	0.535
21			0.391	0.440	0.524
22			0.382	0.430	0.514
23			0.374	0.421	0.505
24			0.367	0.413	0.497
25			0.360	0.406	0.489

Source: Nelson, C.A., Jr., D.W. Armentrout, and T.R. Johnson. 1980. Validation of Air Monitoring Data. EPA-600/4-80-030, June, U.S. Environmental Protection Agency.

$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_{n-2} \leq x_{n-1} \leq x_n$
Criterion Y_{10} applies for $3 \leq n \leq 7$
Criterion Y_{11} applies for $8 \leq n \leq 10$
Criterion Y_{21} applies for $11 \leq n \leq 13$
Criterion Y_{22} applies for $14 \leq n \leq 25$

Table A-2. Critical Values for 5% and 1% Tests of Discordancy for Two Outlines in a Normal Sample

N	5%	1%
4	0.967	0.992
5	0.845	0.929
6	0.736	0.836
7	0.661	0.778
8	0.607	0.710
9	0.565	0.667
10	0.531	0.632
12	0.481	0.579
14	0.445	0.538
16	0.418	0.508
18	0.397	0.484
20	0.372	0.464
25	0.343	0.428
30	0.322	0.402

Source: Barnett, V., and T. Lewis. 1978. Outliers in Statistical Data. John Wiley & Sons, New York, N.Y., p. 311.

Table A-3. Critical T Values for One-sided Grubbs Test when Standard Deviation is Calculated from Sample

No. of observations (n)	Upper significance level (%)					
	0.1	0.5	1.0	2.5	5.0	10.0
3	1.155	1.155	1.155	1.155	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.615	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.275	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.706	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.847	2.699	2.607	2.462	2.331	2.175
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2.997	2.806	2.705	2.549	2.409	2.247
16	3.052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.620	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.967	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	2.730	2.549
30	3.507	3.234	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.544	3.270	3.135	2.938	2.773	2.591
33	3.545	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.342	3.204	3.003	2.835	2.650
38	3.644	3.354	3.214	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.844	2.682
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
44	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768
:	:	:		:	:	::
147	4.219	3.883	3.727	3.509	3.334	3.144

Source: Grubbs, F.E., and G. Beck. 1972. Extension of sample sizes and percentage points for significance test of outlying observations. Technometrics 14(4). For the complete table, see Nelson, C.A., Jr., D.W. Armentrout, and T.R. Johnson. 1980. Validation of Air Monitoring Data. EPA-600/4-80-030, June, U.S. Environmental Protection Agency.

Table A-4. Wilcoxon Signed-Rank Test

No. of pairs (n)	Critical Values			
	$\alpha \leq 0.10$	$\alpha \leq 0.05$	$\alpha \leq 0.02$	$\alpha \leq 0.01$
1				
2				
3				
4				
5	0,5			
6	2,19	0,21		
7	3,25	2,26	0,28	
8	5,31	3,33	1,35	0,36
9	8,37	5,40	3,42	1,44
10	10,45	8,47	5,50	3,52
11	13,53	10,56	7,59	5,61
12	17,61	13,65	9,69	7,71
13	21,70	17,74	12,79	9,82
14	25,80	21,84	15,90	12,93
15	30,90	25,95	19,101	15,105
16	35,101	29,107	23,113	19,117
17	41,112	34,119	28,125	23,130
18	47,124	40,131	32,139	27,144
19	53,137	46,144	37,153	33,158
20	60,150	52,158	43,167	37,173
21	67,164	58,173	49,182	42,189
22	75,178	66,187	55,198	48,205
23	83,193	73,203	62,214	54,222
24	91,209	81,210	69,213	61,239
25	100,225	89,236	76,249	68,257

Source: Documenta Geigy, Scientific Tables, 7th ed., Ciba-Geigy Ltd., Basel, Switzerland (see Nelson, C.A., Jr., D.W. Armentrout, and T.R. Johnson. 1980. Validation of Air Monitoring Data. EPA-600/4-80-030, June, U.S. Enviornmental Protection Agency).

Table A-5. Rank Sum Test (α = p[H_0 is true], for α = 0.10).

n_1	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
n_2	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y	T_IT_Y
4	11-25	17-33	24-42	32-52	41-63	51-73	62-88	74-102	87-117	101-133	116-150	132-168	150-186	168-206	187-227
5	12-28	19-36	26-46	34-57	44-68	54-81	66-94	78-109	91-125	106-141	121-159	138-177	155-197	173-218	193-239
6	13-31	20-40	28-50	36-62	46-74	57-87	69-101	82-116	93-133	110-150	126-168	143-187	161-207	179-229	199-251
7	14-34	21-44	29-55	39-66	49-79	60-93	72-108	85-124	99-141	113-158	131-177	148-197	166-218	186-239	206-251
8	15-37	23-47	31-59	41-71	51-85	63-99	75-115	89-131	104-148	119-167	136-186	153-207	172-228	192-250	212-274
9	16-40	24-51	33-63	43-76	54-90	66-105	79-121	93-138	108-156	124-178	141-195	159-216	178-238	198-261	219-285
10	17-43	26-54	35-67	45-81	56-96	69-111	82-128	97-143	112-164	128-184	146-204	164-226	184-248	204-272	226-297
11	18-46	27-58	37-71	47-86	59-101	72-117	86-134	100-153	116-172	133-192	151-213	170-235	190-258	210-283	232-308
12	19-49	28-62	38-76	49-91	62-106	75-123	89-141	104-160	120-180	138-200	156-222	175-245	196-268	217-293	239-319
13	20-52	30-63	40-80	52-93	64-112	78-129	92-148	108-167	123-187	142-209	161-231	181-254	201-279	223-304	245-331
14	21-53	31-69	42-84	54-100	67-117	81-133	96-154	112-174	129-193	147-217	166-240	186-264	207-289	229-315	252-342
15	22-58	33-72	44-88	56-103	69-123	84-141	99-161	116-181	133-203	152-225	171-249	191-274	213-299	235-326	259-353
16	24-60	34-76	46-92	58-110	72-128	87-147	103-167	120-188	138-210	156-234	176-258	197-283	219-309	242-336	266-364
17	25-63	35-80	47-98	61-114	75-133	90-153	106-174	123-196	142-218	161-242	181-267	202-293	225-319	248-347	273-375
18	26-66	37-83	49-101	63-119	77-139	93-159	110-180	127-203	146-226	165-251	186-276	208-302	231-329	253-357	280-386

Source: Nelson, C.A., Jr. D.W. Armentrout, and T.R. Johnson. 1980. Validation of Air Monitoring Data. EPA-600/4-80-030, June, U.S. Environmental Protection Agency.

Appendix B: FACTORS FOR ESTIMATING CONTROL LIMITS OF SHEWART CHART

Table B-1. Factors for Computing Control Chart Lines for Means and Standard Deviations (in the Absence of Standard Samples)

Observations in samples (n)	Factor for control limits (c)	Observations in sample (n)	Factor for control limits (c)
2	0.7979	14	0.9810
3	0.8862	15	0.9823
4	0.9213	16	0.9835
5	0.9400	17	0.9845
6	0.9515	18	0.9854
7	0.9594	19	0.9862
8	0.9650	20	0.9869
9	0.9693	21	0.9876
10	0.9727	22	0.9882
11	0.9754	23	0.9887
12	0.9776	24	0.9892
13	0.9794	25	0.9896

Source: American Society for Testing and Materials. 1986. Manual on Presentation of Data and Control Chart Analysis. 6th ed. Am. Soc. Test. Mater. Spec. Tech. Publ. 15 D, Part 3, Philadelphia, Pa.

Part VI

GENERAL GLOSSARY

GLOSSARY OF QA/QC TERMS FOR ENVIRONMENTAL MONITORING

Absolute method	A method in which characterization is based entirely on physically (absolute) defined standards.
Acceptance sampling	The procedures by which decisions to accept or reject a sampled area or population are made, based on the results of a sample inspection or survey.
Accountability	Liability to be called to account for an accepted responsibility including a judgement upon the performance of carrying out that responsibility.
Accreditation	A formal process by which a laboratory is evaluated, with respect to established criteria, for its competence to perform a specified kind(s) of measurement. The term also relates to the decision based upon such a process.
Accreditation criteria	A set of requirements used by an accreditation system that a laboratory must meet in order to be accredited.
Accredited laboratory	A laboratory to which accreditation has been granted.
Accuracy	The degree of agreement of a measured value with the true or expected value of the quantity of concern.
Activity	Projects or processes carried out to achieve an objective or a set of objectives of a program.
Adsorption	The surface retention of solid, liquid or gas molecules, atoms or ions by a solid or liquid.
Aliquant	A divisor that does not divide a sample into a number of equal parts without leaving a remainder. The term also refers to a sample resulting from such a divisor (see also aliquot).
Aliquot	A divisor that divides a sample into a number of equal parts, leaving no remainder; also refers to a sample resulting from such a divisor (see also aliquant).
Alternative hypothesis	A statement which provides an alternative or opposite claim to that made in the null hypothesis.

Analyte	The specific component measured in a chemical analysis, also called analate.
Analytical grade	A chemical with a level of purity high enough to permit its use in precise analytical determinations.
Analytical reagent blank	A prepared sample which is expected to contain a negligible or imperceptible amount of the substance of interest, and which is used as a base line for the analytical portion of a method.
Arithmetic mean	The sum of observations divided by their number; also called the average.
Assignable cause	In quality assurance studies, the term represents excess variability (sometimes bias as well), the source(s) of which can be identified and possibly corrected or minimized by appropriate actions. (Also called special cause). See also chance cause.
Attribute	A qualitative characteristic of a population of individuals belonging to a population.
Attribute data	Data pertaining to some selected attribute or attributes of a population. An example is the number of defects of a particular kind, or a number of identified species (variables of interest) observed in a given sample of a population.
Audit	A systematic check to determine the quality of operation of some function or activity. Audits may be of two basic types: (a) performance audits in which quantitative data are independently obtained for comparison with routinely obtained data in for example,a measurement system for environmental monitoring, or (b) system audits of a qualitative nature which consist of an on-site review of a laboratory's quality assurance system, and its physical facilities for sample handling, calibration, analytical measurements and data handling.
Availability	The fraction or percentage of time during which an item is expected to perform satisfactorily (in the reliability sense), taking into account its reliability and its maintainability, or the percentage of uptime of an item or piece of equipment, as contrasted with its percentage of inoperation or downtime.

Average deviation	The arithmetic mean of the absolute values of the deviations of the n individuals, X_i, in a data set from the mean, X, computed from them.
Between-laboratory precision	The variability between results obtained on the same material in different laboratories in interlaboratory analyses. Reproducibility is used as an alternative term for between-laboratory precision.
Bias	A systematic error which is inherent in a method of measurement or caused by some artifact or idiosyncracy of the measurement system. Biases may be both positive and negative and several kinds can exist concurrently, so that only net bias can be evaluated in many cases. Bias is measured as the difference between the limiting mean and the true value measured by a system that is in statistical control. (Also called offset by some metrologists).
Blank	The measured value obtained when a specified component of a sample is not present during the measurement. In such a case, the measured value/signal for the component is believed to be due to artifacts, and hence should be deducted from a measured value to give a net value due to the component contained in a sample. The blank measurement must be made so that the correction process is valid.
Blind sample	A sample submitted for analysis but whose composition is known only to the submitter and not to the analyst. A blind is thus one way to test the proficiency of a measurement process.
Budget	This is a plan which is expressed in monetary and manpower terms, and covers a specific period of time.
Bulk sampling	Sampling of a material that does not consist of discrete, identifiable, constant units, but rather of arbitrary and irregular units.
Calibrant	A substance used to calibrate or to establish the analytical response of a measurement system.
Calibration	Comparison of a measurement standard or instrument with another standard or instrument to report or eliminate by

adjustment any variation (deviation) in the accuracy of the item being compared.

Calibration standard

A standard used to quantitate the relationship between the output of a sensor and a property to be measured. Calibration standards SRMs or primary standards.

Cause-effect diagram

A graphical representation of the causes that can produce a specified kind of error in measurement. A popular one is the so-called fish bone diagram, first described by Ishikawa, given this name because of its suggestive shape.

Central line

The long-term expected value of a variable displayed on a control chart.

Certification

This is the process by which a non-governmental agency or association grants recognition to an individual who has met certain predetermined qualifications specified by that agency or association.

Certified reference material

This is a reference material where one or more of its property values are certified by a technically valid procedure, accompanied by, or traceable to a certificate or other documentation which is issued by a certifying body.

Certified value

The value that appears in a certificate as the best estimate of the value for a property of a reference material.

Chain of custody procedures

These procedures must assure that:

(a) Only authorized personnel handle the sample.
(b) Only the field sampling techniques specified for the measurement program are used.
(c) A record tag is attached to the sample immediately after collection, and the tag should include the following information : program or project identification; sample field number; location; depth; collection date; time; and collector.
(d) Blank samples with and without added chemicals (preservatives) are interspersed with the actual samples.
(e) All record forms are completed.
(f) The transfer of samples is documented.

(g) Transfer procedures provide for proper protection and preservation. For example, if samples are mailed, they should be sent by certified mail with a request for an acknowledgement of receipt.

(h) Signatures must be appended to documents whenever there is a transfer of samples from one work station (or location) to another. All senders and recipients must acknowledge their particular contact or association with the sample(s). Time, date and appropriate comments should be necessary fixtures to appropriate documents.

Challenge examination

This refers to examinations that may be taken to demonstrate that a person has acquired the knowledge of, or the competency in a given field by non-academic means.

Chance cause

A cause for variability of a measurement process that occurs unpredictably, for unknown reasons, and believed to happen strictly by chance.

Check standard

An artifact measured periodically in physical calibration, the results of which typically are plotted on a control chart to evaluate the measurement process.

Chemometrics

A new discipline involving statistics, mathematics, and chemistry and dealing with the interaction of these sciences.

Coefficient of variation (CV)

The CV or the relative standard deviation (RSD) is a measure of precision that is calculated as the standard deviation of a set of values divided by the mean of the values and usually multiplied by 100 to be expressed as a percentage.

Collaborative testing

The evaluation of an analytical method, under actual working conditions , through the participation of a number of typical or representative laboratories in analyzing portions of carefully prepared homogeneous samples.

Collocated samples

These are two or more portions collected at the same point in time and space so as to be considered identical with respect to the variable(s) of interest.

Common cause

A cause of variability of a measurement process, inherent in and common to the process itself, as contrasted to a special or defined cause.

Comparability	The measure of confidence with which one data set can be compared to another.
Comparative method	A method which is based on the intercomparison of the sample with a chemical standard.
Compatibility	Compatibility among data sets relies on the likeness and consistency with which the data sets being compared are acquired (such as similar sampling, laboratory analyses, data procedures and treatments, similar QA/QC protocols, similar reporting units, etc.).
Completeness	This is a measure of the amount of valid data obtained from a measurement system, expressed as a percentage of the number of valid measurements that should have been, or were planned to be collected.
Complex statistical control	Complex statistical control is required when analyses are performed on different occasions. The correctness of analytical results depends on the proper calibration of instrumentation used to observe the chemical or physical changes caused by the analytical process. The natural variability of measurement of the 'standards' will introduce small errors in calibration. Direct action is required to limit the magnitude of this error and ensure calibration control.
Composite sample	A sample obtained by mixing several discrete samples, or representative portions thereof, into one container.
Confidence coefficient	The chance or probability, usually expressed as a percentage, that a confidence interval has of including the population value. The confidence coefficients usually associated with confidence intervals are 90, 95, and 99 percent. For a given sample size, the width of the confidence interval increases, as the confidence coefficient increases.
Confidence interval	That range of values, calculated from an estimate of the mean and the standard deviation, which is expected to include the population mean with a stated level of confidence. Confidence intervals in the same context also may be calculated for standard deviations, lines, slopes and points.
Confidence limit	The outer boundaries of a confidence interval.

Construct validity	This type of validity deals with how conditions, variables, and practices were defined and carried out in a research study.
Contamination	A foreign or unwanted material which renders a sample unfit for meaningful analyses.
Continuing education	This relates to the participation in any educational experience not normally leading to an academic degree . This may range from a one-day workshop (or seminar) to a formal course stretching over a period of several weeks.
Control	This term implies the establishment of a standard for determining the acceptability of a system or product. The criteria for setting the standard may be quite arbitrary. However, when the standard has been exceeded, the system or product is regarded as defective or unacceptable.
Control limit	The limits shown on a control chart beyond which it is highly improbable that a point could lie while the system remains in a state of statistical control (see also, definition for "Statistical Control Chart Limits").
Control chart	A graphical plot of test results with respect to time or sequence of measurement together with limits within which they are expected to lie when the system is in a state of statistical control (see alternative definition for "Shewart control chart" or "statistical control chart").
Control sample	A material of known composition that is analyzed concurrently with test samples to evaluate a measurement process (see also check standard).
Correlation coefficient (r)	A measurement used to express the degree of association between the independent variable and the dependent variable(s) in algebraic relationships. The square of the correlation coefficient is called the 'Coefficient of Determination'.
Cost effectiveness	Maintaining the correct balance between cost of quality and value of quality.
Credentialing	This is a process of giving recognition to an institution or person as having met certain requirements or standards.

Cross sensitivity	A quantitative measure of the response obtained for an undesired constituent (interferrent) as compared to that for a constituent of interest.
Data point	A value for some variable that may be used as an entity for calculation or statistical analysis. It may be a single value or the mean of several individual values.
Data quality	Data quality is demonstrated by evidence of the care taken to maintain sample validity and integrity, as well as precision and accuracy of measurement. The following factors will impact on data quality : project design, sampling, sample preservation, sample handling and submission to the laboratory, sample storage, sample preparation, analysis and measurement.
Data Quality Objectives (DQO)	Those desired outcomes in which the collected data are accompanied with the best achievable and optimum data quality parameters such as precision, accuracy, data completeness and confidence limit values that can be extracted from the monitoring system.
Data validation	The examination of data to determine their correctness.
Defensibility	This relates to data having the capability or power of withstanding any reasonable adversarial inquiry or action, legal or otherwise. There must always be documented proof that sound work was done if defense is not to be deficient and perhaps impossible. While legal defense in its broadest connotation is often the issue, technical reliability is always at stake. Good quality assurance practices and enabling legislation can lend support and defense to data that have been obtained with good precision and accuracy.
Degrees of freedom	The number of elements that can be chosen freely (independently). Mathematically the number of degrees of freedom, f, can be defined as the number of N independent observations in the sample (i.e. the sample size) minus the number k of population parameters which must be estimated from sample observations. In symbols, $f = N - k$.
Dependent variable	A parameter (usually designated the Y value) whose value is determined by a functionally related independent variable.

Depth-integrated sample

A sample which represents the water-suspended sediment mixture throughout the water column so that the contribution to the sample from each point is proportional to the stream velocity at that point.

Design of experiments

A branch of chemometrics that involves statistically based plans to minimize effort and maximize experimental information.

Deterioration

A decline in the quality of a sample over a period of time due to incorrect handling and/or improper preservation techniques.

Deviation

The numerical difference, d, on an observed value from a computed value or true value. The computed value may be the mean of a data set or a value obtained from an equation, expressing or predicting the values of a dependent variable associated with independent variables. The use will make it clear what kind of deviation is indicated. Deviations are computed by the equation :

$$d = X_o - X_c$$

A positive sign indicates that the observed value (X_o) is larger than the compouted or true one (X_o), and conversely.

Double blind

A sample, known by the submitter but submitted to an analyst in such a way that neither its composition nor its identification as a check sample are known to the latter.

Duplicate measurement

A second measurement made on the same (or identical) sample of material to assist in the evaluation of measurement variance.

Duplicate sample

A second sample randomly selected from a population of interest to assist in the evaluation of sample variance.

Dynamic calibration

This is the calibration of a measurement system by use of a calibration material having characteristics similar to the unknown material to be measured. For example, a gas containing ozone (0_3) of known concentrations in an air mixture could be used to calibrate an ozonometer.

Effectiveness

This is the extent to which a program or an operation produces the results that are intended.

Efficiency	This is the ratio of output per unit of input, e.g. 300 analyses per 10 person-hours = 30 analyses per person-hour or 0.033 person-hour per analysis.
Equivalency examination	This refers to examinations used to equate non-formal learning with learning achieved in academic courses or training programs. Such tests may be used by employers or certifying bodies to qualify individuals whose non-formal study and on-the-job learning is deemed equivalent to that expected from a formal program.
Error	The difference between a measured value and the true value of the quantity or parameter measured. It is due to both random and systematic errors.
Eye estimate	The result of an estimation process in which the only instrument used is the eye. The term is used especially to denote a line fitted to plotted data points in which the best fit is judged by the eye.
False negative	See Type II Error.
False positive	See Type I Error.
Figure of merit	A performance characteristic of a method believed to be useful when deciding its applicability for a specific measurement situation. Typical figures of merit include selectivity, sensitivity, detection limit, precision, and bias.
Flow proportional composite sample	A sample obtained by (1) continuous pumping at a rate proportional to the flow; (2) mixing equal volumes of water collected at time intervals which are inversely proportional to the volume of flow; (3) mixing volumes of water proportional to the flow collected during or at regular intervals. This sample will indicate a 'flow' average water quality condition over the period of time of compositing.
Fluvial characteristics	Of or pertaining to a river or rivers; existing, growing or living in or about a stream or river; produced by the action of a stream or river.

Geometric mean

Mathematically the geometric mean (G) can be expressed as the n^{th} root of the product of all values in a set of n values or the antilogarithm of the arithmetic mean of the logarithms of all the values of a set of n values.

$$G = (X_1 \times X_2 \times X_3 \times X_4 \times \ldots\ldots \times X_n)^{1/n}$$

Goal

This is a short-term measurable result, to be achieved within a given time frame (usually not more than a year) and contributes to the overall attainment of objectives.

Good Laboratory Practice (GLP)

An acceptable way to perform some basic operation or activity in a laboratory, that is known or believed to influence the quality of its outputs. GLPs ordinarily are essentially independent of the measurement techniques used.

Good Measurement Practice (GMP)

An acceptable way to perform some operation associated with a specific measurement technique, and which is known or believed to influence the quality of the measurement.

Grab sample

A sample taken at a selected location, depth and time.

Gross sample

One or more increments of material taken from a larger quantity (lot) of material for assay or record purposes. Alternative terminologies are bulk sample and lot sample.

Groundwater

All subsurface water.

Harmonic mean (H)

This is the reciprocal of the arithmetic mean of reciprocals of two or more quantities. It is especially useful when using ratios of different or similar items as data; for example, rates. It is calculated by the expression :

$$H = \frac{n}{(1/X_1 + 1/X_2 + 1/X_3 + 1/X_4 + \ldots + 1/X_n}$$

where, n is the number of quantities averaged. Thus the harmonic mean of the three speeds : 100 km/hr, 110 km/hr and 120 km/hr is 109.4 km/hr.

Homogeneity

The degree to which a property or substance is randomly distributed throughout a material. Homogeneity depends on the size of the subsample under consideration. Thus a mixture of

	two minerals may be heterogeneous at the molecular or atomic level, but homogeneous at the particulate level.
Homoscedasticity	This is the equivalency of variances among given groups. For example, an analysis of variance assumes homogeneity of variances (homoscedasticity), that is, all of the treatment groups have the same variance.
Hypothesis	This is a tentative statement about an expected relationship between two or more variables. It is merely a supposition of what the relationship might be, and serves as a guide in formulating the proposed research.
Implementation plan (Work Plan)	A document which proposes the levels of human and financial resources to carry out programs in a given fiscal year.
Increment	An individual portion of a material collected by a single operation of a sampling device, from parts of an area or site separated in time or space. Increments may be either tested individually or combined (composited) and tested as a unit.
Individuals	Conceivable constituent parts of a population.
Information value	Value of a property, not certified, but provided because it is believed to be reliable, and to provide information important to the certified material.
Input	Resources (human, financial, material) utilized to produce outputs.
Instrument Detection Limit (IDL)	The instrument detection limit is the lowest concentration of analyte that an analytical instrument can detect and which is significantly different from the response obtained from the background instrumental noise.
Intercalibration	The process , procedures, and activities used to ensure that several laboratories engaged in a monitoring program can produce compatible data. When compatible data outputs are achieved and this situation is maintained, the laboratories can be said to be intercalibrated.
Interlaboratory quality control (Proficiency Testing)	A systematic testing program in which a sample is analyzed by a number of laboratories, or a random series of samples are

	analyzed, to determine the accuracy of an analyte or analytes. Actually it involves analysis of uniform samples by a number of laboratories to assess the continuing capability and relative performance of each laboratory. In addition, the testing serves as an external evaluation to verify the effectiveness of a laboratory's internal quality control program.
Intralaboratory quality control	A continuing, systematic in-house regimen intended to ensure the production of analytical data of continuing high quality and validity.
Issue	Questions of importance which have or could have an impact on the realization of plans or the attainment of goals and objectives.
Judgment sampling	A sampling procedure in which elements are selected on the basis of the researcher's knowledge or opinion.
Laboratory sample	A sample, intended for testing or analysis, prepared from a larger sample or otherwise obtained. The laboratory sample must retain the composition of the larger (gross) sample. Often reduction in particle size (mostly solid samples) is necessary in the course of reducing the original quantity.
Level of significance	The probability which is set by the investigator that a true null hypothesis will be rejected, designated by the Greek letter alpha (α) and which specifies the probability of making a Type I error.
Licensure	This is the process by which an agency of government grants permission to persons meeting predetermined qualifications to engage in a given occupation and/or use a particular title, or grants permission to institutions to perform specified functions.
Limiting mean	The value approached by the average as the number of measurements, made by a stable measurement process, increases indefinitely.
Limit Of Linearity (LOL)	The upper limit of concentration or amount of substance for which incremental additions produce constant increments of response. It can be regarded as the upper level of reliable measurement for all practical purposes.

Limit Of Quantitation (LOQ)	The lower limit of concentration or amount of substance that must be present before a method is considered to provide quantitative results. By convention, LOQ = 10 s_o, where s_o is the estimate of the standard deviation at the lowest level of measurement.
Limit of Detection (LOD)	This is the minimum concentration of a substance that can be determined to be statistically different from a blank or background value at a specified level of confidence. In water quality laboratories, the confidence level of 95% is adopted as a standard for LOD for all analytes.
Littoral	Pertaining to the shore line; either as existing, taking place upon, or adjacent to the shore.
Maintainability	The probability that an item that has failed (in the reliability sense) can be restored (i.e., repaired or replaced) within a stated period of time.
Measures of central tendency	Measures of the tendency of values in a set of data to be centered at some location. Measures of central tendency are the median, the mode, the arithmetic mean, and the geometric mean.
Measures of dispersion (or of variability)	Measures of the differences, scatter, or variability of values of a set of numbers. Measures of dispersion or variability are the range, the standard deviation, the variance, and the coefficient of variation.
Median	The middle value of a set of data when the set is ranked in increasing or decreasing order.
Method	An assemblage of measurement techniques and the order in which they are used.
Method of averages	A mathematical procedure to adjust the parameters of a function so that the algebraic sum of the residuals is zero.
Method of least squares	A mathematical procedure to adjust the parameters of a function so that the sum of the squares of the deviations of the observed quantities from the respective calculated values, is a minimum.

Metrology	The study of "Weights and Measures", including such elements as gravimetric and volumetric measurements and analyses.
Milestone	A significant step or event in the conduct and control of a project.
Mode	The value or values occurring most frequently in a set or sample of data.
Network design	The activity involved in determining the placement of sampling points, the calculation of sampling frequencies, and the selection of water quality variables to be measured, all conducted in a hydrologically and statistically sound manner.
Normal distribution	A frequency distribution that is bell-shaped, symmetrical, and of infinite extent. It is mathematically described by an expression due to Karl Friedrich Gauss, and is often called the Gaussian distribution.
Null hypothesis	A statement made in support of a proposal or claim prior to the use of experimentation to test the claim.
Objective	This is a statement of intended output to be achieved in a time of 3 to 5 years. The objective should express in concrete terms, the quantity and quality of results expected and the period in which they are to be achieved.
Observation	The value noted of some variable or instrumental reading as the result of experimentation or of a natural occurrence. Ordinarily, this is a single data point.
OC Curve	Operating characteristic curve. A graphical depiction of the discriminatory power of a particular statistical test.
Outlier	A value which appears to deviate markedly from that for other members of the sample in which it occurs.
Output	The results of the work done in projects and processes.
Parameter	This denotes in water quality monitoring, any physical, chemical or biological entity that can be determined qualitatively and/or quantitatively. In other scientific fields

	(besides water pollution), the term 'variable' is used in place of 'parameter'.
Parameter	A characteristic or measure of a population, often symbolized by a Greek letter.
Pareto analysis	A statistical approach relating to ranking of assignable causes according to the frequency of occurrence.
Performance Indicator (PI)	This is a measure used to indicate the level of achievement or success within a given task or project. The PI could be either a target date or a measurement.
Performance audit	A process to evaluate the proficiency of an analyst/ laboratory by evaluating the results obtained on known test materials. One type of performance audit involves the use of a second (or dual) instrument independently maintained, calibrated, and operated at a given monitoring site or in a given laboratory by an auditor.
Performance Monitoring (PM)	Performance monitoring encompasses technical level activities that are performed routinely, while the total operational system is in service, to determine and verify the proper and satisfactory operation of a process, and to demonstrate and document the actual level of quality achieved. Typical activities include : field/laboratory blanks, duplicate samples, duplicate analyses, spike recovery, analysis of control/reference materials, etc.
Plan	A plan is a rational definition of actions to be taken in order to achieve a desired state of affairs.
Planning	This process involves : (a) forward thinking and directing the change of present activities; (b) setting objectives and selecting a course of action to achieve them; (c) establishing priorities for resource allocation; and, (d) assigning responsibility for results, and ensuring the evaluability of the efficiency and effectiveness of these results.
Poisson distribution	A model with a discrete distribution used to describe random phenomena in which the probability of occurrence is small but constant.

Pollution	The condition caused by the presence of substances of such character and in such quantities that the quality of the environment is impaired.
Population	A generic term denoting any finite or infinite collection of individual things, objects or events in the broadest concept.
Position description	A detailed statement of the requirements of a position. This is sometimes referred to as job description.
Position qualifications	Any quality, knowledge, ability, experience, or acquired attributes that fit a person for a particular position.
Power of a statistical test	This is the probability that the null hypothesis will be rejected when in reality it is false and should be rejected. Power is defined as 1-ß(i.e., as 1 minus the probability of making a Type II statistical error). The power of a statistical test depends on the level of significance that is used, the sample size, the true value of the population parameter being estimated or tested, and whether a one-tailed or a two-tailed test is used.
Practical detection limit	The practical detection limit is the lowest concentration of analyte in a real sample matrix that a method can detect reliably, and is statistically different from the response obtained from a blank carried through the complete method. It is calculated in the same manner as the method detection limit.
Precision	Denotes the agreement between the numerical values of two or more measurements on the same homogeneous sample made under the same conditions. The term is used to describe the reproducibility of the measurement or method. It can be expressed by the standard deviation.
Preservative	A substance added to a sample in order to maintain the original integrity of the sample.
Preventive maintenance	An orderly program of positive actions for preventing failure of equipment and ensuring, in so far as possible, that the equipment is operating with the reliability required for quality results.
Primary standard	A substance or artifact, the value of which can be accepted (within specific limits) without question when used to establish

the value of the same or related property of another material. Note that the primary standard for one user could be the secondary standard for another user.

Primary reference standard

A homogeneous material with specific properties, such as identity, purity, and potency, that has been measured and certified by a qualified and recognized organization.

Probability

The likelihood of the occurrence of any particular form of an event, and estimated as the ratio of the number of ways or times that the event may occur in that form to the total number of ways that it could occur in any form.

Probable error

A once popular way to express random error, but little used today. It is the value that defines the 50% portion of the error distribution in a normal distribution. Half the errors will be smaller and half larger than the probable error. It is calculated by :

$$P.E. = (0.6747)\,[(e_1^2 + e_2^2 + e_3^2 + e_4^2 + \ldots + e_n^2)/(n-1)]^{1/}$$

The probable error is thus $0.6747s_p$ or P.E. is approximately $2s_p/3$, where, s_p, is an approximation for sigma(σ), the population standard deviation.

Probability distribution

A distribution of events that indicates the probability of occurrence of each event.

Procedure

A set of systematic instructions for using a method of measurement or of sampling or of the steps or operations associated with such.

Process

A continuous or repetitive type of operation which has a measurable volume or output.

Proficiency examination

This refers to the measurement of an individual's competency to perform at a certain job level - a competency made up of knowledge and skills, and related to the requirements of the specific job. Such testing is , therefore, not only a measure of the knowledge gained through work experience, training, and didactic instruction but also an assessment of job capabilities.

Proficiency testing

A systematic testing program in which a sample, or a random series of samples is analyzed by a number of laboratories to determine the accuracy of an analyte or an analysis.

Program audit

Program audits evaluate the completeness of the quality management plan including the documentation and implementation of Quality Management strategies and the documentation of the Quality Assurance program.

Project

A specific task of a non-repetitive nature with an identified beginning and end.

Project quality

Project quality is demonstrated by evidence that :

(a) data quality needs have been defined and met;
(b) the correct samples were taken from the environment of interest;
(c) the samples taken represent the specific attributes to be studied;
(d) sample validity is not degraded;
(e) the impact of environmental variability on data interpretation has been accounted for;
(f) the impact of the field activity can be assessed;
(g) the impact of laboratory procedures, including sample storage and analysis, can be assessed;
(h) all sources of error have been minimized, and appropriate controls have been established to ensure this;
(i) data have been interpreted with due regard for the uncertainties introduced by all aspects of the investigation; and,
(j) the impact of environmental variability within field and laboratory activities does not preclude the attainment of project objectives.

Project quality planning

Project quality planning includes activities undertaken during project development to ensure that the project's quality needs will be met by the operational services available (field, laboratory, data management, or others), and that all critical facets of the project will be accompanied by appropriate QA/QC activity. Project quality planning requires multi-disciplinary cooperation to ensure that all personnel are aware of their roles.

Propagation of error	How the individual errors of measurement are propagated in a final computed result.
Protocol	A procedure specified to be used when performing a measurement or related operations, as a condition to obtain results that could be acceptable to the specifier.
Quality	This is an estimation of acceptability or suitability of an object, item, tangible or intangible thing for a given purpose. It involves the totality of features and characteristics of a product or service to satisfy a given purpose. For air and water pollution measurement systems, the product is pollution measurement data and the characteristics of major importance are accuracy, precision and completeness. The relative importance of accuracy, precision and completeness depends on the particular purpose of the user.
Quality assessment	This the overall system of activities whose purpose is to provide assurance that the quality control activities are being done effectively. It involves a continuing evaluation of performance of the production system and the quality of the products produced.
Quality assurance	This relates to the system for integrating the quality planning, quality control, quality assessment, and quality improvement efforts of various groups in an organization to enable operations to meet user requirements at an economical level. In air and water pollution measurement systems, quality assurance is concerned with all activities that have an important effect on the quality of the measurements, as well as the establishment of methods and techniques to assess the quality of the measurements. The more authoritative usages differentiate between quality assurance and quality control; quality control being the system of activities to provide a quality product, and quality assurance being the system of activities to provide assurance that the quality control system is performing adequately.
Quality assurance manual	This document is an orderly assemblage of management policies, objectives, principles, guidelines and general procedures by which an agency (or laboratory) outlines how it intends to produce quality data.

Quality assurance plan

This document is an orderly assemblage of detailed and specific procedures by which an agency (or laboratory) delineates how it produces quality data for a specific project or measurement method. A given agency or laboratory would have only one quality assurance manual, but would have a quality assurance plan for each of its projects or measurement methods, or for a family of similar methods.

Quality circle

This is a small group of individuals with related interests, and which meets at regular intervals to consider problems or other matters related to the quality of outputs of a process and to the correction of problems or to the improvement of quality.

Quality control

This is the overall system of activities whose purpose is to control the quality of a product or service so that it meets the needs of users. The aim is to provide quality that is satisfactory, adequate, dependable and economic.

INTERNAL - The routine activities and checks, such as periodic calibrations, duplicate analyses, and use of spiked samples, included in normal internal procedures to control the accuracy and precision of a measurement process.

EXTERNAL - The activities that are performed on an occasional basis, usually initiated and performed by persons outside of normal routine operations, such as on-site system surveys, independent performance audits and interlaboratory comparisons, to assess the capability and performance of a measurement process.

Quality control reference sample (working standard)

A material used to assess the performance of a measurement or portions thereof. It is intended primarily for routine 'intralaboratory' use in maintaining control of accuracy and would be prepared from, or traceable to a calibration standard.

Quality management

Quality management encompasses activities undertaken by Management in operational areas to establish and maintain accountability for the quality of operation. It will include assignment of responsibilities, definition of tasks, protocols and procedures, and audit mechanisms, to determine and maintain the desired level of product quality. In addition, it includes

	documenting the mechanisms for achieving product quality, and verifying that product quality is indeed being achieved.
Random error	This is an error due to variability causes. The net random error of a measurement process can cause either positive or negative errors, both small and large within limits. Random errors will average out to zero for a large number of independent measurements. The causes of random error are assumed to be indeterminate or non-assignable. The distribution of random errors is generally assumed to be normal (Gaussian).
Random sample	A sample that is obtained in such a manner that all items or members of the population have an equal chance of being selected in the sample. A randomization procedure is used in the selection process.
Range	The difference between the largest and the smallest values in a data set. Ordinarily, it is the absolute value and does not take into account the time sequence of the respective observations. In the case of duplicates, the order of observation may be considered in special cases, i.e. the difference of the second minus the first may be recorded to determine whether there is any significant difference due to order of measurement.
Reference Material (RM)	A material or substance which has one or more properties sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for the assignment of values to materials.
Reference method	A method which has been specified as capable, by virtue of recognized accuracy, of providing primary reference data.
Relative error	An error expressed as a percentage of the true value or accepted reference value. All statements of precision or accuracy should indicate clearly whether they are expressed in an absolute or relative sense. (This becomes complicated when the absolute value is itself a percentage, as is the case for many chemical analyses).
Relative Standard Deviation (RD)	This is the coefficient of variation (CV) expressed as a percentage. Thus :

$$RSD = 100 \left(\frac{s}{X}\right)$$

where, s is the standard deviation and X is the mean for the data values.

Reliability

The ability of an item or system to perform a required function under stated conditions for a stated period of time.

Reliable Detection Limit (RDL)

The Reliable Detection Limit is the recommended lowest level for qualitative decisions based on a single analysis, and it provides a much lower statistical probability of false negative determinations than the Method Detection Limit (MDL). It is suggested that the RDL be made equivalent to a concentration which is twice that of the MDL.

Reliable Quantitation Level (RQL)

The Reliable Quantitation Level is the lowest level for quantitative decisions based on a single analysis. It is suggested that the RQL be equivalent to a concentration which is four times the MDL.

Repeatability

This refers to the standard deviation of measurements made in close succession (i.e., the short term standard deviation). In ISO (Interna(Standards Organization) terminology, it represents $2x(2)^{1/2}$ times the standard deviation (s_a) which corresponds to the expected difference of two successive observations with 95% confidence. Some North American methods quote the reproducibility as 2 times this standard deviation(s_a).

Replicability

The precision, usually expressed as a standard deviation, measuring the variability among replicates. It can also be expressed as an upper confidence limit for the difference between two replicates.

Replicate

A counterpart of another, usually referring to an analytical sample or a measurement. It is the general case for which a duplicate is the special case consisting of two samples or measurements.

Replicate samples (spatial)

This refers to two or more samples taken simultaneously in a given cross-section of the sampling site under study.

Replicate samples (temporal)	This relates to two or more samples taken at the same place sequentially at specified intervals over a specific period of time.
Representative sample	This is a sample taken to represent a lot or population as accurately and precisely as possible. A representative sample may be either a completely random sample or a stratified sample, depending on the objective of the sampling and the conceptual population for a given situation.
Reproducibility	The standard deviation of observations made over a protracted period of time (the long-term standard deviation). In ISO (International Standards Organization) terminology, it is $2x(2)^{1/2}$ times this standard deviation (s_b). Some North American methods quote the reproducibility as 2 times this standard deviation(s_b).
Responsibility centre	An organizational unit in which a manager has the delegated authority to manage financial and non-financial resources.
Routine method	A method used in recurring analytical problems.
Ruggedness tests	These relate to a series of tests performed to empirically determine the sensitivity of the measurement system to variations of certain factors that are suspected to affect the method.
Safety	This is a quality of being devoid of whatever exposes one to danger or harm. It represents freedom from danger or hazards.
Sample	This refers to a portion of a population. It may consist of an individual or groups of individuals. It may refer to objects, materials, or measurements, conceivable as part of a larger group that could have been considered.
Secondary standard	A standard whose value is based upon comparison with some primary standard. Note that a secondary standard, once its value is established, can become a primary standard for some other user.
Segment	A specifically demarked portion of a lot or site, either actual or hypothetical.

Selectivity	This is the ability of a methodology or of instrumentation to respond to a desired substance or constituent and not to others.
Sensitivity	This is the ability of an analytical method to detect small quantities of the measured component. There is no numerical value associated with this definition. An alternative definition describes the sensitivity as the change in measured value resulting from a concentration change of one unit.
Sequential composite sample	A sample obtained either by continuous, constant pumping of water or by mixing equal volumes of water collected at regular time intervals. This sample will indicate an average water quality condition over the period of time of compositing.
Sequential triplicate sampling	This is an activity in which the collection of three sequential samples from the same location, is designed to assess the efficiency and effectiveness of the overall water quality monitoring program. The operation serves to evaluate whether the water sampling performed is representative; if it helps in discerning contamination and/or analytical problems; and, reveals problems associated with water quality data management. On the whole, the sequential triplicate sampling program covers the entire quality assurance picture by providing some information on the combination of field, laboratory and data management components. Sequential triplicate sampling may not define the exact source(s) of a problem, but can be used as an aid in determining the cause(s) of the problem.
Simple statistical control	Simple statistical control is observed in all properly defined systems. Replicate analyses which do not follow the expected distribution, i.e., too many values in the tail of the distribution, provide evidence of lack of statistical control. This is often due to inadequate definition of the procedure or the quality of components used in the procedure. Ordinary care in following the defined procedures will be sufficient to maintain quality.
Special cause	A cause of variance or bias that is external (not inherent) to the measurement system.

Spiked sample

This is a normal sample of material (gas, liquid, or solid) to which is added a known amount of some substance of interest. The extent of the spiking is unknown to those analyzing the sample. Spiked samples are used to check the performance of a routine analysis or the recovery efficiency of a method.

Split sample

A replicate portion or sub-sample of a total sample obtained in such a manner that it is not believed to differ significantly from other portions of the same sample (mother sample).

Standard

A substance or material, the properties of which are believed to be known with sufficient accuracy to permit its use to evaluate the same property of another. In chemical measurements, it often describes a solution or substance, commonly prepared by the analyst to establish a calibration curve or the analytical response function of an instrument.

Standard

That which is established by authority, custom, or general consent : as a rule for the measure or judgement of quantity, value or quality; as a model or example; as an acceptable way to conduct a measurement or a test.

Standard addition

A method in which small increments of a substance under measurement are added to a sample under test to establish a response function or, by extrapolation, to determine the amount of a constituent originally present in the test sample.

Standard deviation

This is a measurement of the dispersion or spread of data points around the mean value of the data set obtained by repetitive testing of a homogeneous sample under specified conditions. It is calculated from the square root of the 'variance' of a set of values.

Standard deviation of the mean

This term is also alluded to as the 'standard error of the mean'. It represents the standard deviation of means of size, n, about the grand mean, i.e., the grand average of all means of that size. It is calculated by:

$$\text{S.D. of mean} = (\sigma \text{ or } s)/(n)^{1/2}$$

Standard error of the mean

See 'standard deviation of the mean'.

Standardization	The process whereby the value of a potential standard is fixed by measurement with respect to a standard(s) of known value.
Standard method	A method or procedure of testing developed by a standards-writing organization, based on consensus opinion or other criteria, and often evaluated for its reliability by a collaborative testing procedure.
Standard Operations Procedure (SOP)	A procedure adopted for repetitive use when performing a specific measurement or sampling operation. It may be a standard method or one developed by the user.
Standard Reference Material (SRM)	A material produced in quantity , of which certain properties have been certified by the National Bureau of Standards (now called the National Institute of Standards and Technology), the National Research Council of Canada (NRCC), or other agencies to the extent possible to satisfy its intended use. The material should be in a matrix similar to actual samples to be measured by a measurement system, or should be used directly in preparing such a matrix. Intended uses include: (a) standardization of solutions, (b) calibration of equipment, and (c) monitoring of the accuracy and precision of measurement systems.
Standard Reference Sample (SRS)	A carefully prepared material produced from or compared against an SRM (or other equally well characterized material) such that there is little loss of accuracy. The sample should have a matrix similar to actual samples used in the measurement system. These samples are intended for use primarily as reference standards (a) to determine the precision and accuracy of measurement systems, (b) to evaluate calibration standards, and (c) to evaluate quality control reference samples. They may be used 'as is' or as a component of a calibration or quality control measurement system.
Static calibration	Artificial generation of the response curve of an instrument or method by use of appropriate mechanical, optical, electrical, or chemical means. Often a static calibration checks only a portion of a measurement system. For example, a solution containing a known amount of an inorganic nitrite compound would simulate an absorbing solution through which has been bubbled a gas containing a known amount of nitrogen trioxide (N_2O_3). Use of the solution would check the analytical portion

	procedure involving the sulphanilamide reagent (to give the diazo derivative) coupled with the addition of N-(1-naphthyl) ethylenediamine dihydrochloride, but would not check the sample handling and the flow-control parts of the continuous-flow apparatus.
Statistic	A measure computed from observations in a sample, usually designated by a Latin letter.
Statistical control	A measurement system is in a state of statistical control when the process is stable. Although each individual measurement can be expected to differ somewhat from each other during repeat measurements (determination of precision), the pattern of deviations must be stable (i.e., a constant dispersion) and the averages of a large number of measurements should zero in on a particular value, called the limiting mean. The individual values will of course be randomly distributed around this limiting mean, if a large number of measurements are considered. Only in a state of statistical control is a statistical treatment meaningful.
Statistical control chart (Shewart Control Chart)	A graphical chart with statistical control limits and plotted values (usually in chronological order) of some measured parameter for a series of samples. The chart provides a visual display of the pattern of the data, enabling early detection of time trends and shifts in level. For maximum usefulness in control, such charts should be plotted in a timely manner, i.e., as soon as the data are available.
Statistical control chart limits	These are the limits on statistical control charts that have been derived by statistical analysis and are used as criteria for action, or for judging whether a set of data does or does not indicate lack of statistical control.
Statistical hypothesis	A tentative statement about one or more parameters of a population. Null and alternative hypotheses are two type of statistical hypotheses.
Statistics	A branch of mathematics concerned with the collection, analysis, presentation, and interpretation of data.
Strata	Segments of a lot or population that may vary with respect to the property under study.

Strategic planning	The process of deciding on the objectives of the organization and the main strategies for achieving these objectives.
Stratified sampling	A sampling procedure in which the population is divided into sub-populations, called strata, and then elements are randomly selected from each strata.
Sub-sample	A portion taken from a sample. A laboratory sample may be a sub-sample of a gross sample; similarly, a test portion may be a sub-sample of a laboratory sample.
Systematic errors	Errors which can, at least in principle, be ascribed to definite causes and which contribute a constant error or bias to results. Systematic errors never average out and the magnitude is independent of the number of repetitions of the measurement process.
Systematic sampling	A sampling procedure in which a consistent method is used, such as selecting every k^{th} element or one element from every k^{th} time period.
System audit	An on site-inspection or assessment of a laboratory or field's quality control system. More generally, system audits evaluate the day-to-day implementation of Quality Assurance and Quality Control protocols and procedures.
Target population	The population to which inferences are to be made.
Technique (analytical)	A physical or chemical principle utilized separately or in combination with other techniques to determine the composition (analysis) of materials.
Test portion	This is sometimes called a specimen , test specimen, unit or aliquot and is that quantity of a material of proper size for measurement of the property of interest. Test portions may be taken from the sample directly, but often preliminary operations, such as mixing or further reduction in particle size, are necessary.
Test-retest form of reliability	A method to determine the reliability of a procedure or an instrument by taking two measures of the same objects or things using the same procedure or instrument within a reasonable time period.

Test statistic	The statistical method selected to test the null hypothesis.
Tolerance interval (engineering)	The engineering tolerance interval represents limits of acceptability of some characteristic of a population, usually prescribed by a design engineer.
Tolerance interval (statistical)	That range of values, calculated from an estimate of the mean and the standard deviation, within which a specified percentage of individual values of population (measurements or sample) are expected to lie with a stated level of confidence.
Trace	A concentration that is below the detection limit of the standard analytical method used.
Traceability	The ability to trace the source of uncertainty of a measurement or a measured value.
Training	Formal or informal instruction designed to provide competence of a specific nature.
Trend analysis	An a priori test to determine if there is a relationship between the independent and dependent variables and if there is, what type of trend there is.
Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. The measure of trueness is usually expressed in terms of bias.
Two-tailed test	A statistical test of the nondirectional hypothesis in which the critical region lies in both tails (ends) of the sampling distribution.
Type I error	A decision made to reject a null hypothesis when in actuality it is true and should not have been rejected. This is also called alpha (α) error or false positive.
Type II error	A decision made not to reject a null hypothesis when in actuality it is false and should have been rejected. This is also called beta (ß) error or false negative.
Type III error	A decision to correctly reject a false null hypothesis but incorrectly interpreting the direction of the difference.

Uncertainty

The range of values within which the true value is estimated to lie. It is a best estimate of possible inaccuracy due to both random and systematic errors.

Universe

A term used interchangeably with population.

Validation

The process by which a sample, measurement method, or a piece of data is deemed to be useful for a specified purpose.

Variable data

Data obtained as the result of measurements of a variable.

Variance

Mathematically, this is the sum of the squares of the difference between the individual values of a set and the arithmetic mean of the set, divided by one less than the number of the values.

Alternatively, the variance can be defined as the value approached by the average of the sum of the squares of deviations of individual measurements from the limiting mean. Algebraically, it may be expressed as:

$$\sigma^2 = \frac{\sum (X_i - m)^2}{n}$$

as n becomes very large (n ------ ∞).

Ordinarily the population variance is not known but only its estimate, s^2, which is calculated by the use of the expression :

$$s^2 = \frac{\sum (X_i - X)^2}{n - 1}$$

Warning limits

The limits shown on a control chart within which most of the test results are expected to lie (within a 95% probability) while the system remains in a state of statistical control.

Water quality criteria

Scientific data evaluated to derive the recommended limits for water uses (EPA, USA).

Water quality guideline

A numerical concentration or narrative statement recommended to support and maintain a designated water use (Environment Canada).

Water quality objective	A numerical concentration or narrative statement which has been established to support and protect designated uses at a specified site (Environment Canada).
Water quality standard	An objective that is recognized in enforceable environmental control laws that are established by a level of government.
Weighting	The weight of an observation is its relative value among others of the same set. It represents the relative confidence in that observation in comparison with other observations of the same quantity. In the context of the definition, an observation may be an individual value or the mean of a set of values.
Weight of evidence	An approach based on "Weight of Evidence" looks at all information gathered by acceptable scientific methods and scientifically based approaches. All interpretation or conclusions are based on scientific evidence and good professional judgement.
Within-laboratory precision	The variability between replicate results obtained on the same material within a single laboratory (intralaboratory analyses). It is sometimes referred to as "repeatability".
Working standard (secondary standard, house standard)	A material having a property that has been calibrated against a primary reference standard.
Youden plot	A graphical presentation of data, recommended first by W.J. YOUDEN, in which the result(s) obtained by a laboratory on one sample is plotted with respect to the result(s) it obtained on a similar sample. Depending on the relation of the plotted point to the "true value", it can be decided whether discrepant results are due to random errors (imprecision), systematic errors (bias) or both causes.

ACKNOWLEDGEMENT

The preceding definitions were taken as either modified or unmodified extracts from various sources. These sources include:

Gaskin, J.E. 1988. Quality Assurance Principles and Guidelines for the Handling and Management of Water Quality Data. Water Quality Branch, Inland Waters Directorate, Conservation and Protection, Environment Canada, Ottawa.

Gaskin, J.E. 1988. Quality Assurance Principles and Guidelines for Water Quality Sampling. Water Quality Branch, Inland Waters Directorate, Conservation and Protection, Environment Canada, Ottawa.

Gaskin, J.E. 1989. Quality Assurance Management Plan. Water Quality Branch, Inland Waters Directorate, Conservation and Protection, Environment Canada, Ottawa.

Gaskin, J.E. 1990. Responsibility and Accountability. Water Quality Branch, Inland Waters Directorate, Conservation and Protection, Environment Canada, Ottawa.

Inhorn, S.L. 1978. Quality Assurance Practices for Health Laboratories. American Public Health Association, Washington, D.C.

Keith, L.H. 1991. Environmental Sampling and Analysis – A Practical Guide. Lewis Publishers, Boca Raton, Florida.

Ontario Ministry of the Environment. 1985. Quality Management Plan. Laboratory Services Branch.

Taylor, J.K. 1986. A Collection of Abstracts of Selected Publications Related to Quality Assurance of Chemical Measurements. U.S. Department of Commerce, National Bureau of Standards, Gaithersburg, Maryland.