UN environment programme

QUALITY ASSURANCE FOR FRESHWATER QUALITY MONITORING

Technical Guidance Document

Prepared by UNEP GEMS/Water Capacity Development Centre for the
United Nations Environment Programme

ISBN No: 978-92-807-4031-8

Job number: DEW/2529/NA

Disclaimers

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory or city or area or its authorities, or concerning the delimitation of its frontiers or boundaries. For general guidance on matters relating to the use of maps in publications please go to http://www.un.org/Depts/Cartographic/english/htmain.htm

Mention of a commercial company or product in this document does not imply endorsement by the United Nations Environment Programme or the authors. The use of information from this document for publicity or advertising is not permitted. Trademark names and symbols are used in an editorial fashion with no intention on infringement of trademark or copyright laws.

The views expressed in this publication are those of the authors and do not necessarily reflect the views of the United Nations Environment Programme. We regret any errors or omissions that may have been unwittingly made.

Cover photo © Top left – Deborah Chapman; Middle left – Patrick Cross; Bottom left – Environmental Research Institute (ERI), UCC; Bottom middle – Patrick Cross; Bottom right – Deborah Chapman

Suggested citation: United Nations Environment Programme (2023). *Quality Assurance for Freshwater Quality Monitoring - Technical Guidance Document*. Nairobi

https://doi.org/10.59117/20.500.11822/42666

GEMS Water

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

ACKNOWLEDGEMENTS

This technical guidance document has been compiled by the UNEP GEMS/Water Capacity Development Centre with contributions from GEMS/Water at Environment Canada, Greg Beechinor, Deborah V. Chapman, Linda Doran, Lucia Hermida Gonzalez, Dmytro Lisniak, Stuart Warner and Peter Webster. We are grateful to Peter Webster for his extensive contributions and for reviewing the draft document and providing helpful advice for its improvement.

FOREWORD

This technical guidance document is intended for scientists and practitioners who work with freshwaters in the field and in the laboratory but who do not have specialist knowledge of quality assurance. It provides an introduction to the key concepts and approaches that can be used in Quality Assurance and Quality Control. Without appropriate and adequate quality assurance measures at each step, from taking a water sample to storing and sharing the data, the outputs of the monitoring activities may not reflect, as fully as they should, the true quality condition of the water body. Although the main focus is on freshwaters, the general principles outlined are applicable to most types of aquatic monitoring including drinking water, wastewater, and industrial effluents.

This document is part of a series of guidance documents that address various aspects of monitoring and assessment of freshwater. It is recommended that the other guidebooks in the series are consulted for more detail on related topics. Guidance documents in the series being released in 2022/23 include:

- Introduction to Freshwater Quality Monitoring and Assessment
- Water Quality Monitoring and Assessment in Rivers, Lakes and Reservoirs
- Water Quality Monitoring and Assessment of Groundwater
- Freshwater Quality Monitoring with Biota
- Freshwater Quality Monitoring using Particulate Matter
- Water Quality Data Handling and Assessment

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AOAC	Association of Official Analytical Collaboration
APHA	American Public Health Association
AQC	Analytical Quality Control
AWWA	American Water Works Association
BOD	Biochemical Oxygen Demand
BSI	British Standards Institute
CEN	European Committee for Standardization
cfu	colony forming units
CoA	Certificate of Analysis
COD	Chemical Oxygen Demand
CRM	Certified Reference Material
DL	Detection Limit
DQOs	Data Quality Objectives
EPA	Environmental Protection Agency
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPS	Global Positioning System
HPLC	High Performance Liquid Chromatography
ICP	Inductively Coupled Plasma (mass spectrometry instrument)
ILAC	International Laboratory Accreditation Cooperation
ISO	International Organization for Standardization
LIMS	Laboratory Information Management System
LoD	Limit of Detection
LoQ	Limit of Quantitation
MDL	Method Detection Limit
MPN	Most Probable Number
MSDS	Material Safety Data Sheet
NIWA	New Zealand National Institute of Water and Atmospheric Research
QA	Quality Assurance
QAP	Quality Assurance Plan

QC	Quality Control
QM	Quality Manager
QMS	Quality Management System
RCA	Root Cause Analysis
ROS	Regression on Ordered Statistics
SDG	Sustainable Development Goal
SOPs	Standard Operating Procedures
TDS	Total Dissolved Solids
US EPA	United States Environmental Protection Agency
WEF	Water Environment Federation

CHAPTER 1

INTRODUCTION TO QUALITY ASSURANCE FOR FRESHWATER MONITORING

Time, effort, and resources put into the collection and water analysis of water samples are all wasted if the data produced by the monitoring activities are not credible and defensible. Quality assurance contributes to ensuring that these criteria are met. Data that will be shared with other laboratories, organizations or databases must also be comparable and compatible. Often the body or organization with responsibility for carrying out the monitoring programme will not necessarily have control over the use of their data after they are made accessible to third parties. It is vital, therefore, that these expectations are considered during the development of monitoring programmes before they are implemented. The application of Data Quality Objectives (DQOs), Quality Assurance (QA) principles, and relevant Quality Control (QC) process tools, helps to ensure these requirements can be met. In the real world, it is quite common to find that monitoring programmes commence without sufficient thought or attention having been given to ensuring the data generated are of high quality and fit for purpose. Similarly, it is not unusual for the design of programmes to evolve over time in response to changes in the natural environment or, more often, to changes in the availability of resources, particularly funding!

Data collected through the quality assurance chain may also be used to support interdisciplinary research in other related sectors. For example, it is recommended that sex-disaggregated and, where possible, age-disaggregated data are recorded. Indicating the sex of the collector as well as others involved in the chain provides useful information. Tracking gender data in water quality management provides accurate evidence-based records which provide useful information, not only towards the formulation of gender-responsive policies and strategies, but also crucial towards establishing baselines and monitoring gender parity, thereby supporting progress towards the achievement of Sustainable Development Goal (SDG) 5.

This chapter introduces some important definitions and terms, and highlights the importance of DQOs in the quality assurance process. The concept of a Quality Assurance Plan (QAP) for a monitoring programme is briefly introduced. More in-depth information on the various topics covered in this guidebook is available in the cited references and in the other guidebooks in this series.

1.1 Definition of key terms

Quality can be defined as fitness for use, conformance to requirements, and the pursuit of excellence

A *Quality Management System* (QMS) includes all activities that determine the quality policy, objectives and responsibilities, together with their implementation. The QMS documents all relevant operational details, management activities and responsibilities. Its development and maintenance would generally be the function of a dedicated *Quality Manager* (QM). Quality assurance and quality control are two key aspects of quality management.

Quality Assurance comprises all the planned and systematic activities implemented within the quality system that can be demonstrated to provide confidence that a product or service will meet all the requirements for quality.

Quality Control describes the operational techniques and activities used to fulfil the requirements for quality. Quality assurance is process oriented and focuses on defect prevention, while quality control is product oriented and focuses on defect identification and remediation. In practice, the difference is that QA sets the rules and standards to achieve product quality, while QC inspects and tests the product against those predetermined rules and standards.

Precision and *Accuracy* are important aspects in the quality assurance of monitoring measurements, in the field and in the laboratory. *Precision* is the likelihood of the method giving the same value if the same sample is analysed more than once. Precision can be expressed as the standard deviation. Less precision is reflected by a larger standard deviation. *Accuracy* is the nearness of the measured value to the true value.

Data Quality Objectives are the criteria that define the quality of data to be managed.

1.2 Data Quality Objectives

A key element of planning any monitoring programme is the need for DQOs to be established prior to the commencement of monitoring. The DQO process should, therefore, be used during the planning stage of any monitoring programme that requires data collection. It is a means of developing consensus amongst stakeholders, and it is important where data collection will require substantial resources. After data collection, the DQO process should guide the use of data in decision-making. The process comprises seven key steps as described below and illustrated with a hypothetical example.

Step 1. State the problem, i.e., what it is necessary to monitor

Example "How is the water quality in the catchment of interest changing over time, given that point and non-point sources in the river basin are continuing to release pollutants into the watercourses"?

Step 2. Identify the goal(s) of the study, i.e., the objective(s) and expected output(s)

Example The National Environmental Agency proposes to monitor key water quality variables at a number of specified locations and to compare the monitoring results with available historic data and national targets. If the data comparison shows a statistically significant deterioration in water quality, the Agency will increase efforts to identify the sources of pollution and to develop initiatives to reduce pollutant loadings.

Step 3. Identify the data and information needed to answer the monitoring programme question(s)

This step should include an assessment of all relevant analytical parameters (both field and laboratory measurements) required to provide the necessary information to answer the objectives of Step 1.

Step 4. Identify the boundaries of the study

The characteristics of interest are specified, and the spatial and temporal limits and the scale of inference are defined.

Example The spatial scope of the proposed programme encompasses the main rivers, their tributaries, and any associated lakes. Sampling locations were selected during the monitoring programme design phase. It is anticipated that approximately "*n*" sampling locations will be included as part of this programme. The temporal boundaries for the programme are three full successive years. The climate is seasonal with a defined drier summer period and wetter winter period. Monitoring will be undertaken on a fortnightly basis. Summer climatic conditions and river flow may cause severe stress on the lake system.

Although this example is hypothetical, it is typical of conditions often found in many European and African countries where summer droughts are common. The effects of such seasonality must be factored into any monitoring programmes. Flow monitoring is important

in the river system in order to determine overall loading rather than simply concentration. Such an approach is more informative, especially when large flow variations are exhibited by river systems.

Step 5. Define the analytical approach

In this step, the variables of interest need to be identified and the type of inference specified. This enables the logic for drawing conclusions from the findings to be developed.

Example Water quality data collected as part of this programme will be compared with available historical data from within the catchment. They will also be compared with relevant national and international water quality criteria to assess and/or characterize the overall condition of water quality. Should the data show a meaningful deterioration in water quality from the earlier time period or, for example, where water quality targets for nutrients or other chemical variables are not being met, the National Environmental Agency will increase its efforts to identify the most significant sources of pollution and develop initiatives to reduce pollutant loadings.

One of the key issues here is the choice of applying a statistical significance of a change in the decision-making. To apply such a rationale, it is generally recommended that a minimum of 20 samples are available and that any seasonal variability is considered. For this reason, a three-year period would be more desirable than perhaps a more intensive one-year study to account for year on year and seasonal impacts.

Step 6. Develop performance criteria for new data or acceptable criteria for existing data being considered for use

The purpose of this step is to ensure, as far as possible, comparability of both new data (resulting from the application of this monitoring programme) and any historic data against which they will be compared. In practice, mismatched data are often found where the methodology used has been insufficiently sensitive to meet the current DQOs or where data have been reported in inconsistent units, e.g., mg l^{-1} as NO_3 rather than as N.

Example Standard Operating Procedures (SOPs) in support of sampling and analysis will be detailed in the project QAP. Field sampling personnel will be trained and appropriately qualified prior to project commencement. Internationally recognised procedures will be applied to sample collection, preservation, holding times and environmental factors and only laboratory services accredited to ISO 17025 (International Organization for Standardization [ISO] 2017) will be used for sample analysis. Field and laboratory activities will be assessed throughout the term of the project using suitable internal and external quality assurance methodologies. At least 5 per cent of the samples collected and analysed will be quality control samples and acceptance criteria for these QC samples will be detailed in the QAP. These QA criteria will also be used to assess the acceptability of the data from the earlier period of sampling.

Step 7. Develop the sampling plan

The sampling plan should ensure that resources are efficiently and effectively applied and that sampling and analysis criteria are met. The completeness of the monitoring programme (in terms of number of samples collected) should be assessed on an ongoing basis and any deficiencies should be corrected. Sampling locations should be chosen to ensure they are as representative of the waters being studied as possible, taking into account access and safety issues. The use of Global Positioning System (GPS) coordinates and photographs of monitoring locations can ensure that the correct stations are monitored. Particular attention needs to be paid to the regularity of the monitoring programme to ensure statistical validity.

The selected design option that results from the DQO process, should be incorporated into the QAP. Adequate oversight of sampling and analysis activities will help ensure that the programme design is implemented as intended.

1.3 The Quality Assurance Plan

A QAP is a document that is generally produced by the monitoring programme team and which is intended to ensure that the data generated from a monitoring

programme are of the highest quality possible. Depending on the context, a QAP describes the activities of an environmental data acquisition project, whether the information is generated from direct measurement activities, collected from other sources, or compiled from computer databases and similar information systems.

The QAP integrates all the technical and quality aspects of the monitoring programme, including planning, implementation and assessment. Its purpose is to document planning results for the water quality monitoring programme and to provide a programme-specific "blueprint" for obtaining the type and quality of data needed for a specific decision or use. The QAP documents how QA and QC are applied to the monitoring programme to assure that the results obtained are of the type and quality required. A QAP is prepared either as part of, or after, the monitoring programme planning process. Once the DQOs have been defined, the next important part of creating a QAP is to define the roles and responsibilities of team members. In all cases, the QAP should be completed and approved by all stakeholders before monitoring is started.

The key benefit of having a QAP is that it communicates to all stakeholders the specifications for implementation of the monitoring programme design and provides greater assurance that the quality objectives for the programme are more likely to be achieved. It does not necessarily guarantee that all the monitoring programme objectives will be met in every case (as unexpected problems can arise at any stage), but the likelihood of these being met is much higher with a QAP than without one. Planning in advance helps to eliminate approaches that will not work well (or not at all), and this can potentially reduce the cost associated with wasted time and having to repeat activities. Implementation as prescribed in a QAP, using the appropriate QC practices, increases efficiency and provides for early detection of problems in the field and in the laboratory. The QAPs can therefore save time and money by enabling decisions to be made in a timely and efficient manner.

The key elements of a QAP are listed in **Box 1.1** and described in more detail below. The United States Environmental Protection Agency (US EPA) provides useful guidance for developing QAPs in United States Environmental Protection Agency [US EPA] (2006) and US EPA (2002).

BOX 1.1 THE MAIN ELEMENTS OF A QUALITY ASSURANCE PLAN

- Problem definition / background information
- Project / task description
- Quality objectives and criteria
- Special training /certification
- Documents and records management
- Monitoring experimental design
- Field sampling methods
- Sample handling and custody
- Analytical methods (suitability and performance criteria)
- Quality control procedures
- Key equipment
- Assessment and response actions
- Data review, verification and validation

The *problem statement* explains the background of the monitoring programme and the reasons for initiating monitoring. It also includes uses and/or designated uses and impairment of the water resource, if applicable. The intended usage of data must be stated together with the outcomes expected from the information that will be collected (e.g., to show that best management practice is effective, watershed characterisation, collection of background data, environmental education, etc.). The type of data to be collected must also be described (e.g., screening, definitive, characterisation, baseline or background). If applicable, the technical or regulatory standards or criteria with which data will be compared should also be cited.

The *project or task description* should summarise the work to be performed, define the geographical, spatial and temporal boundaries of the study, and briefly describe the monitoring design and how

monitoring data will assist in achieving the monitoring programme objectives.

The *quality objectives and criteria* identify the performance and measurement criteria for all the information to be collected, the acceptance criteria, including monitoring programme action limits and laboratory detection limits, and the range of anticipated concentrations of each variable of interest (including field and laboratory). These include:

- Data precision, accuracy and measurement range. The degree to which sample results are repeatable should be defined, together with decision error limits, if applicable. Note that monitoring programmes that are based on authoritative, rather than statistical, sampling design will not have quantitative decision error limits.

- Data representativeness. The degree to which the data accurately represent the environmental condition at the sampling location should be outlined, i.e., how well the monitoring characterises the physical condition should be reviewed and assessed where practicable.

- Data comparability. The degree of confidence that one data set can be compared with another at the sample location, or with a sample taken at another location, should be defined.

- Data completeness. This is a measure of the amount of valid data required in order to develop conclusions, i.e., estimate how many measurements are needed to meet each monitoring objective. It reflects the actual number, and extent of sampling, compared with what was anticipated to be sampled or measured. Gaps in monitoring, especially if associated with extreme environmental conditions, may have a marked influence on the data assessment.

A general description of *training requirements* and needs should be included in the QAP. This should include, where applicable, a description of special personnel or equipment required. Barriers and challenges faced should be mentioned. Closely scrutinizing and analyzing the differentiated challenges that women and men face in the field may help advance gender equality in water quality management.

All data reporting information should be identified and all *programme documents*, reports and electronic files that will be produced should be listed. Quality assurance records and reports should be included with a list of the information and records to be included in the data reports, e.g., laboratory and field raw data, field logs, laboratory records, results of QC checks, problems encountered.

The *monitoring design strategy* should indicate the size of the area, volume or time period to be represented by the monitoring, i.e., detail the type and total number of sample types/matrix or test runs/ trials expected and needed. It should also specify any monitoring of covariates that will be required such as rainfall and discharge. More specifically these include:

- Rationale or criteria for selection of sampling sites. The monitoring design strategy should be described and justified, indicating the size of the watershed area, discharge volume, or time period to be represented by the monitoring. If applicable, appropriate validation study information for non-standard sampling situations should be described.

- Project monitoring locations and watershed boundaries. A map should be included that delineates the watershed boundaries or the drainage area being monitored. The geographic locations of sample locations (including GPS coordinates) need to be identified on a map or in a table. If other data sources are to be obtained and compiled, the sources should also be listed.

- Sample design logistics. This includes sample numbers and sampling frequency, including for any covariates such as rainfall and discharge. It should be clearly stated whether the variable is required for informational purposes only or whether it is essential.

- Health and Safety considerations. A key characteristic of the selection of monitoring stations must be the safety of such locations. A full Site Risk Assessment should be undertaken for each monitoring location, which examines factors

such as access routes, vehicle and equipment security, and staff safety (especially for staff working alone and in relation to specific safety and other concerns that women might have). Some locations may have specific requirements, such as boats and depth sampling equipment.

The sampling equipment and *sample collection methods* should be clearly set out, and the SOPs to be applied should be included. Procedures should be described in sufficient detail for preparation and decontamination of sampling equipment, for collection of monitoring samples, and for sample preservation methods (as required). The following information should be included:

- Sampling equipment (including selection and preparation of sample containers)
- Sample collection methods (including volumes to be sampled)
- Field data collection equipment (including field kits and sensors)
- Standard operating procedures (including references for additional detail)
- Standard references

The individuals responsible for any corrective action should also be identified.

Sample handling and custody must be clearly described and documented to specify how samples will be physically handled, transported and received. Documentation for the handling of sample information and chain-of-custody should be described. The maximum allowed holding time from collection to analysis, including any relevant laboratory preservation procedures, must also be included. Information must also be provided on sample archiving, storage and retrieval, and the staff responsible for this. An example of the sub-sections to be included in this section of the QAP are:

1. Sample supplies (e.g., bottles)
2. Physical handling
3. Transportation
4. Reception at laboratory
5. Hold times prior to analysis
6. Documentation
7. Sample archiving, storage and retrieval

For all the *analytical methods* defined in the monitoring programme, a variety of international reference sources exist. Using an accredited laboratory will ensure that the analytical methodology applied will be suitable for its intended purpose and appropriate for the sample matrix (e.g., water, wastewater, sediment). The performance criteria, including the limit of detection, sensitivity and linearity (see Chapter 10) will have been determined by the accredited laboratory and any known interferences will have been identified. Analytical methods should be capable of achieving at least one tenth of the concentration of interest (if possible) and no less than one third of the relevant concentration (see example in **Box 1.2)**.

Analytical Quality Control (AQC) procedures are fundamental to ensuring that data quality is not compromised by factors such as analytical bias, inaccuracy or poor precision. At its simplest, AQC involves the use of a series of control charts (see Chapter 5) to assess the ongoing performance of each batch of analyses. It can use duplicates of real samples, surrogate controls such as a bottled mineral

BOX 1.2 EXAMPLE OF SELECTING AN ANALYTICAL METHOD

If the national target for ortho-P is 50 µg l^{-1}, then the test method applied should ideally be capable of achieving a limit of detection (LoD) of at least 5 µg l^{-1} P and no more than 16 µg l^{-1}. In practice 10 µg l^{-1} may be the most practicable LoD, but all efforts should be taken to improve this where possible.

For some analytical methods, such as chromatographic assays for pesticides, it may not be possible to achieve the above criteria. In such cases, the best available methodology not exceeding excessive cost (BATNEEC) should be applied.

water, or synthetic standards of the determinand. Whatever the source material, compliance with predefined limit values is necessary before data can be considered to be of suitable quality.

Equipment that is central to the monitoring programme should be specified. Such equipment should be maintained in accordance with the manufacturer's service schedule and only by trained personnel. Records should be kept of any changes made which may affect the instrumental performance.

Where ongoing quality checks identify any irregularities or deficiencies, there should be an agreed programme for *follow-up action*, such as to withdraw or replace defective equipment, to reanalyze samples where AQC results are out of specification, or similar corrective actions. Responsibilities for the assessment and verification of corrective actions need to be set out in the QAP.

Before any data are released for inclusion in the overall dataset their quality should be *reviewed* by a suitable competent person. In practice, this will generally be the laboratory manager or quality manager.

CHAPTER 2

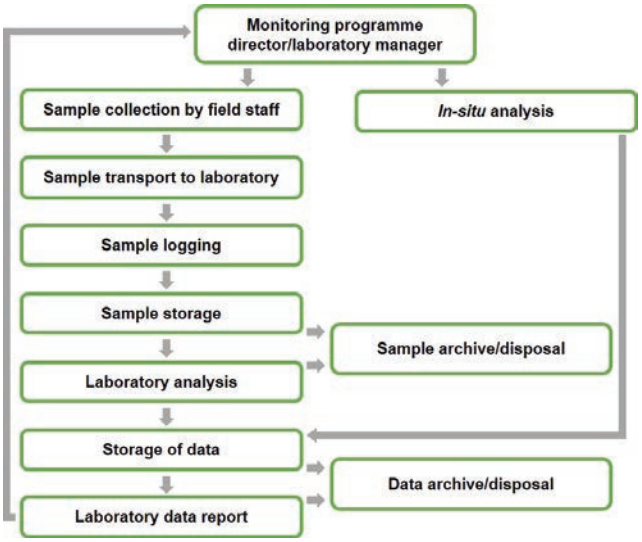
SAMPLE LIFE-CYCLE

This chapter provides an overview of the key steps involved in the sampling and analytical process and focuses on some of the issues associated with field sampling and data quality and suggests approaches for dealing with those issues.

2.1 Defining the Quality Management System

A typical workflow for sampling and analysis in a water quality monitoring programme is shown in **Fig. 2.1**. The whole cycle of sampling, analysis, data storage and reporting should be managed within an overall Quality Management System (QMS). The QMS embodies the Data Quality Objectives described in the previous chapter. In the context of this guidebook, the term "Management System" refers to the quality, administrative and technical systems that govern the operations of a water quality monitoring laboratory. The management system covers policy, staff, facilities, equipment, procedures, documentation, records and quality control. The relevant International Organization for Standardization (ISO) standard for this is ISO 17025:2017 (ISO 2017). The collection of sex-disaggregated data on the staff in the QMS is also recommended. This provides a legitimate source of data to help track progress towards gender equality in water quality management.

Figure 2.1 The elements of a Quality Management System. (adapted from ISO 2017)



In many situations, the management system and monitoring programme will be under the control of a central laboratory. However, in the case of multiple facilities, individual laboratories may have their own management systems, in which case common performance requirements should be agreed.

A QMS should be considered and prepared while the monitoring programme is being designed and it should be in place before the monitoring programme commences and samples are ready to be analysed. In some circumstances, the QMS may need to be adapted to accommodate new circumstances or programme requirements.

2.2 Defining the need for sampling and analysis

Once the requirement for sampling and analysis has been identified as part of a monitoring programme there are several issues that should be addressed and resolved with the analytical laboratory before monitoring commences. It is important to determine whether:

- Appropriate detail has been provided. Is there enough detail in the request to determine the kind of monitoring data that are required? For example, has the required fraction of phosphorus (P) been specified, i.e., Total P, Orthophosphate PO_4. What are the likely ranges of concentrations expected or the limits of detection required? Will samples require filtration?

- The laboratory is capable. Can the laboratory carry out the sampling and analysis required? Do the staff have the skills and expertise necessary to complete the work?

- The laboratory has sufficient capacity. Does the laboratory have sufficient time and resources (e.g., personnel and equipment) to complete the analysis required in a timely and competent manner?

If any requirements cannot be met, or if the detail provided is insufficient, it will be necessary to negotiate with the monitoring programme designers or agency that has requested the sampling and analysis work. The agreed requirements and deliverables should be recorded.

The laboratory can make a useful contribution to the general monitoring programme design because laboratory staff will often have valuable insights that might help save time, energy and resources. The laboratory will be able to assist in detailing the field supplies needed for sampling and will also be able to contribute to the monitoring programme design by providing information on the laboratory turnaround time for samples which, in turn, could influence the timeframe for sampling, analysis and reporting. The laboratory may also be able to anticipate issues that may arise for the monitoring programme, so that they can be resolved before they become critical.

It is very important that both the managing and coordinating laboratories gain an understanding of the context that has led to the need for monitoring and analysis. This may enable useful information to be provided to the "clients" requesting the monitoring and the personnel designing the monitoring programme.

Communication among all stakeholders is a key factor in the efficiency of running any monitoring programme and it is integral to the application of a quality system. It is extremely important that the monitoring programme director and field staff share information with the laboratory throughout the period of the programme, so that the sampling and laboratory staff can adjust their activities or approach if required. Such information may arise from changes in the monitoring programme design which can be a result of, for example:

- the addition or removal of sampling locations,

- the need for additional field supplies or additional analyses,

- a requirement for faster reporting of results, or

- additional complexity in the programme, such as the inclusion of sediment or biological samples.

Changes to monitoring programmes should be agreed with all relevant parties and documented under the QMS.

2.3 Sample collection

The second step in the sample life cycle is that of sample collection. Locations, with GPS coordinates and possibly photographs, as well as sampling techniques should have been agreed and defined in the QMS and at the monitoring programme design stage. Once the scope and requirements for sampling and analysis have been defined, the laboratory and field staff will have to assemble sampling equipment, sample bottles and associated field consumables, and supply them to the field team. This includes equipment and consumables for any *in situ* analyses. The analytical laboratory should ensure that the correct sample containers are provided to field staff, or the specifications for sample containers to be used in the field must be precise, e.g., glass or polyethylene containers. There are several international sources, such as ISO (2018) and Rice, Baird and Eaton (2017) that provide guidance on the preservation and handling of water samples. In general, there is good agreement between the various sources, but they do not always recommend the same types of containers and preservatives, so it is necessary to check with the analysing laboratory which approach they follow.

BOX 2.1 A SIMPLE CHECKLIST FOR SAMPLE COLLECTION CAN HELP REDUCE UNNECESSARY ERRORS

- Sample containers and waterproof labels.
- Field kit consumables.
- Preservatives (where appropriate).
- Sampling instructions.
- Calibrated field equipment.
- Field record sheets or electronic records.

The analytical laboratory should also be able to provide assistance with field recording sheets and sampling records, and any other relevant information. A simple checklist (**Box 2.1**) can reduce errors by ensuring nothing is forgotten when going out into the field.

There are many factors that can affect the quality and integrity of samples collected in the field. For example, the detail and quality of field notes can affect later interpretation of results, so they must be legible and complete. Field equipment must be calibrated prior to use and maintained regularly. Appropriate containers and clear legible labelling are essential for subsequent sample handling in the laboratory. Proper sample preservation directly influences the stability of the variables that will be analysed back in the laboratory, but it should be clearly indicated when preservatives are hazardous. Freezing or cooling may be necessary for some types of sample and others, e.g., those for microbiological analysis, may have to be analysed within a specific time period following collection. Ensuring appropriate field quality controls are applied *in situ* can assist in determining the quality of the data subsequently generated. There may be a legal requirement to follow a chain of custody for some types of monitoring, which must be adhered to, especially when samples are transferred between persons or organizations. It is good practice to ensure analytical request forms are completed because it minimizes the risk of mistaken or omitted analyses.

One of the most important sources of error is sample contamination, or the introduction of artefacts to the sample after it has been taken (**Box 2.2**). Possible contamination can be assessed using field blanks or sample trip blanks, or by the submission of "blind" samples for which the concentration is known only to the submitter. A *field blank* is a sample of clean matrix (typically distilled or deionized water provided by the laboratory) and quality assured to be free of any of the substances (organic, inorganic or both) that are to be analysed in the real samples. The sample in its container is taken into the field and exposed to the atmosphere at the sample site for an equivalent period of time as the other samples. A *trip blank* is a container of laboratory reagent water that is transported, unopened and that is in the same storage and transportation conditions, to and from the field, as the other sample containers. Its purpose is to identify contaminants introduced into samples during transit to and from the laboratory. There is still the possibility that contamination can be introduced in the laboratory, so a trip blank can be compared with the laboratory blank. Contaminants found in the trip blank,

BOX 2.2 POTENTIAL SOURCES OF CONTAMINATION OF WATER QUALITY SAMPLES

- The sampling equipment itself – has it been cleaned between uses?
- The sampling supplies (e.g., contaminated preservatives). Some plastic bottles have been known to leach Zinc and may require to be acid rinsed before being used.
- The field technician taking the samples – are they a smoker? Cigarette smoke contains nitrogen compounds, metals and hydrocarbons which could contaminate a sample if the sampler's fingers come into contact with the sample or the interior of the sampling equipment. They should use personal protective equipment when handling samples.
- The environment in which the sample is transported – e.g., possible contamination from fuel and volatile compounds which may be stored in the vehicle. Samples stored in the vicinity of fuel cans can easily be contaminated with hydrocarbon residues.
- Poor quality laboratory reagents.

but not in the laboratory blank, may indicate a problem with transportation, storage or container preparation, among others.

Other alternatives to assess possible contamination or laboratory error include the use of a quality control sample that is either submitted blind, where the laboratory is not told that there are quality control samples included in the submission, or non-blind, where the quality control samples are identified in advance to the laboratory. Blind submission (using a surrogate sample identification) is generally the most effective approach for thorough checking of the laboratory's processes, because it incorporates everything from sample cataloguing to data reporting. It excludes any bias that would be introduced into the processing because it prevents the laboratory from taking extra precautions when the laboratory staff know that they are quality control samples. All approaches should be documented in the Quality Control or Sampling Manual comprising part of the overall QMS.

2.4 Transport to the laboratory

Once samples have been collected, the next key step is transporting them to the laboratory. This is a critical step because during transport samples may be out of the control of the monitoring programme director, the field staff and the laboratory. There is always a high risk that the samples will be compromised through loss or breakage, especially if using external courier services. Efforts should be taken to ensure that any sample packaging is suitable for the transport of samples.

It is very important that samples are kept cool until they are analysed, and that they are not allowed to exceed their ambient temperature. Less than 10 °C is the typical transportation temperature for most samples. Temperature control loggers can be used for critical samples. Packing, such as bubble-wrap, minimises the risk of glass bottles breaking during transit. Wherever possible the mode of transportation should be selected to ensure the samples reach the laboratory in the shortest time possible. This aspect should have been considered during the monitoring programme design phase, especially where long transit times are anticipated.

Figure 2.2 Coolboxes used for sample transport. © Deborah Chapman

Portable coolboxes are a practicable way to transport samples and are relatively cheap (**Fig. 2.2**). There are many commercial brands available, but they need to be robust and able to withstand rough handling. Commercial 12-volt options are available with inbuilt temperature control. Ice packs and freezer blocks generally stay frozen for several hours stored in a cool box and can keep samples cool during transit without the need for a refrigerated vehicle. However, it can take up to several hours for samples in a cool box at an ambient temperature of about 20 °C to cool down to around 4 °C. If it is critical that the sample temperature is recorded, a replicate sample (not for analysis) should be taken, and a temperature logger used to verify the sample temperature during transport. This approach may be needed if sample analysis results are likely to be used in legal (infringement) proceedings or prosecutions. More detailed guidance on approaches to ensure the integrity of samples after collection is available in Rice, Baird and Eaton (2017).

2.5 Sample reception and recording

Receiving the samples at the laboratory is an important step in sample processing. It is essential that the field staff responsible for submitting the samples to the laboratory inform the laboratory whether there have been any holding time issues (such as transport delays which may result in samples exceeding recommended hold-times) and give information about the mode of transport of samples to the laboratory, so that the laboratory can track samples that are in transit and obtain estimated arrival times. On receipt of samples, the laboratory should record on the sample submission record, the condition of the test samples and the storage container, sign any custody documentation and, having logged-in the samples, ensure they are suitably stored to maintain their integrity. Sample integrity checks include:

- Ensuring that the correct number of samples were received.
- Checking the temperature inside the cool box upon receipt to ensure the samples have been kept at <10 °C.
- Recording and communicating any sample breakage, holding-time violations, preservation issues and labelling issues.

Any issues regarding sample integrity should be communicated to the monitoring programme director, field sampler and laboratory supervisor so that the appropriate corrective action can be carried out. This may involve repeating the sampling.

Once sample integrity checks have been carried out, the samples can be logged with their unique identification (ID) numbers. The numbering system should have been agreed before sampling commenced. Giving the samples a unique identification number ensures that samples can be readily identified during processing in the laboratory and ensures the traceability of all samples, and hence the credibility of the monitoring data. It also allows the laboratory to organize its work and prioritize workflows easily and is useful for the corresponding paperwork, data storage and other information.

In most situations, samples will need to be stored for a period of time prior to and during analysis. Hence it is necessary to ensure that the appropriate preservative has been added to the samples (if required), the sample storage method used is specific to the sample type and measurements parameter(s) and that samples are stored in a temperature-controlled environment. The most typical storage temperature is "refrigeration" which is generally recognized as being 4 °C ± 2 °C. Frozen storage (at ≤ 20 °C) is a specialized storage technique that may be used for certain types of samples and for specific analyses, e.g., Total P. In general, extended storage at "room temperature" (ca. 20 °C) should be avoided, especially where the room may experience large fluctuations in temperature and humidity which may rapidly compromise the sample. If necessary, sub-samples should be taken and the bulk sample returned to cold storage as soon as possible. Once stored in the laboratory, the samples should be analysed as quickly as possible.

2.6 Laboratory analysis

For many water quality laboratories, a key function is the analysis of samples to check compliance with standards and regulations, such as for drinking water and effluent discharge samples. Such regulations often define the analytical accuracy and the rigor of the methods to be applied. Method choice, validation and suitability are covered in Chapter 10. The need for a rigorous quality assurance process is particularly important when water quality monitoring results may need to be defended in legal prosecutions, such as when regulations or license conditions are breached. The QMS should have clear instructions on reporting to the relevant enforcement staff any apparent breaches of regulatory limits. This is especially important with respect to the enforcement of drinking water quality regulations.

There are many different analytical techniques, and associated equipment available, that vary in accuracy, precision and complexity. Examples are:

- Electrometric probe methods, such as those for pH, conductivity, biochemical oxygen demand (BOD).

- Titrimetric assays, such as those for alkalinity and total hardness.
- Gravimetric assays, such as those used for suspended solids.
- Chromatography – gas chromatography (GC), gas chromatography/mass spectrometry (GC/MS), high performance liquid chromatography (HPLC), and ion chromatography used for trace organics.
- Spectroscopy – both visible and atomic spectrometry used for nutrients and metals.
- Microbiological culture techniques.
- Taxonomy and enumeration.

The actual analyses conducted within the laboratory will generally depend on the determinands to be analysed. The performance characteristics, such as sensitivity, precision, limit of detection and accuracy of these methods, should be known before their selection and use. Approaches to assessing these characteristics are covered in Chapter 9.

2.7 Data storage

Once the required analyses have been completed the results must be stored in a database so that they can be integrated into a laboratory data report and shared with other users of the data.

Errors can (and frequently do) occur in the storage and treatment of data, particularly if they have to be manually transcribed from instruments to notebooks, from technician notebooks to formal record books, and from records to computer spreadsheets or databases. Errors can be minimized by reviewing and auditing transcribed data against original readouts or datasheets immediately after transfer into electronic format. Every step of data transfer or handling has its own potential for introducing errors and should therefore have its own quality control and assurance mechanisms. Sample values should be stored together with information about their credibility and quality. This could include a simple comment or flag to indicate valid or invalid, but this is often missing from many data record systems.

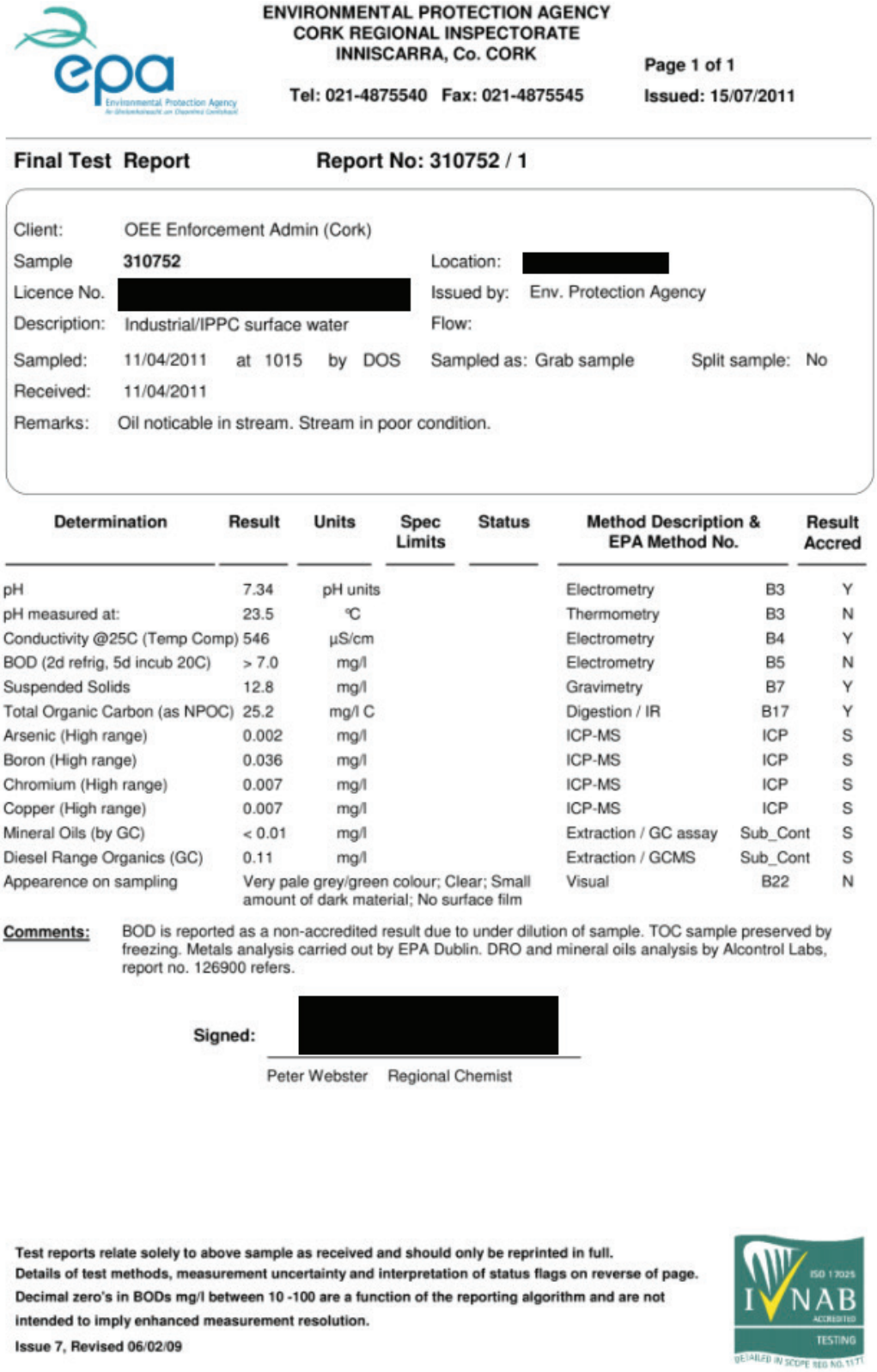
The way in which results from analyses are recorded and stored depends on the nature of the analytical techniques used and the facilities available in the laboratory. For most test methods, manual recording of results is generally necessary. A fully manual system comprises visual instrumentation output, manual transcription and calculation/integration, and manual hard copy report generation. A partially automated system comprises manual transcription to a computer spreadsheet or database, computer integration of data from multiple analyses, and manual manipulation and generation of reports from spreadsheets or databases. A fully automated system uses a laboratory information management system (LIMS), which performs electronic transfer of results to computer spreadsheet or database, computer integration of data from multiple analysis systems and computer-generated test reports.

2.8 Data retrieval and reporting

When data have been compiled and quality checked, they should be sent to the monitoring programme director, laboratory manager or other nominated administrator for sign-off before being issued to the client as a Test Report or Certificate of Analysis (CoA). This can be in hard copy (paper) or digital form depending on the client's requirements. Hard copy reports can be generated using a standard format, such as tables generated from document templates or spreadsheets. **Fig. 2.3** shows an example of a Certificate of Analysis issued by the Irish Environmental Protection Agency.

Where data are to be provided electronically it is important to obtain information about the format in which they will be required, such as Excel spreadsheet, database, etc. For some situations, such as where several laboratories share data with a central database, or where a laboratory regularly reports results to a client's own database, it may be possible to upload the data directly through a web-based platform. These uploads are likely to have particular data formatting requirements. Where work is being undertaken for third party clients, it

Figure 2.3 Example of a Certificate of Analysis. Note the comments to the "client" relating to particular analyses. © Peter Webster



ENVIRONMENTAL PROTECTION AGENCY
CORK REGIONAL INSPECTORATE
INNISCARRA, Co. CORK

Tel: 021-4875540 Fax: 021-4875545

Page 1 of 1
Issued: 15/07/2011

Final Test Report **Report No: 310752 / 1**

Client: OEE Enforcement Admin (Cork)
Sample 310752 Location: [redacted]
Licence No. [redacted] Issued by: Env. Protection Agency
Description: Industrial/IPPC surface water Flow:
Sampled: 11/04/2011 at 1015 by DOS Sampled as: Grab sample Split sample: No
Received: 11/04/2011
Remarks: Oil noticable in stream. Stream in poor condition.

Determination	Result	Units	Spec Limits	Status	Method Description & EPA Method No.		Result Accred
pH	7.34	pH units			Electrometry	B3	Y
pH measured at:	23.5	°C			Thermometry	B3	N
Conductivity @25C (Temp Comp)	546	µS/cm			Electrometry	B4	Y
BOD (2d refrig, 5d incub 20C)	> 7.0	mg/l			Electrometry	B5	N
Suspended Solids	12.8	mg/l			Gravimetry	B7	Y
Total Organic Carbon (as NPOC)	25.2	mg/l C			Digestion / IR	B17	Y
Arsenic (High range)	0.002	mg/l			ICP-MS	ICP	S
Boron (High range)	0.036	mg/l			ICP-MS	ICP	S
Chromium (High range)	0.007	mg/l			ICP-MS	ICP	S
Copper (High range)	0.007	mg/l			ICP-MS	ICP	S
Mineral Oils (by GC)	< 0.01	mg/l			Extraction / GC assay	Sub_Cont	S
Diesel Range Organics (GC)	0.11	mg/l			Extraction / GCMS	Sub_Cont	S
Appearence on sampling	Very pale grey/green colour; Clear; Small amount of dark material; No surface film				Visual	B22	N

Comments: BOD is reported as a non-accredited result due to under dilution of sample. TOC sample preserved by freezing. Metals analysis carried out by EPA Dublin. DRO and mineral oils analysis by Alcontrol Labs, report no. 126900 refers.

Signed: [redacted]
Peter Webster Regional Chemist

Test reports relate solely to above sample as received and should only be reprinted in full.
Details of test methods, measurement uncertainty and interpretation of status flags on reverse of page.
Decimal zero's in BODs mg/l between 10 -100 are a function of the reporting algorithm and are not intended to imply enhanced measurement resolution.
Issue 7, Revised 06/02/09

is essential that the reporting formats meet their needs. This may require the preparation of client-specific reporting formats, whether as hard copy or as electronic transfer of data. The format of any test reports, how they will be stored and for how long, client confidentiality, etc. should be specifically defined within the laboratory's QMS.

Some of the potential issues that may be raised upon reporting and/or receipt of the data report are missing data, turn-around times not being met for some parameters, and data integrity issues. Requests may also be made for re-checks of some or all results, although in practice this may be impractical due to sample degradation. Changes in the formatting of the data (e.g., special requirements by regulators or lawyers for legal cases, etc.) may also be requested. Any anomalies raised must be communicated back to the laboratory manager. All issues need to be addressed before the work is considered complete. Relevant protocols should be set out in the QMS documentation covering how the validity of any issues will be established; once established, what corrective action(s) are necessary; and whether a high degree of urgency will need to be placed on the issue(s). Once the issues have been resolved, a revised or new report can be produced. This should indicate that it is a replacement for the original report, which should be retained. Finally, it is good practice to follow-up and ensure that the monitoring programme director or client has received the data successfully and that they fulfil their requirements.

2.9 Archiving and disposal of samples and extracts

It is generally not appropriate or economically viable to retain water samples indefinitely that have a very short shelf life or hold-time, or to store samples or extracts derived from samples. An agreed archive time might need to be set for such samples. All stakeholders in a monitoring programme, or the organizations requesting sampling and analyses, should be informed of the standard archiving time period. If required, it may be necessary to make special provisions to hold samples longer than the standard archive time. Specific legal requirements may also need to be considered. Therefore, when archiving samples, consideration should be given to the sample stability, setting a standard archive time, legal and contractual requirements, and any special archive times. Once the samples are ready for disposal, they need to be disposed of safely according to local regulations and such that they are not adversely affecting the environment.

Hard copy data records can require a large amount of storage space and a very efficient filing and cataloguing system. Electronic data records are now much more practical. However, neither paper nor electronic records last forever. They can degrade over time, depending on storage conditions. Electronic records may become obsolete with changes in software systems, e.g., historic data may be stored on floppy disks or other (now largely obsolete) media. Where this is the case, efforts should be made to retrieve the data and retain it on more modern media.

It is generally recommended that original laboratory records (e.g., field sheets, laboratory notebooks, etc.) are retained for at least five years, although longer periods may apply. As with samples, it is not viable to store such records indefinitely. A standard archive time should be set and documented for different record types. The archive time should consider aspects such as contractual requirements, legal requirements and public access to information, and storage space restrictions. Electronic records such as databases and spreadsheets should be regularly backed up on secure media in a methodical manner using version control software (if possible).

CHAPTER 3

QUALITY ASSURANCE OF FIELDWORK

Sampling is the first practical step in carrying out chemical, physical and biological determinations of water quality. The goal of sampling should be to obtain a representative sample and to supply it to the laboratory in a condition that ensures it remains representative of the water quality at the sample site.

3.1 The need for quality control in fieldwork

Errors caused by improper sampling, sample pre-treatment, transport and storage cannot be corrected later and resampling is the only alternative if samples are compromised (see ISO 2014). These issues are dealt with in more detail later in this chapter.

Quality control procedures provide a means of minimizing and detecting sampling error, and hence a means of rejecting invalid or misleading data resulting from the sampling process (**Box 3.1**). All quality control procedures should be documented, and a system of record-keeping should be established. The effort expended on sampling and field quality control depends largely on the objectives of the monitoring programme; however, it is recommended that at least 2 per cent of the analytical effort should be devoted to quality control for sampling. There are essentially three main approaches to quality control in the field: (i) collection of *replicate samples* as a check on the precision of sampling, (ii) use of *field blank* and *trip samples* to monitor potential sources of sample contamination, and (iii) use of *spiked samples* to assess sample stability during transport and storage.

BOX 3.1 THE OBJECTIVES OF QUALITY CONTROL FOR FIELDWORK ACTIVITIES IN SAMPLING

The three main objectives in quality control for fieldwork activities when sampling are:

- To monitor and detect sampling errors in order to reject invalid or misleading data.
- To demonstrate that sampling errors have been controlled adequately.
- To indicate the variability in results arising from sampling and to illustrate the importance of this error.

There are several principal sources of sampling error. The level of knowledge of the sampling personnel is very important and training of staff is essential to ensure the quality of samples. Personnel should be able to provide input into issues relating to the suitability of sample sites, such as access and safety issues, as required. Using an incorrect sampling location due to inadequate site description, incorrect georeferenced coordinates or difficulty with access can also lead to errors. **Fig. 3.1** shows the incomplete mixing of a pollutant discharge in a receiving river flowing slowly through a lowland area. It is several kilometres downstream (Fig. 3.1C) before mixing is likely to have occurred throughout the width of the river. The analytical results obtained from samples taken from the bank of the river, especially at 200 m downstream, could give very different values. Sampling from the river bank only provides accurate

Figure 3.1 **Phases of mixing of an effluent discharged into a receiving river. *A (left)*. At the discharge point. *B (middle)*. At 200 m downstream. *C (right)*. Approximately 2 km downstream. © Peter Webster**

information about the potential pollutant impacts on the whole river if the discharge is well mixed across the complete width of the river.

Samples can also be compromised due to inadequate labelling, or incomplete or incorrectly completed sampling protocols, including sample handling on-site and inappropriate transport and storage. Any deviation from the sampling procedure, such as collection of non-representative samples that do not satisfy the monitoring objectives, can mean that data are not fit for purpose.

Another source of error is contaminated sampling equipment leading to carryover of substances from one sample collection to the next. This may be caused by inadequate rinsing or cleaning of the sampling and field equipment between samples, i.e., buckets, samplers, field filtration equipment, etc. The use of unsuitable sampling devices and containers can also lead to sample errors, such as the use of devices or containers that may introduce contaminants to the samples through leaching or abrasion of the container material, or from lubricants in pumps.

Contaminants may be introduced to the sample from the environment during the sampling procedure. For example, contaminants may arise from materials from the bank and sediment of the water body; from abrasion of bridge railings; from sampling devices, tubes, sample containers and lids coming into contact with soil (**Fig.3.2**); and from filling and storing of samples in air contaminated by pollutants, such as exhaust fumes, or from outgassing of preservatives, or volatilisation from strongly contaminated samples. Cross contamination can also occur from preservative chemicals or from mixing up sample bottle lids.

Figure 3.2 **Filling a sample bottle from a Ruttner grab sampler. Neither the sampler nor the bottle must touch the ground until the bottle is filled and the lid is tightly closed. © Deborah Chapman**

There are several technical and personnel requirements that need to be taken into account when sampling. To take a sample correctly, adequate and cleaned equipment (such as sample containers, sampling devices, filtration equipment, a homogenizer, an intermediate container (funnel/spoon), and measurement equipment for on-site analysis) should be available for each sampling campaign. Regular maintenance of all equipment should be guaranteed. The sampling vehicle(s), such as vans, trailers and boats, should be equipped appropriately for sampling (e.g., adequate storage space for sample equipment and containers, anchor and winch on a boat). In areas where alien species (such as the European Zebra mussel or water hyacinth) or potentially zoonotic spores (such as *Cryptosporidium* and *Giardia*) may be present, and potentially could be transferred from one sampling station to another, suitable decontamination procedures may be necessary between sampling locations.

Sampling personnel should have relevant training in the correct use of the equipment, associated quality procedures and related safety requirements. This should take the form of essential and relevant job-training prior to commencing the first sampling campaign, followed by regular training subsequently. Participation in any training should be documented. The competence of staff should be subject to periodic assessment, with refresher training provided (where necessary) or whenever new procedures or equipment are introduced. Regular exchange of information between sampling personnel and laboratory personnel improves the quality of sampling and testing.

Standardized procedures or operating instructions (often known as SOPs) should be prepared for all activities. Each person responsible for collecting water samples should carry an up-to-date sampling manual with them during the sampling campaign. The sampling manual should provide specific guidance and detail for all quality assurance procedures (**Box 3.2**).

BOX 3.2 ESSENTIAL CONTENTS OF A SAMPLING MANUAL

- Sampling procedures (including any pre-treatment) for each medium and variable to be sampled including the type of sample to be collected (e.g., grab, integrated, composite).
- The cleaning procedure and shelf-life for bottles, containers and closures used for each variable, including the amount and type of preservative to be added (if necessary).
- The name and description of preservation reagents (including their usual colour), together with the appropriate safety measures in case of a splash or spill, e.g., contact with skin and eyes.
- Details for *in situ* measurements and for operation or download procedures for any online continuous monitoring sensor equipment.
- The types of bottles or containers, their closures (lids and/or covers) and the specific purposes for which they are to be used.
- The frequency and order of sampling e.g., from least polluted to most polluted.
- The conditions of storage, transport and delivery of samples and the maximum time that can elapse before analysis should commence for each variable (e.g., holding time) and temperature limitations.

It is also recommended that the manual provides guidance for alternative and appropriate sampling actions when unusual conditions are identified (e.g., very high water levels), and a contingency plan for emergency conditions.

3.2 Prior to fieldwork

Prior to any field work it is essential to prepare all equipment and sample requirements and to have them ready before departure, including sensors and test solutions for any *in situ* measurements (e.g., temperature, oxygen, pH, conductivity, turbidity) together with any necessary filters and preservatives for sample pre-treatment (such as homogenization or filtration). Associated materials (e.g., labels for sample containers, disposable gloves, distilled or deionised water for rinsing probes, sample preservatives and any necessary dispensers) as well as safety information, protective clothing and equipment, such as a First Aid kit and lifejackets, should also be assembled.

Field staff may also need to ensure valid access authorizations (e.g., keys to on-site equipment housing or gates to sample locations, access permits, ID cards) are available. Finally, the sampling vehicle should be checked for operating and traffic safety, cleanliness, and a functioning refrigeration unit (if applicable).

On arrival, and while at the sampling location, it is important to verify the accuracy of the sampling location and exact position for sampling. This can be done from geographical coordinates, sampling point number, and position in the water body (close to bank, centre of stream, etc.). Risk Assessment documentation should be checked to determine whether there have been any significant changes from previous visits that affect the suitability of the location. Any changes should be noted and reported. Environmental conditions at the time of sampling should be recorded on the field data sheet, including the date, time, weather, and any observations about the condition of the water. Wherever possible, the water level or flow should be recorded from a nearby gauging station or, if necessary, a subjective assessment should be used, such as low flow, moderate flow, high flow, flood, etc.

Field personnel should ensure the correct sampling equipment and sample containers are selected for each type of sample required and they should ensure that the sample containers are correctly and completely labelled with markers that are not water soluble so that the labels cannot be accidentally erased. Pre-printed labels prepared in the laboratory can be used to avoid such errors. They must also ensure that the equipment used at each sampling has been cleaned. This can be achieved by either having several pre-cleaned sampling devices available, or pre-rinsing sampling equipment three times with water from the sample location, or with deionized water, to prevent any carryover of residues from one sample location to the next. In certain situations, such as high concentrations of suspended material or visible presence of oils and fats, sampling devices should not be pre-rinsed with water from the sample location because residues may adhere to the sampling equipment.

3.3 During the field measurements

Particular care should be taken with measurements and analyses performed on-site or *in situ* and to the correct recording of these results. Guidance is provided in ISO (2009) regarding analytical quality control for water analysis and in ISO (2003a) regarding online sensors. *In situ* measurements should be performed before any water samples are taken, because they might provide information that affects the subsequent collection of water samples and because the sampling process might disturb the water column and lead to further samples being unrepresentative. Field measurements can either be performed directly in the water body, or in a grab sample that is discarded after the measurements. Sensors or electrodes should not be used in combination in grab sample containers, because some sensors could contaminate the sample.

It is important to ensure that field instruments are regularly calibrated. Manufacturer's instructions detail the functional tests, calibrations and operational procedures for each variable. On-site verification tests for each piece of field measurement equipment should be performed *in situ* before commencing a series of measurements. Back-up instruments should be available in case of breakage or malfunction for any critical applications.

Parameters such as temperature and pH should be measured *in situ* because they can change if samples are stored. However, it may be desirable, or simply more practicable, to undertake the analysis of other substances of interest in the field, rather than in the laboratory, particularly where long transport times may be needed. There are a wide range of proprietary test kits available but, in general, their sensitivity fails to match that of laboratory analysis. Nonetheless, they may present a viable option for more remote field locations. Guidance on their selection and limitations is available in ISO (2003b) and British Standards Institute [BSI] (2009).

Grab samples are probably the most common type of field sample. Direct filling of the grab sample into the sample container (**Fig. 3.3**) is recommended

Figure 3.3 **Taking a grab sample directly into the sample container. © Deborah Chapman**

where possible because it reduces errors arising from outgassing, sample degradation, and adsorption or contamination from liquid or solid deposits on sampling equipment. The construction material of the sample container used should be selected according to the requirements of the variable and its analysis (see ISO 2020). To avoid changes due to excessive air input, the sample containers for the determination of certain volatile variables should be filled until they are overflowing, allowing at least two volumes to overflow, and then immediately stoppered and checked to ensure the absence of air bubbles.

A *composite sample* typically comprises several grab samples taken over a specified period of time and subsequently mixed, or taken by an automatic sampling device that collects samples continuously on either a time- or volume-basis. For automated systems, the use of flow proportional sampling is recommended. Without cooling or preservation some deterioration of determinands may occur, e.g., loss of ammonia. Such systems are not suitable for the monitoring of volatile substances.

The use of *automatic sampling systems* should be recorded on the sample field sheet. All the parts of the pump that carry water, i.e., the hoses or pipes and associated sampling equipment, should be manufactured from a material that does not change the sample (i.e., adsorb the variables of interest or contribute contaminants). Submersible pumps are preferred to vacuum pump systems in order to minimize outgassing of volatile substances. The pump should be self-lubricating to prevent the leakage of lubricant into the water or into the sample. The siting of auto-samplers and the frequency and volume of sampling events should be sufficient to minimise the likelihood of settlement of suspended particles in the sample tubing.

Depending on the requirements of the monitoring programme it may be necessary to provide alternative samples such as *replicates* and *duplicates*. While this approach can be useful for assessing sampling homogeneity, there is a clear difference between replicate and duplicate samples, which becomes important when samples are being compared for regulatory purposes. The two terms are often interpreted as being identical but it is important to distinguish which approach applies. Splitting a sample, which involves dividing the main sample into two or more sub-samples, can be affected by how well the sample is homogenized, i.e., how well it is shaken and how it is divided. For example, two 500 ml sub-samples could be made by combining five 100 ml aliquots in each sub-sample or by simply pouring a full 500 ml into each sub-sample. This approach is generally considered to generate *duplicate* samples which should be identical in every respect. Dividing samples taken using composite samplers (whether flow or time proportional samples) is challenging to do in a way that would stand up to rigorous cross-examination in a court of justice. Splitting a sample in this manner is quite different from taking two sequential, but discrete samples. Even though these may be taken quite close together in time, they could give rise to differences in quality, especially in very dynamic aquatic systems. These would be considered to be *replicate* samples.

If several grab samples are required to ensure one sample of sufficient volume, or if different sample containers have to be filled with one sample, the homogeneity of the sample should be ensured. This is especially important in samples containing suspended particles and for the determination of variables that may be attached to particles. If necessary, the grab sample should be gently homogenized between filling each sample container using an appropriate clean stirring rod or mechanical stirrer at low speeds to avoid excessive sample aeration. Using a 'bucket' to

sample is easy but it is essential that all subsamples can be filled out of one bucket and that the bucket grab sample is taken from a homogeneous location.

As water samples can be subject to very fast changes due to biological activity and chemical processes, appropriate preservation or pre-treatment measures should be performed immediately after sampling. The type of sample pre-treatment and preservation agent should be defined for each variable, or variable group, and included in the SOP. Sampling staff should adhere to these guidelines at all times.

In some cases, the objective of sampling may include the determination of soluble components (e.g., metals, nutrients, dissolved organic carbon). For this, it is advisable to separate the dissolved component from any particulates at the sampling site prior to transport to the laboratory. In this way, changes in the composition that may otherwise occur after sampling, and prior to any treatment in the laboratory, can be minimised. However, this may not always be practicable, for example, when it is raining heavily. If separation is not possible on-site, the components should be separated immediately after receipt in the laboratory. Where separation (e.g., filtration, sedimentation or centrifugation) is required, it should be done before any sample preservation. The selection of the separation method depends on the instructions relating to the recommended methods for each variable.

When both total and dissolved metal analyses are required from a single sample, a sub-sample for dissolved metal analysis must be filtered before acidification of the remainder which can then be used for total metal analysis. Filtration for dissolved metal analysis should be done immediately after sampling on-site, because significant losses by adsorption to container walls are likely to occur in a very short time, especially in samples with high contents of suspended matter.

For on-site filtration, portable filtration devices using membrane filters (e.g., 0.45 µm pore diameter) can be used. Rinsing the filter or syringe several times with distilled water (in the laboratory or on-site) is necessary prior to filtering the sample when the analysis is for certain variables, such as dissolved organic carbon, depending on the construction material of the filtration unit or syringe. Deionised water blanks should be run through the filter regularly prior to its use for sampling, and then analysed along with the filtered field samples in order to check for carry-over. Guidance on preservation of the individual parameters is documented in ISO (2018). Sample containers for chemically preserved samples should be marked accordingly. Quality assurance samples used to check for issues due to transportation, stabilization and storage should be treated with the same processes as test samples. In addition, identification information on sample labels for quality assurance samples should ensure anonymity for these samples.

The accuracy of the results of the water quality sampling can only be verified by taking replicate samples. The precision of sampling can only be determined indirectly, provided the precision of the other steps of the analysis is known. A range of protocols, mainly involving sample subdivision and replicate analysis, are available to identify and quantify errors associated with sampling. These are extensively documented in ISO (2014) which provides schematics for a range of approaches to assessing accuracy and precision using replicate samples. In general, the use of such detailed sample subdivision and analysis, is only practicable when there is clear evidence of significant variations between expected and actual analysis outcomes, which cannot be explained by instrumental variance alone.

CHAPTER 4

QUALITY ASSURANCE IN THE LABORATORY

This chapter gives an overview of the elements of Quality Assurance in the analytical laboratory. It discusses each of the key elements (**Box 4.1**) in more detail. Although laboratories may not choose to undertake the rigors of achieving ISO 17025 accreditation (ISO 2017), it is highly desirable that they have suitable documentation in place, in the form of a Quality Manual and SOPs, at the very minimum.

BOX 4.1 THE KEY ELEMENTS OF QUALITY ASSURANCE IN THE LABORATORY

- *Documentation.* This is central to the QA/QC process and details how the laboratory operates.
- *Staff training.* This ensures that staff are competent to undertake the various tasks assigned to them.
- *Laboratory facilities.* Suitable premises and environmental conditions are essential to producing satisfactory outcomes.
- *Equipment.* Equipment should be suitable for purpose, calibrated and well maintained to ensure satisfactory performance.
- *Reagents and standards.* All reagents should be of analytical quality grade and within their use-by dates.
- *Sample receipt, storage and disposal.* Procedures for sample receipt, storage and disposal should be documented.
- *Reporting of results.* Reporting of results should be done in a timely manner as agreed with the "client".
- *Internal Quality Control (IQC).* All procedures should be subject to the use of suitable internal QC standards.
- *External Quality Control (EQC).* Where practicable the use of external, independent, performance testing should be undertaken.

4.1 Documentation

The laboratory should have a Quality Manual which describes in detail the policy on quality and the quality management structure. It describes or refers to the procedures which constitute the working quality system and should outline at the very least:

- Communication channels and the reporting structure within the laboratory.
- Job descriptions and responsibilities for each member of staff.
- The role of QA in the laboratory and who is responsible for each area/activity.
- The records of routine operations that should be kept.

The quality manual should be available to all staff and they must be instructed to read it and to use it to guide them in all aspects of their work. It may be completely self-contained or it may link to technical documents such as SOPs, which describe in detail every procedure conducted in the field and laboratory. These procedures include sampling, transportation, analysis, use of equipment, quality control, calibration,

production of reports, etc. They are the laboratory's internal reference manual for the specific procedure to which they are dedicated and must document every step of the procedure so that anyone of the appropriate training grade should be able to apply the procedure when following the SOP. The technical procedural documents should also contain, or refer to, the operating details for all the instrumentation used to carry out the analyses.

The quality system records maintained in the laboratory should comprise:

- All original observations, raw data, calculations and derived data in the form of worksheets, notebooks, instrument output, etc. These must be dated and should ideally be traceable to the person who made the observation or measurement, and if relevant, to the equipment used.
- Records of installation, maintenance, calibration and checks carried out on instruments and other equipment. This should be in the form of an individual equipment log for each major item of equipment, or a composite log for smaller items such as balances and thermometers.
- Copies of reports issued by the laboratory.
- Records of staff qualifications, training and review of training.
- Records of all audits and reviews of the quality system, including records of corrective and preventative actions taken.
- Records of all complaints from data users and the responses to non-conforming analyses, and details of follow-up and any corrective action taken.
- Records of suppliers and contractors.

4.2 Staff training

The key purpose of staff training is to ensure that laboratory staff provide the same degree of competency as an experienced analyst. There are many ways in which this can be achieved, but the simplest is by demonstration and practice whereby the trainee watches the trainer doing the test (as many times as is necessary). The trainer then watches the trainee doing the test until they are happy it can be done competently, following which both trainer and trainee analyse the same test samples (five is generally sufficient) and the results are compared for precision and accuracy with the method requirements. The test samples should have a range of differing concentrations of the test parameter. Once the required level of competency is demonstrated, the trainee is considered trained and their training record can be formally signed off by the trainee, the trainer and the Quality Manager. Training should be refreshed at regular intervals depending on how often the test procedure is carried out. Specific efforts should be made to target both women and men for training. Setting quotas has been seen to be a successful approach. The Irish Environmental Protection Agency (EPA), for example, sets a target of at least 40 per cent women and 40 per cent men in research teams (Environmental Protection Agency [EPA] undated).

4.3 Laboratory facilities and equipment

It is important that there are sufficient resources such as space, staff, equipment and supplies for the volume of analyses anticipated, and that adequate space is available to allow work to be undertaken without risk to personnel or to the analytical sample. Sufficient equipment should be available to allow the procedures to be conducted effectively and the environment in which the work is to be conducted must be well controlled and kept clean and tidy. There should also be sufficient storage space for glassware, chemicals, samples and consumables, and there should be adequate numbers of appropriately trained staff available to undertake all the required tasks.

Equipment calibration and maintenance records should be kept for all equipment, and these should be monitored. This reduces the likelihood that malfunctioning equipment will be used for analyses (thereby leading to poor quality data) and allows any problems with equipment to be more quickly diagnosed and corrected. Checks on the reliability of equipment must be performed regularly. The following

are examples of simple check procedures which can be used to assess equipment performance:

Balances. A separate check weight, for which the mass was determined immediately after the balance was serviced and calibrated, should be used. This should be kept in its own container when not in use. If weighing over a wide range, high or low mass check weights may be necessary. The check results must be recorded, and the balance recalibrated if there is any drift. Balances should also be calibrated annually by an approved service agent.

Ovens. These should be temperature profiled to identify potential hot spots and a suitable recording thermometer placed in the position of greatest use.

Pipettes. Several volumes of water (at a known temperature) should be pipetted into a beaker on the calibrated balance. The density of the water at the known temperature is used to assess the accuracy of pipette volume delivery. Variable volume pipettes should be checked regularly. These have a tendency to drift, especially if any liquid has entered the piston mechanism. Positive displacement pipettes are very prone to volumetric errors due to impacts from overfilling. Their accuracy can be checked using water at room temperature and measuring the additive weights of 10 successive aliquots. Micropipettes can be checked in a similar manner but using a heavy solvent such as Bromoform (using suitable protective measures).

Thermometers. Thermometers (e.g., alcohol in glass) should be checked against a calibrated reference thermometer of a suitable range which is used solely for this purpose. A calibrated reference thermometer must be calibrated professionally on at least an annual basis.

4.4 Reagents and standards

Reagents should be logged in a chemical register and their associated Certificate of Analysis (if relevant) should be kept in the appropriate file. They must be of an analytical grade quality. Care should be taken to ensure that reagents are stored according to their material safety data sheet (MSDS). The storage conditions of reagents should be checked, especially for those that must be stored away from the light or at a controlled temperature. Solvents and acids should be kept in suitable (ideally ventilated) storage cabinets. Reagents should be labelled with their expiry date because, in general, reagents more than three years old should be replaced even if unused. The shelf-life of reagents should, therefore, be checked regularly and any that are outdated, or have been improperly stored, should be discarded. All reagents must be disposed of in accordance with the MSDS information. Standard stock solutions (and method-specific standards) should be transferred and stored in suitable labelled containers rather than being left in laboratory glassware such as volumetric flasks (volumetric glassware should never be refrigerated).

4.5 Sample receipt, storage and disposal

Samples should be inspected on receipt at the laboratory to ensure that none are compromised. If any irregularities are noted, they should be recorded and reported on the test report. Samples should then be recorded and logged. The samples must be clearly and uniquely marked (i.e., with a sample number) to ensure that no confusion exists about the identity or source of any sample. They should also be labelled with their "analyse by" date. All samples should be stored in a way that minimises deterioration or contamination. For example, drinking water samples should be stored away from sewage samples wherever possible. Any sub-sampling or splitting of samples to allow for different storage conditions, or sample pre-treatment to increase stability, must also be recorded. The condition of each sample and its storage location should be recorded together with, where appropriate, the analysis to which it is to be subjected.

Disposal of samples should be done when the sample exceeds its stable storage time. With some analyses which are required for legal or for regulatory reasons, there may be a requirement to store a suitable aliquot of the sample safely, for a given time, to allow for re-examination if it is considered necessary. Most aquatic samples can be safely disposed of into

drains together with domestic wastewaters, provided there are no local restrictions. Samples or reagents containing hazardous chemicals or biological agents may require specialised disposal routines.

4.6 Reporting of results

All data should be examined by an experienced analyst to determine whether the results are fit to report. Data should be examined at many stages in the quality assurance system and no data should be reported from assays that are out of control (see Chapter 6). Many laboratories have a system which requires checking of data records and countersigning of analytical reports to act as a safeguard against incorrect data leaving the laboratory. This type of system is only effective when consistently applied.

It is important to ensure that results are reported accurately and in a manner that aids their interpretation. It is often necessary to include information which may have a bearing on interpretation, such as unusual field conditions that affected the nature of the sample, or a variation in the analytical procedure that was applied. All such information must be available to the reporting analyst.

Reports must be prepared according to an agreed procedure, and they must accurately reflect the findings of the study. They should include reference to all calibration and quality control data and to any problems that were encountered during the study (e.g., rejected analytical batches, loss of sample, etc.).

4.7 Internal quality control

Internal Quality Control comprises a range of operational techniques used by the laboratory for continuous assessment of the quality of the results of individual analytical procedures. The focus of IQC is on the assessment of both accuracy and precision. Whereas QA strives to achieve quality by regulating procedures based on management techniques, IQC focuses on the individual method and tests its performance against mathematically derived quality criteria. Further guidance on analytical quality control is available in ISO (2009).

There are usually several different analytical methods available for determining the concentration of any variable in a water sample. The choice of method is crucial to ensure that the results of the analysis are appropriate to meet the needs of the monitoring programme objectives. Validation of a method to determine its performance characteristics must be carried out before it can be put into routine use (see also Chapter 9) and thereafter IQC checks should be carried out on each batch of analyses to ensure performance is satisfactory. These are known as validity checks. They should be carried out on a very frequent basis, for example, one QC standard per 20 samples.

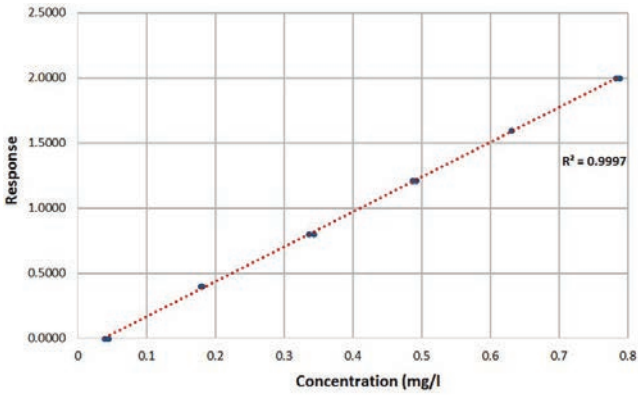
The most widely used approach to method validation is to use a series of solutions, in duplicate, comprising:

- Blank (typically deionised water).
- Low standard (ca. 20 per cent of calibration range).
- High standard (ca. 80 per cent of calibration range).
- Low concentration sample of suitable matrix.
- Spiked low concentration sample (spiked to ca. 50 – 80 per cent of range).

The precision and reproducibility of analyses is then assessed and compared to the target performance. The overall approach is outlined in detail in Irish National Accreditation Board (2019) and similar guidance may be available from other national accreditation bodies. If the method is known to provide a linear response, a sample in the 10–20 per cent range, and one in the 60–80 per cent range, should be analysed. If precision is checked at only one concentration of the variable, it is impossible to detect with certainty whether precision is deteriorating at other concentrations.

If a calibration curve (**Fig. 4.1**) is being used, standard solutions should be analysed from time to time within the required range of concentration. The ideal calibration curve is one which is linear within its most useful range, with a regression coefficient of 0.95 or better. Deviation of individual calibration points from the line of best fit can be used to assess the

Figure 4.1 **Example of a calibration curve for ammonia (NH_3) (mg l^{-1}). Source: P. Webster**



precision and stability of the calibration, which should be within the mean precision limits for the method. If IQC suggests any unacceptable deviation from this stability the calibration should be re-run. In practice, however, many analytical systems will be calibrated immediately prior to use.

Using *reference materials* provides some monitoring of accuracy, but they are only useful if the reference material to be used will be stable in storage for a sufficiently long period of time. The reference material can be prepared in bulk, in-house, by taking previously analysed samples and mixing them and subsequently aliquoting the resulting pooled mixture. Typically, four aliquots would be analysed over five batches (preferably by different analysts) to determine the mean concentration of the variable, and the standard deviation and coefficient of variance at that concentration level. This approach is lengthy and a more practical alternative may be to use, for example, a commercially bottled groundwater as a surrogate, spiked if necessary for trace ions. It can remain stable for many months for the common ions at mg l^{-1} concentrations.

Certified reference materials are matrix-matched materials with assigned target values and assigned ranges for each variable, reliably determined from data obtained by repeated analysis. Target and range values are generated from data produced by several laboratories using different analytical methods. The declared mean (target value) will be a close approximation of the true concentration of the variable in the reference material. The mean and standard deviation become the basis of the acceptance criteria for the analytical method and may be used to draw up control charts (see Chapter 8).

4.8 External quality control

External Quality Control is a way of establishing the accuracy of the analytical methods and procedures by comparing the results of analyses made in one laboratory with the results obtained by others conducting the same analyses on the same material. The general objective of EQC is to assess the accuracy of analytical results measured in participating laboratories and to improve inter-laboratory comparability. For an individual laboratory, participation in an EQC exercise is the only way to ensure that accuracy is independently monitored. This is usually carried out using proficiency testing or rechecking/retesting of samples by a reference laboratory. There are several commercial, external proficiency testing schemes which facilitate the analysis of unknown test samples between multiple laboratories. Where practicable, the use of these is encouraged because it provides a comparison of any systematic errors with other laboratories.

Wherever possible, laboratories should participate in relevant EQC programmes for each variable routinely analysed. This approach supplements IQC as part of the laboratory's normal procedures. Participation in relevant EQC programmes, and maintenance of adequate performance in those programmes, is often a requirement for laboratory accreditation.

CHAPTER 5

VISUALISING QUALITY AND MEASURING UNCERTAINTY IN THE LABORATORY

This chapter introduces approaches to visualising quality using charts and graphs, and to the sources and assessment of measurement uncertainty for different types of water quality analyses.

5.1 Control charts

Quality variation within an analytical process is determined by running special control samples or standards with each batch of test samples. The concept of control charts is founded on the assumption that quality variations are attributable to both "assigned causes" and "chance causes". Although every process displays some degree of variation, some display controlled variation that is natural, while others display "uncontrolled" variation. If a process is operating within "statistical control", the distribution of chance would result in most measurements of the control samples tending towards following a normal distribution, with the bulk of measurements centred on the true value with a small proportion of more extreme values. All processes, including analysis, can fall into one of four states:

- The ideal state, in which the process is in statistical control and produces 100 per cent conformance, is stable, and is predictable.
- The threshold state, in which the process is now characterized by being in statistical control but still producing the occasional non-conformance.
- The brink of chaos state, which reflects a process that is now not in statistical control, but is also not producing defects, e.g., incorrect results. It is, however, becoming unpredictable. The lack of defects can lead to a false sense of security but, because such a process can produce non-conformances at any moment, it is only a matter of time before it occurs. This situation needs action to prevent non-conformances.
- The state of chaos, where the process is very much *not* in statistical control and produces unpredictable levels of non-conformance.

Every process falls into one of these four states at any given time but will not remain in that state constantly. All processes will inevitably tend to migrate towards the state of chaos. Typically, some form of improvement effort starts to be applied when the process approaches (or more often reaches) the state of chaos. In such circumstances, it would have been better to have initiated improvement plans as the process transferred from the threshold state towards the brink of chaos, i.e., when multiple non-conformances became more evident. Simple control charts are very robust and effective tools to use as part of the strategy used to detect this natural process degradation. The common types of charting tools are Shewhart Control charts and Range charts, as well as other charting tools that are useful in data analysis (see, for example, Montgomery 2019).

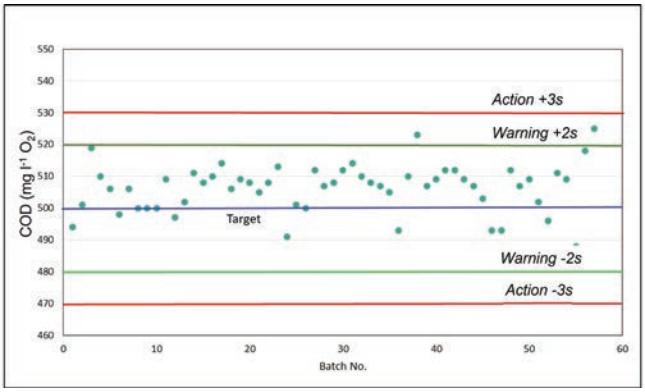
5.1.1 Shewhart charts

The Shewhart chart, also known as Levey-Jennings chart, is the most commonly used form of control chart for laboratories (ISO 2013). Use of the Shewhart control chart is based on the assumption that the analytical process is in "statistical control" i.e., the data are likely to display a bell-shaped, normal distribution, which is characterized by the mean value and the standard deviation. Analyses can only generate a relatively small sample of all the possible measurable values of the test variable and so the true mean value (μ) is substituted with the calculated mean value of all the results ($\bar{\chi}$) and the true population standard deviation (σ) is substituted by the calculated sample standard deviation (*s*).

To construct the control chart, such as the one shown in **Fig. 5.1**, all the measurements of a quality characteristic in control samples analysed at different times (i.e., the raw data), should be plotted sequentially. A centre line, which is the target line, is then drawn at the expected value of the control sample. The standard deviation, *s*, is calculated using all the sample results. Upper and lower control limits indicate the threshold at which the process output is considered statistically "unlikely". These "warning lines" and "action lines" are drawn typically at two and three standard deviations from the centre line. Ideally, charts should also record the date and the analyst initials to assist in checking any performance issues.

Provided the distribution is normal, 95% of normally distributed results should be between the warning lines ($\bar{\chi} \pm 2s$) and 99% of results within the action lines at ($\bar{\chi} \pm 3s$). A single result outside the warning lines should lead to careful review of data from that analytical batch and two or three subsequent batches. Results occurring outside the ± 2*s* warning lines more frequently than once every 20 consecutive analyses of control samples should prompt detailed checking of the analytical method and rejection of the analytical data for that batch. Any result outside the 3*s* action limits should prompt detailed checking of the analytical method and rejection of the analytical data. Two successive results outside the 3*s* warning limit on either side should also result in rejection of the data.

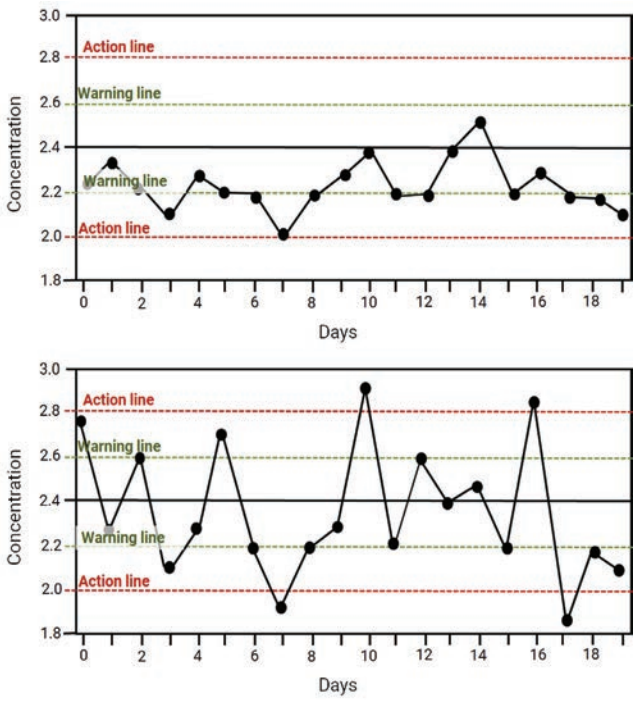
Figure 5.1 **A hypothetical example of a typical control chart for analyses of quality standards for chemical oxygen demand (COD) in wastewater. A standard with an expected value of 500 mg l^{-1} was analysed which each batch of samples. Source: P. Webster**



The scatter of the analysis results for the reference material around the target line **(Fig. 5.2**) provides an indication of the precision of the method, while the mean of the analysis results relative to the target value indicates whether there is any bias (i.e., consistent deviation) in the results. Fig. 5.2A shows poor accuracy (showing what appears to be a negative bias of approximately 0.2 from the target value of 2.4) and Fig.5.2B shows a process which is significantly out of control and suffers from poor precision (repeatability). Both these situations require corrective action to be taken to restore the process to a stage of being in "statistical control".

In addition to the generally applied 2*s* and 3*s* exceedance rules described above, there are additional criteria, such as the four "Western Electric rules" which can be applied to assess an "out of control" situation and may be used to enhance the

Figure 5.2 **Examples of control charts showing A. poor accuracy (top) and B. poor precision (bottom). (after Briggs 1996)**



sensitivity of control charts (Montgomery 2019). In this case, the chart is divided in three different zones (**Fig. 5.3**):

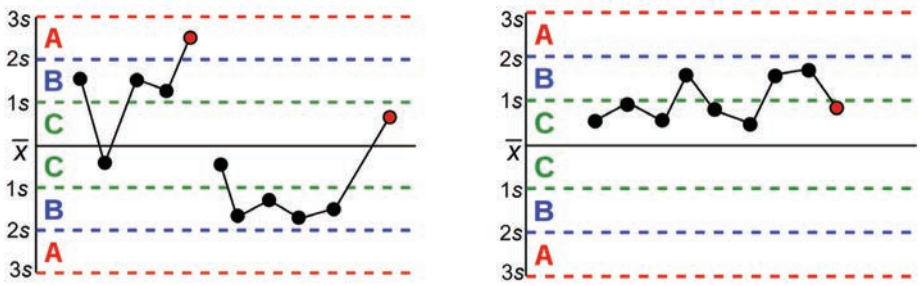
Zone A: between 2*s* and 3*s* from the centre line.
Zone B: between 1*s* and 2*s* from the centre line.
Zone C: within +/– 1*s* of the centre line.

For example, rule 3 of the Western Electric rules states that if four out of five consecutive points fall beyond the 1*s*-limit (i.e., in zone B or beyond) on the same side of the centre line, action is required. Similarly for rule 4, where nine consecutive points fall on the same side of the centre line (bias), the situation is out of control and necessitates corrective action.

Another commonly applied set of criteria are the Nelson rules (Nelson 1984) which propose a further eight sensitizing rules as additional criteria for "out of control" analyses as follows:

1. One point in the control chart is more than 3*s* in either direction.
2. Nine (or more) points in a row in the graph are on the same side of the mean indicating bias to be present.
3. Six or more consecutive points increasing or decreasing (trend).
4. Fourteen or more points in a row alternate in direction (increasing / decreasing) indicating instability.
5. Two (or three) out of three points in a row are more than 2*s* from the mean in the same direction.
6. Four (or five) out of five points in a row are more than 1*s* from the mean in the same direction (significant bias).
7. Fifteen points or more all within 1*s* (greater variation would be expected). Limits are possibly not tight enough.

Figure 5.3 **Control charts showing the principles of Rules 3 (left) and 4 (right) of the Western Electric Rules. The chart on the left shows that action is required and the chart on the right shows that the situation is out of control. (after Montgomery 2019)**



8. Eight points in a row exist, but none within 1*s* of the mean, and the points are in both directions from the mean (**Fig. 5.4**). This is not normal behaviour.

For Shewhart control charts the initial control chart centre line will typically be set at the expected value of the mean. However, in time, sufficient data may have been collected to merit a review of this value. It is common to review chart limits after each 20–30 results. If there is no appreciable change in the calculated centre line mean value, the new chart can be prepared using the original expected value. However, as information on the long-term performance is improved with increasing sample numbers, it is common to apply the "pooled" standard deviation because this reflects the actual variability in the process, given that the standard deviations will be broadly similar. Using this approach, and comparing the individual changes to the standard deviation between batches, provides a more accurate assessment of any long-term change in bias in the test method.

The pooled standard deviation is given by the formula below (Montgomery, 2019):

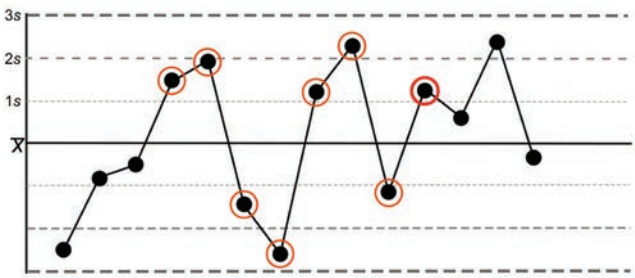
$$S_{pooled} = \sqrt{\left((n1 - 1)\, s_1^2 + (n2 - 1)s_2^2\right) / (n1 + n2) - 2}$$

Where n^1, n^2 are the numbers of samples in respective batches, and s^1, s^2 are their batch standard deviations. An example of the calculation of a pooled standard deviation is given in **Box 5.1**.

Control charts are frequently constructed on graph paper and recorded in paper format. This approach is both simple to prepare and easy to maintain. Commercial LIMS generally incorporate this charting feature, but there are also many proprietary products for AQC charting. One simple approach, although one that also needs security of access and change control, is to use MS Excel or other spreadsheet products. In the "Assay" (results) column, the use of Excel's conditional formatting function can be used to flag results out of specification (TRUE or FALSE for LCL < Assay < UCL) as in the example in **Fig. 5.5**.

If any of the quality control procedures indicate that a method is out of control, or that a problem exists, corrective action must be taken to determine the source of the problem. Calculations, records, standard solutions, reagents, equipment calibration, performance, and quality control materials should all be checked to identify possible contributing factors. Ideally, data should not be reported where the data quality is suspect. However, there can be circumstances (e.g., BOD analysis involving a 5-day incubation period) where repeat analysis is not practicable. Although suspect, the information may still be of use to the client. In this situation, results *MUST* be reported as being associated with an unsatisfactory control situation. Analyses, and any accreditation, should be suspended until the problem is rectified.

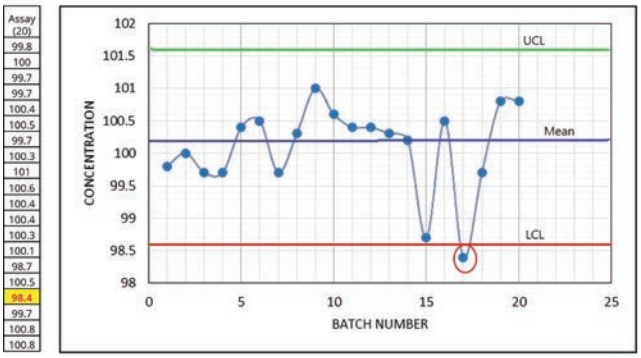
Figure 5.4 **A control chart showing the application of Nelson Rule 8. There are eight points in a row, of which none are within 1s of the mean and the points are in both directions from the mean.**



BOX 5.1 EXAMPLE OF POOLED STANDARD DEVIATION

- Consider two batches of 20 results from QC data for COD.
- In the first batch n = 20 and s = 4 mg l^{-1}.
- In the second batch there are 15 results but s for this batch has risen to 6 mg l^{-1}.
- The S_{pooled} is calculated as follows:
- Sp = Sqrt [(19 × 16) + (14 × 36)] / (33) = 4.94 mg l^{-1}

Figure 5.5 **Example of a quality control chart produced with MS Excel using the data given on the left. The lower control limit (LCL) is 98.58, the upper control limit (UCL) is 101.58 and the mean is 100.21. One data point is below the LCL. Source: P. Webster**



5.1.2 Moving average charts

Although very easy to use, one disadvantage of Shewhart charts is that they can take some time before an "out of control" situation becomes apparent, despite applying the sensitizing rules described above. One means of improving this situation is to use a moving average chart (**Fig.5.6**) where the average value for, typically, the last four results is calculated and plotted sequentially. The first charted point would be results 1 to 4, the second 2 to 5 and so on. The larger the value of "*n*" (4 in this case), the greater the smoothing of variability, but the longer it takes to discern a pattern. If control limits are required these should be set to ± (2 ÷ √n) and ± (3 ÷ √n), respectively and recalculated as new results are input, until *n* = 6 at which point the control values remain constant. This approach is very common in financial analyses but is also quite well suited to laboratory analyses.

5.1.3 Range charts

If the number of QC samples varies each time, the range of results can either be displayed on the Shewhart chart directly (as a vertical line through each plotted point) or on a separate R-chart (Range

Figure 5.6 **Example of a moving average chart. Source: UNISTAT Ltd (undated).**

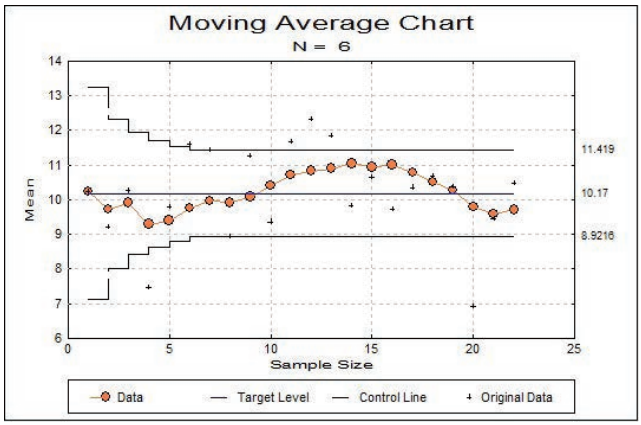
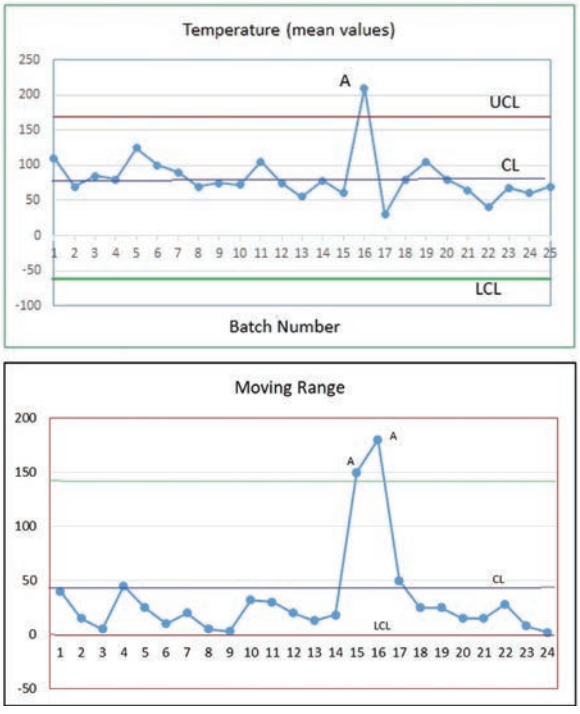


chart) (**Fig. 5.7**) where the expected range (R) would have a value of (Max – Min) = 0. Whereas the control chart using the mean shows any underlying trend in the process, the R chart shows the degree of within-batch variability. For large batches of samples

Figure 5.7 **Example of mean values of temperature (top) and moving range chart (bottom) Source: P. Webster**



(typically more than 20), such as those obtained from an automated analyser, it is generally considered good practice to include at least one quality control sample per 20 test samples.

In Fig. 5.7 there was a significant shift in the temperature pattern of a process at point 'A', as indicated by both the mean value of approximately 220 °C (top graph) and the two successive large range differences (bottom graph). Clearly, something has gone wrong. Most laboratory analyses seldom show such dramatic changes, but it can happen!

As with the 2*s*, 3*s* limits for Shewhart charts, control limits for Range charts are generally set using the following formula depending on the number of samples in the range, where *R* represents the average of the ranges and D_3 and D_4 values are control chart constants (**Table 5.1**):

Upper limit (UCL) = $\bar{R} \times D_4$ Centre line = $\bar{R}$
Lower limit (LCL) = $\bar{R} \times D_3$

Control chart constants can be obtained from Wheeler and Chambers (2010) and ISO (2013). An example of their use is given in Box 5.2. Because automated analysis batch sizes can vary considerably, this approach is generally only applied where there is a consistent number of QC standards in the batch.

For varying numbers of control standards, an alternative and much simpler approach is to set a maximum acceptable range. In practice, this tends to be set at 2*s*. If, for example, the calculated standard deviation for a 10 mg l^{-1} control standard was 0.2 mg l^{-1}, then the maximum acceptable range could be set to 0.4 mg l^{-1}.

Table 5.1 Control chart constants for a Range chart

Sample size (n)	D_3	D_4
2	0.000	3.267
3	0.000	2.574
4	0.000	2.282
5	0.000	2.114
6	0.000	2.004
7	0.076	1.924
8	0.136	1.864
9	0.184	1.816
10	0.223	1.777

Data from ISO (2013)

BOX 5.2 EXAMPLE OF THE USE OF RANGE CHARTS

If the typical range for a series of repeat analyses of the control standard is 0.2 mg l^{-1} then the LCL for a batch containing five standards would be:

$\bar{R} \times D_3 = (0.2 \times \text{zero}) = 0.000$

The target range would be 0.2 while the UCL would be

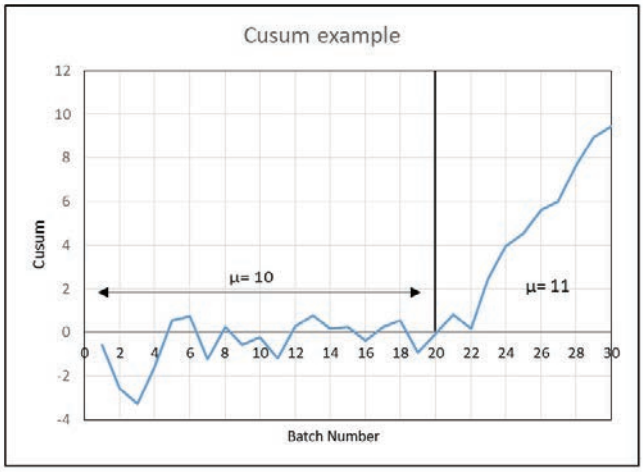
$\bar{R} \times D_4 = (0.2 \times 2.115) = 0.423$

5.1.4 Other charting techniques

A very effective tool for detecting small shifts in the process mean, which can be missed in the random noise of a Shewhart chart, is the Cumulative Sum (CuSum) chart. Further information is available in ISO (2021). This approach involves summing the absolute differences between the sample measurement and the target value, regardless of whether greater or smaller. The running total (T) is plotted against successive measurements. A positive slope implies that the operating mean exceeds the target, while a negative slope suggests it is below the target. Sharp shifts in direction are a sign that there is a problem.

The example data in **Fig. 5.8** show the results of analysis of a sample with an expected mean value μ_0 of 10 mg l^{-1} and a known process standard deviation (*s*) of 1 mg l^{-1} (data are provided in Appendix A). When plotted as individual differences from 10 mg l^{-1}, it is very evident from the difference data that a significant shift in performance is evident after batch 20 with the mean value for batches 21–30 being 11 mg l^{-1} rather than 10 mg l^{-1} previously. This represents a 10 per cent shift in performance, with only 1 of 10 results showing a measured value below 10 mg l^{-1}. The CuSum data can also be displayed in tabular format using spreadsheet functions to calculate and present

Figure 5.8 **An example of a CuSum Chart. Source: Data from Montgomery (2019)**



the information (as shown in Montgomery (2019) and Appendix A). The data can be recalculated to show the one sided upper and lower CuSum limit values. These values are defined by the following equations where C_i^+ and C_i^- represent the positive or negative exceedances of *K*, and the starting values are: $C_0^+ = C_0^- = 0$

The tabular CuSum equations are:

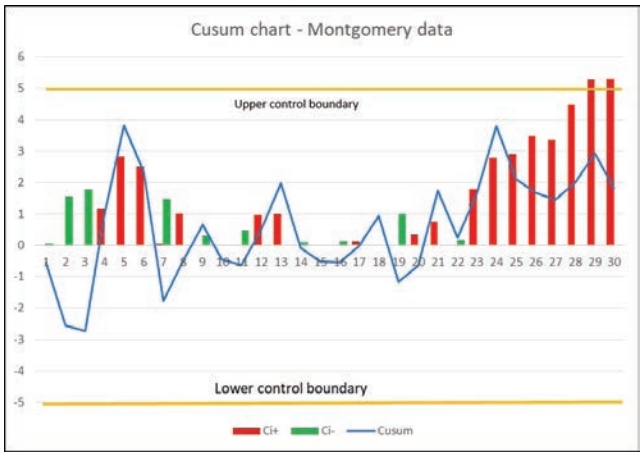
$C_i^+ = \max[0, \chi_1 - (\mu_0 + K) + C_{i-1}^+]$

$C_i^- = \max[0, (\mu_0 - K) - \chi_i + C_{i-1}^-]$

The output depends on the size of the shift that needs to be controlled. Assuming a 1*s* shift is the maximum acceptable, then *K* would be 0.5 × *s* (in this case 0.5 × 1 = 0.5). *K* is usually called the reference value (or the allowance, or the slack value), and it is chosen to be halfway between the target μ0 and the out of-control value of the mean μ1 that needs to be detected quickly. Thus, if the shift is expressed in standard deviation units, then *K* is typically set at one-half the magnitude of the shift. A 'decision interval' *H* is typically set at 5 × *s* so would equate to (5 × 1) = ± 5. The decision interval "*H*" is a multiple of the expected process standard deviation which, if exceeded, will trigger an "out of control" situation. *H* is generally set at 4 or 5 to ensure practical run lengths before "out of control" situations are triggered. Values that are too small (< 4) will result in reactions to small process deviations, whereas larger values (>5) will only trigger action when there are major changes in the process performance.

In the table shown in Appendix A, N+ and N– are simply counters of the number of measurements on each side of zero, and are reset each time the zero is crossed. The graphical output of the dataset available in Montgomery (2019) and given in Appendix A is shown in **Fig. 5.9.** This plot shows the system as having very poor control. There are very regular changes in either direction. Those clustering around the zero CuSum indicate stability, but large swings shown by the elevated values for C_i^- (green) and C_i^+ (red) indicate potential problems. By batch 29 there is an "out of control" situation ($C_i^+ = 5.28$), but it is evident that this began around batch 22–23.

Figure 5.9 **An example of a CuSum chart. By batch 29 there is an "out of control" situation ($C_i^+ = 5.28$), but it is evident that this began around batch 22–23. See Appendix A for data. Data from Montgomery (2019)**

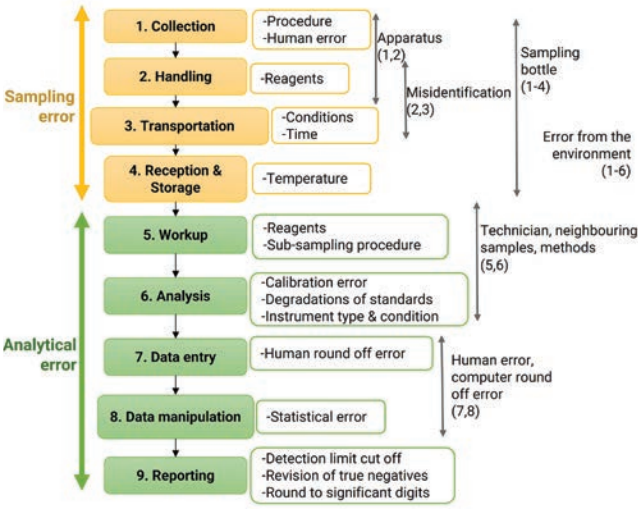


5.2 Measuring uncertainty

There are many potential sources of measurement error at all stages from sample collection to data handling. Some of these can be controlled to some extent and others are totally random in the way they influence the test results. Sources of uncertainty generally arise from: pollutant distribution, sampling, analytical preparation and analysis, and data management and interpretation (**Fig. 5.10**).

The acceptable level of uncertainty depends on the use to which the data will be put. As a general rule,

Figure 5.10 **Sources of error and uncertainty in the chain of activities from sample collection to data storage. Possible causes of each type of error in the chain are indicated in boxes with white backgrounds and potential causes contributing to multiple sources of error are indicated with vertical arrows.**



the aim would be to achieve better than a maximum uncertainty of 20% comprising bias (10%) and standard deviation (5%). Most analytical methods will meet the bias comfortably, but some could have difficulty with the standard deviation. If this is the case, it should be discussed with the person requesting the analyses to determine what is acceptable to them. For trace organics, a broader threshold of typically 30% is applied, with 10% standard deviation.

5.2.1 Approaches to assessment of uncertainty

It is important to realize that any particular analysis will never achieve the "true" result. Even with all the care in the world, the result obtained will only be a good estimate of the true value because there will always be some uncertainty at each stage of the process. There are essentially two distinct approaches to evaluating what influences the overall uncertainty of a test result: the bottom-up approach and the top-down approach.

In the *bottom-up* approach, each source of variation is analysed in turn and its contribution to the overall uncertainty is evaluated. This is particularly useful where processes are relatively straightforward but can be very complex to implement, especially for analyses which involve a large number of individual stages, such as weighing, sample dilution, etc. If there are several possible options, such as choice of pipette or flask volumes, then the overall uncertainty will differ (albeit only slightly), between one sample and the next. It is not practical to recalculate this uncertainty each time and, in any event, this is rarely necessary unless specified by the client or regulatory body.

The *top-down* approach assumes that whatever the sources of variation, it may not be possible to say exactly how they interact or what is their relative contribution. The best guide as to how they all affect the results is to use routine analysis of an AQC sample or a Certified Reference Material (CRM) (see Chapter 4) because these will have been subject to the same elements of uncertainty (positive and negative) as real samples, at least as far as laboratory procedures are concerned.

Many national regulatory and accreditation bodies have country-specific guidance on the application and calculation of measurement uncertainty. These should be consulted where available. Fig. 5.1 shows a hypothetical example of a typical control chart (see section 5.1) for chemical oxygen demand (COD), where a control with an expected value of 500 mg l^{-1} has been analysed which each batch of samples. The average (mean) value is 505 mg l^{-1} (close to the expected 500 mg l^{-1}) and the standard deviation (s) (a measure of the spread of results) is approximately 8 mg l^{-1} (i.e., less than the 10 mg l^{-1} target value assigned). The relative standard deviation of the COD analysis is 8/500 (1.6%). To accommodate the expectation that 95% of results will fall within this range, a coverage factor of 2s is used, and therefore the estimated relative uncertainty associated with this test is ± 3.2%. If additional information on bias is available, e.g., from inter-laboratory performance tests, this can also be factored into the overall estimate of measurement uncertainty. In the example in Fig. 5.1, the mean is approximately 505 mg l^{-1} which

equals a bias of 1%. To be more certain that this is accounted for in the uncertainty estimate, a relative uncertainty of, for example, ± 4% could be applied to measurements at this range. Some uncertainty could be concentration-dependent, therefore using a quality control sample of 50 mg l^{-1} may also be desirable, using the same approach. A graduated relative uncertainty could be applied to intermediate concentrations. At very low concentrations, an absolute uncertainty will apply because relative values will be too small to determine. This threshold can be determined at the performance testing stage, e.g., ± 0.1 mg l^{-1} or 5%, whichever is greater.

A worked example of the calculation of measurement uncertainty for the measurement of total solids in water and wastewater is given in Appendix B.

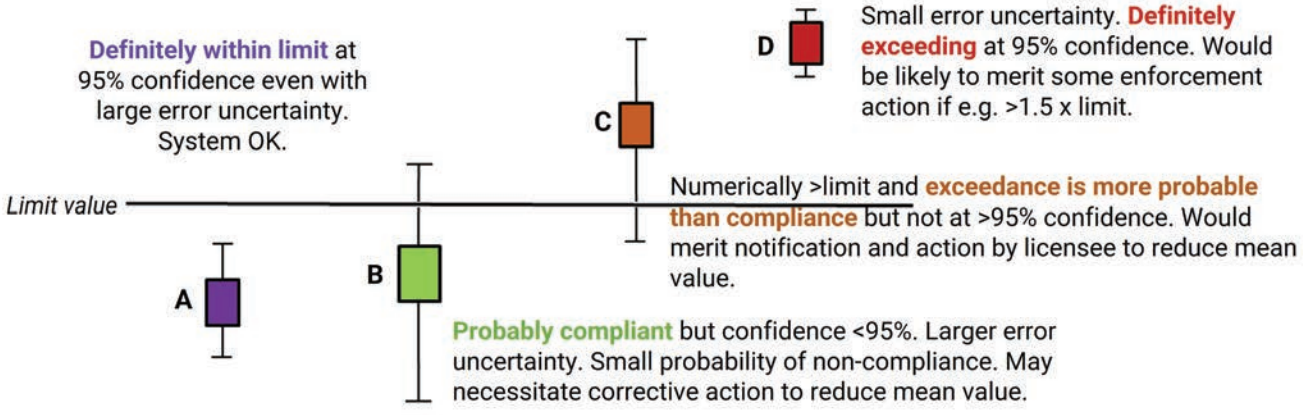
5.2.3 Interpreting exceedances

Reporting the estimated measurement uncertainty is critical if results are to be used for regulatory purposes. It can be very difficult to provide standards of custody, sample handling and analysis which can stand up to scrutiny under oath in a court of law. However, having an estimate of uncertainty can provide reassurance that a parameter has exceeded the regulatory limit with sufficient confidence to merit prosecution.

Assessing whether a value exceeds a regulatory limit requires consideration of the uncertainty of measurement associated with the result(s) (**Fig. 5.11**). Assessment based on single measurements (which is the most common situation) is difficult because there is no information on the range of repeatability of the actual measurement. Variation often has to be inferred from historic data from replicate analyses.

A basic introduction to the terms used in assessing uncertainty and how uncertainty influences decision making can be found in Bell (2001), Crumbling *et al*., (2001), International Accreditation New Zealand (2021) and Standing Committee of Analysts (2018). More detailed and mathematical treatment of the subject can be found in Ellison and Williams (2012), ISO (2012a), Magnusson *et al*. (2017), Singapore Accreditation Council (2008) and Vetter (2001).

Figure 5.11 **Example of determining whether a value exceeds a regulatory limit for four different scenarios (A to D). They all have different mean values, confidence and uncertainty. Assuming multiple measurements, the mid-point of the box indicates the average concentration measured, the box boundaries represent the range of measurements and the whiskers indicate the range of measurement uncertainty. If there is overlap with the limit value then action may be required. Source: P. Webster**



CHAPTER 6

LABORATORY DATA MANAGEMENT

This chapter briefly discusses database structures, data management, and data interpretative tools in the context of ensuring data quality in a water quality monitoring programme. It covers some common reasons for errors in water quality data storage and retrieval and suggests some simple processes or mechanisms for carrying out data quality checks. It also introduces interpretation of data relationships and discusses ways in which these can assist in the management of data.

There are many software tools that can assist in the process of data management. Data are generally stored in one of two systems: spreadsheets (such as MS Excel) or databases (such as MS Access and Oracle). They are fundamentally different in the way in which data are stored as shown in **Table 6.1.** Before selecting which to use, their design and how they can assist in data quality control before and during data entry, should be carefully considered. Some of the key capabilities of spreadsheets and relational databases for data quality management are given in **Table 6.2**. There are also external computer programs or scripts that can read data from any input format, transform them into a database import file, and validate them at the same time.

A wide range of commercial statistical packages are available to assist in data analysis, but they can be expensive. In contrast, a broad range of free-to-use (but occasionally quite complex) routines and scripts have been written for use with the R programming language (R Core Team 2018). At present, there are over 5,000 individual packages covering a wide range of scientific areas including trending, statistics and graphics.

Table 6.1 **A comparison of key features of spreadsheets and databases for storing water quality data**

Spreadsheets	Databases
Designed to store key numeric points rather than details	Designed to hold a larger collection of organized information in one space
Made of rows and columns that simulate a paper worksheet	Made up of several tables, that look like a spreadsheet, where each row is a unique record within the dataset
Manage and calculate information which can be used to create charts and graphs	Each table holds only one type of information
Ideal for calculations and data analysis	Tables are related to each other, enabling records of one table to be associated with records of another table
Tempting as a quick and easy way to store organized information	Suited to complex grouping and aggregation
BUT	*BUT*
Risk of not updating formulae etc. when new data are input	More complex to design and to prepare reports
Not suited to concurrent use	

Table 6.2 A comparison of data quality management features available in spreadsheets and databases

Microsoft Excel	Relational databases
Conditional formatting, e.g., when certain criteria are not met	Validation rules for every individual column
Drop-down selection lists for allowed values	Enforcing of certain data types
Validity checks with error message	Full inclusion of controlled vocabularies through relations
Scripting language (VBA) for background automatization	Scripting languages for background automatization included for some products

Water quality data go through many steps during which errors may occur (see Fig. 5.10), but once errors have entered a water quality dataset, it is hard to remove them. Data contamination can result from wrongly entered, or otherwise altered, values that negatively affect data integrity. This can make it rather difficult to work with the data and generally decreases confidence in the whole dataset. Therefore, it is necessary to take actions that prevent errors from entering the dataset in the first place. This involves applying quality assurance and quality control protocols. Every anomaly encountered during these steps should be documented and finally stored with the data, providing information about their credibility and their quality. This is known as the metadata. Unfortunately, many data storage systems are simply not designed to handle information in this way, so retention of paper records or electronic storage of information in other formats may also be required to support any subsequent data usage.

6.1 Avoiding errors

6.1.1 Before data entry

Standardized terms should be defined and enforced for all relevant fields and table names in the spreadsheet or database. For example, it should be defined whether a given column can hold only numbers or text, or even dates. If abbreviations or codes are to be used, they should be easily understood and defined at the beginning in a Data Dictionary, such as for water body types e.g., RIV for River, RES for Reservoir, GW for Groundwater, and so on. These defined vocabularies will allow only the entry of values that are defined within them. If you only allow RIV, RES and GW as a water body type and make the system respond with an error if anything different is entered, no keyboard entry errors can occur which could result in erroneous outputs during data searches.

Measurement units should also be specified, for example as mg l^{-1} or µg l^{-1} and situations where units are reported in mixed format should be avoided. Where relevant, any required metadata should be stipulated and defined and no exemptions should be allowed. For example, if a column is only allowed to hold the words “YES” and “NO” it should not be able to accept “Maybe”. There should be only one piece of information in each cell of the spreadsheet or database. If there are multiple pieces of information embedded in a single data cell it will cause severe problems during data analysis.

A designated person should be assigned responsibility for data quality assurance and data quality control. Ideally, this should be the person who enters the data and they should be educated in quality control and assurance methods.

6.1.2 During data entry

In general, there are two types of errors that can occur in the process of entering data into a storage system: errors of commission and errors of omission. Errors of commission are the result of incorrect or inaccurate data being included in the dataset. This may happen because a malfunctioning instrument has produced faulty results or data were mistyped during entry, or other problems such as lack of clarity resulting in data being entered into incorrect spreadsheet columns. Errors of omission result from data or metadata being

omitted, for example when data are inadequately documented, when there are human errors during data collection or entry, and when there are anomalies in the field that affect the data. Some examples of omission errors would be a measurement that is unintentionally forgotten and not entered, a line that is accidentally skipped during data entry and the value ends up in the wrong row of the spreadsheet (which is very common) and the GPS unit ran out of power and the sampling station coordinates were guessed or written down incorrectly.

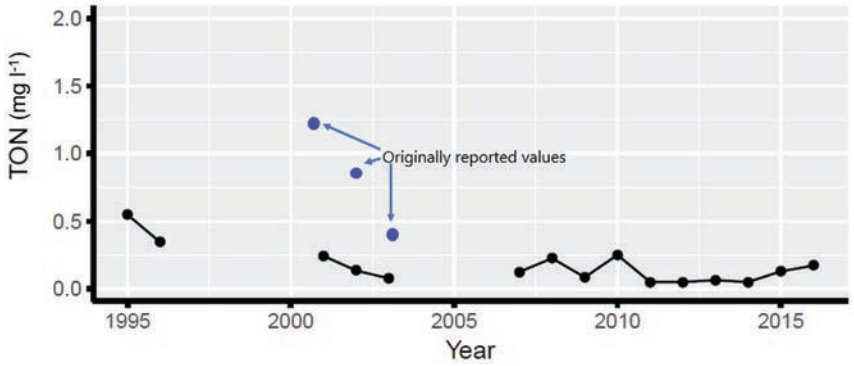
6.1.3 After data entry

After data have been entered into the database, basic quality assurance measures can be taken. If water quality data are in spreadsheets or databases, they should be checked to ensure they have been entered into the appropriate columns. They should also be checked for any impossible values, such as water temperatures above the boiling point, or negative concentrations. Where data are used to produce calculated values (e.g., total hardness from Ca/Mg analysis, anion/cation balances, conductivity to total dissolved solids (TDS) ratios) the calculation should have plausibility checks built in, such as per cent ion balance differences.

Data should also be checked for anomalous values (outliers), and it must then be verified whether they are justified (see section 6.2.1). They could be atypical values as a result of data entry or unusual field conditions, e.g., flood or drought. The goal is not to eliminate outliers, but to identify any potential data contamination. If an outlier is found, it must be further investigated and the findings, whether it is credible or not, should be flagged within the dataset. Therefore, there should be a column in the database where such findings can be documented.

Unfortunately, it is common within historic data sets to find that outlier (or unusual) values are present without any clear indication of the cause. In such circumstances, professional judgement may be required to decide how the data need to be treated. **Fig. 6.1** shows archived data from the Finn River, Ireland. It was suspected that some data had been reported as NO_3 rather than as N. The original data submitted could not be found by the submitting authority and only the archived data were available. Recalculating the suspect data points as mg l^{-1} N gives a data series that is more likely to have been representative of the years 2001–2003.

Figure 6.1 **Time series of total oxidised nitrogen (TON) for the period 1995–2016 for the Finn river, Ireland. The original reported mean values erroneously reported as Nitrate (NO_3) between 2001 and 2003 are shown in blue. The corrected values as Nitrogen (N) for the same years are show in black. Source: P. Webster**



6.2 Data quality

6.2.1 Outlier sources and detection

Table 6.3 presents some possible sources of outliers in a water quality data set, together with some options for how to deal with them. If the first possibility was definitely the cause, i.e., a simple typographical error, the corrected value should be kept because removing it would suggest no measurement was made or recorded. However, for the second possibility, i.e., the result was scientifically impossible, the value should be removed because including an erroneous value in the analyses will give invalid results. In the latter case, a note should be made explaining why the value was removed.

If the answer to all of the questions in Table 6.3 is "No", there are two remaining possibilities:

- the suspect value came from the same set of measurements as the other values, but was an extreme value, or
- the suspect value could have been due to an error, such as bad pipetting, holes in filters, etc.

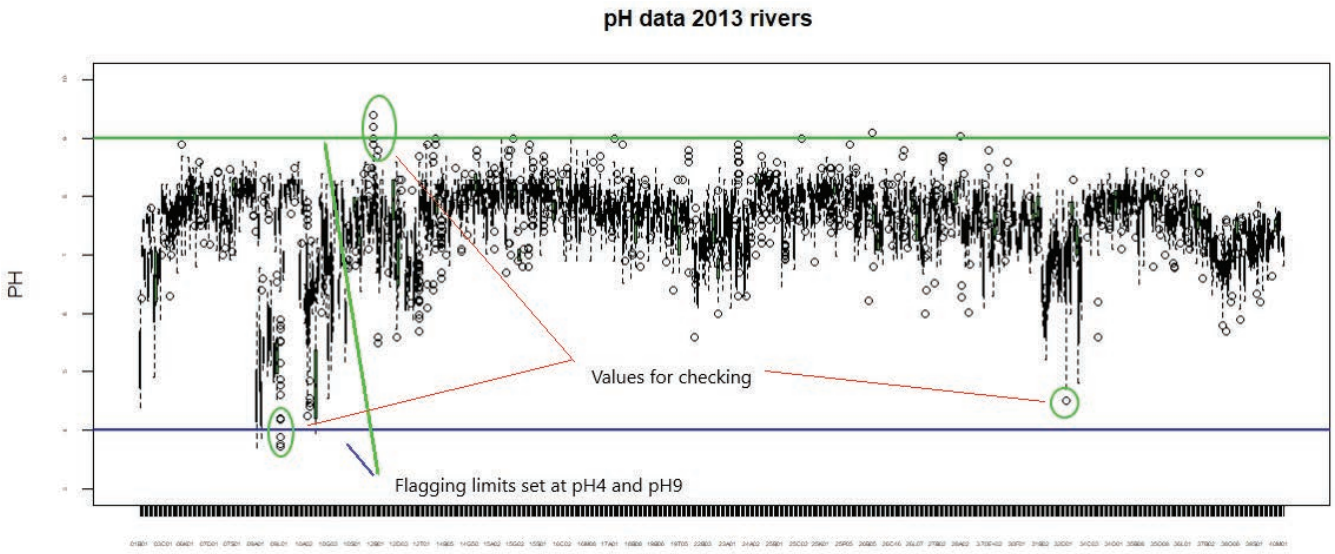
Grubbs' test (also known as the Extreme Studentized Deviate method) is one of the most popular ways to detect outliers (see ISO 2010a).

Using simple graphics can be very helpful in the process of data clean-up. **Fig. 6.2** shows a plot of a very large number of values for pH in rivers. As a result of the large number of monitoring stations used in this example to generate the chart (almost 2300 stations), it is impossible to identify individual stations. Nevertheless, it is apparent that there are some extreme values which can readily be found in the spreadsheet holding the data. From the metadata stored with the spreadsheet, it was determined that none of the extreme values were outliers and that there were genuine reasons for the unusual values — acidic mine drainage was responsible for the values around pH4 and samples taken downstream of an industrial discharge were responsible for the values above pH9.

Table 6.3 Possible sources of outliers in a water quality data set and some potential approaches to managing them

Question	Example	Response
Is the result an obvious typographical error?	pH 95 rather than pH 9.5	Correct it and note the original entry.
Is the result scientifically impossible?	Suspended Solids of -6 mg l^{-1}	Remove it or have it rechecked by the submitter.
Is the assumption of a normal distribution of the data dubious?	Skewed data	Retain but outlier tests like Grubbs test are less robust.
Is the outlier scientifically interesting?	An unusually polluted sample	Retain the data as an outlier until after evaluating whether the finding is potentially important.
Does the laboratory QC data suggest an analytical problem?	AQC bias failure	Flag the result and do not use it for data assessments.
Is there an Outlier policy?		Removing data should not be done in an *ad-hoc* manner.
Is there more than one outlier?	3.1, 3.5, **6.1**, 3.6, 3.4, **5.7**, 3.4, 3.1, **5.8**	Removal of outliers may mask a genuine problem. Possible seasonality may be indicated.

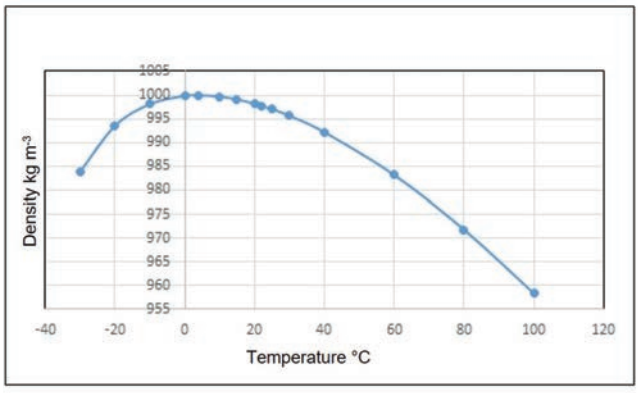
Figure 6.2 **Example of general "outlier" output for pH values (created with R) for more than 2,000 monitoring stations. Values outside the pH range 4–9 should be considered as suspect values (outliers) and checked. Source: P. Webster with data courtesy of the Environmental Protection Agency of Ireland**



6.2.2 Data relationships

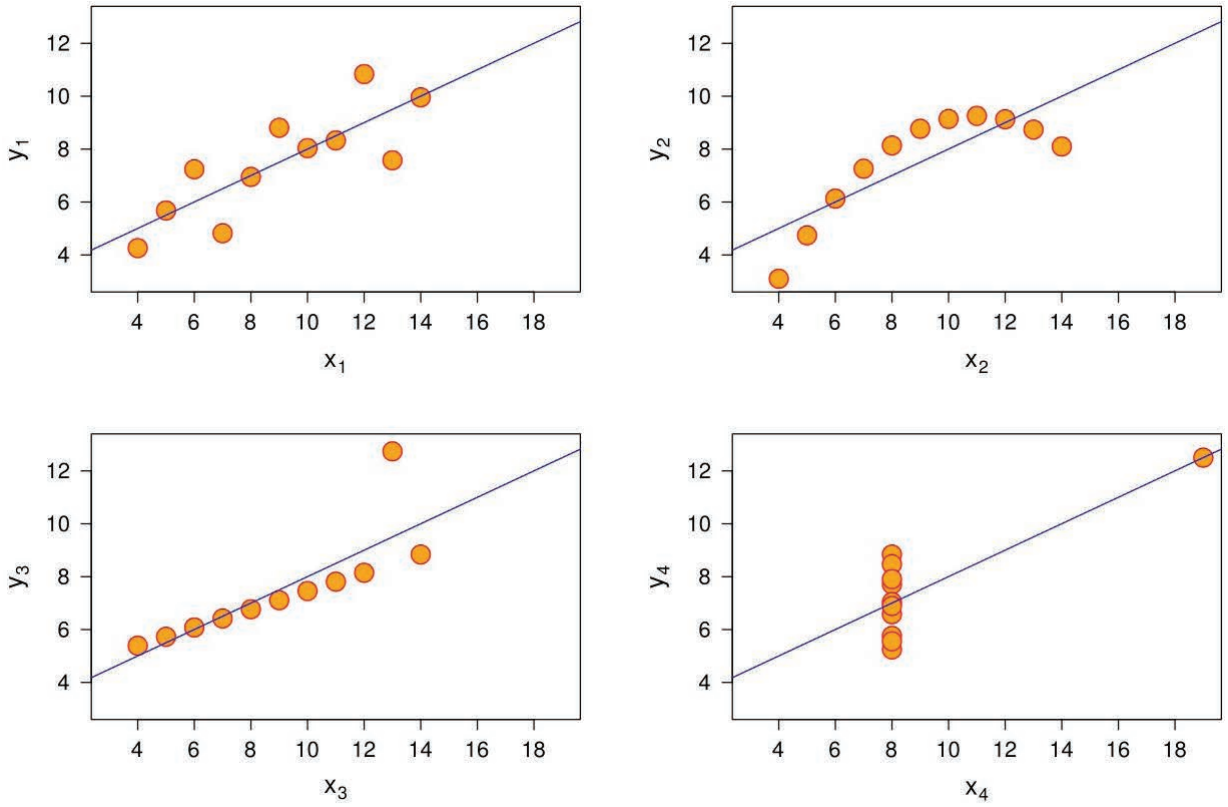
Unusual or suspect data may be apparent if there is a deviation from an anticipated or known relationship between two variables. There are essentially two ways to examine the relationship between two variables. *Correlation analysis* is suitable for situations where the relationship of one variable is not dependent on another. **Fig. 6.3** illustrates the correlation between freshwater density and temperature. In a *regression analysis*, the relationship between two variables may be one of functional dependence, i.e., one of the variables (the dependent variable) is a function of the magnitude of the second variable (the independent variable), while the reverse is not necessarily true. For example, a person's blood pressure might be a function of age, but age is not determined by blood pressure. Note, however, that age is not the only biological determinant of blood pressure, but it is just one determining factor. Hence, the term dependent does not necessarily mean that there is a cause-and-effect relationship between the two variables.

Figure 6.3 **Correlation of freshwater density with temperature.**



Linear regression techniques attempt to calculate the best fit and the correlation coefficient (r), which is usually expressed as its square (R^2). Thus, a correlation of 0.9 (strongly positive) will have an R^2 of 0.81. It is, however, essential to look at the scatterplot of the relationship. In the example shown in **Fig. 6.4**, all four plots have the same R^2 value, but they look totally different. Outliers can have a significant impact on regression calculations. In the case of Anscombe's Quartet (Anscombe 1973) shown in Fig. 6.4, all have the same calculated statistical properties but widely differing data plots. This example illustrates the necessity of examining data relationships visually and not simply relying on what would appear to be a good inter-relationship between the variables based on statistical properties.

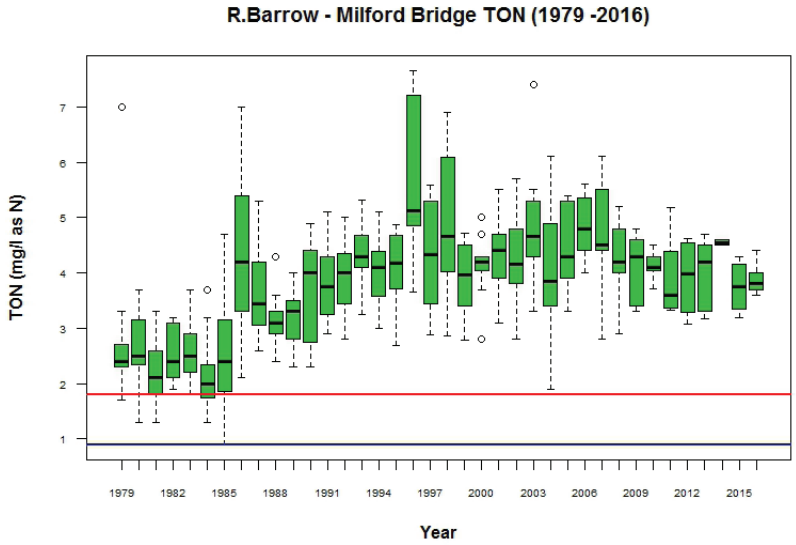
Figure 6.4 **Anscombe's quartet (Anscombe 1973) illustrates four datasets which are markedly different, but where the usual statistical properties (mean, variance, correlation and regression line) are the same. By Schutz. Licensed under CC BY-SA 3.0.**



6.2.3 Data trends

Plotting trends or time series can help in identifying what may appear to be unusual results. For a comparison of individual locations over time, and identification of outliers, the use of boxplots can be particularly useful (**Fig. 6.5**). For trend analysis, the freely distributed Finnish Meteorological Institute package "MAKESENS" (Finnish Meteorological Institute, 2002) allows the assessment of trends in stations using a combination of Mann-Kendall coefficient and Sen's slope. It provides both numeric data summary and graphical output. Other open access and free products include the Time Trends package developed by the New Zealand National Institute of Water and Atmospheric Research (NIWA undated).

Figure 6.5 **A boxplot of total oxidised nitrogen in the River Barrow, Ireland, 1979–2016. Outliers are indicated by small circles and horizontal lines represent guidelines for high (blue) and good (red) status. Source: P. Webster with data courtesy of the Environmental Protection Agency, Ireland**



CHAPTER 7

LABORATORY ACCREDITATION

In recent years there has been an increased drive towards laboratories becoming accredited to ISO 17025 (ISO 2017) for their test procedures. This approach is not only quality driven but also serves as a valuable marketing tool in the commercial water quality monitoring sector. Laboratory accreditation provides a third-party evaluation of the technical competence of a testing laboratory using internationally recognized processes and standards, and experts in the appropriate technical field. Some regulatory bodies have also now begun to specify that laboratories returning data to them must be accredited for the required data, such as drinking water quality data. ISO 17025 (ISO 2017) is the global quality standard that outlines all the requirements that testing and calibration laboratories have to meet if they wish to demonstrate that they operate a QMS, that they are technically competent, and that they are able to generate technically valid results. Chapter 1 of this guidebook referred to the preparation of a Quality Assurance Plan (QAP) containing Data Quality Objectives (DQO), the purpose of which is to set out in writing the overall approach to the issue of data quality for specific monitoring programmes. Much of this content forms the basis of the quality documentation required to achieve laboratory accreditation (**Box 7.1**).

Accreditation bodies are connected globally by the International Laboratory Accreditation Cooperation (ILAC), which provides consistency in the methods and processes used when accrediting laboratories. The assessment of the ability of a laboratory to produce precise and accurate data, involves the regular and documented assessment of the following:

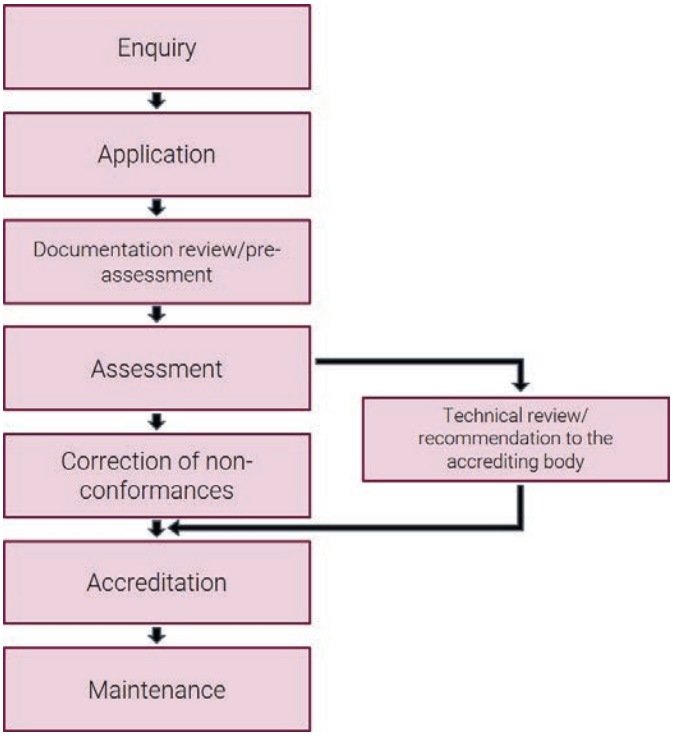
- Technical competence of staff.
- Validity and appropriateness of test methods.
- Traceability of measurements and calibration to national standards.
- Suitability, calibration and maintenance of test equipment.
- The testing or analytical environment.
- Sampling, handling and transportation of test items.
- Quality assurance of test and calibration data.

BOX 7.1 ESSENTIAL ELEMENTS OF THE QUALITY ASSURANCE PLAN AND DATA QUALITY OBJECTIVES DOCUMENTS THAT SUPPORT LABORATORY ACCREDITATION

- The scope of the monitoring programme in terms of spatial coverage and timeframe.
- Data to be collected, DQO and acceptance criteria.
- Staff training.
- Documents and records management.
- Sampling methods, together with sample handling and custody.
- Analytical methods, together with relevant QC and performance testing.
- Essential equipment.
- Assessment and response actions.
- Data review, verification and validation.

Each national accreditation body has its own specific documentation and guidance, but the process is usually similar and generally follows the scheme shown in **Fig. 7.1.**

Figure 7.1 Steps in a typical laboratory accreditation process



7.1 Audits

There is a difference between the processes of *assessment* and *audit*, in the context of quality assurance. *Assessment* is the evaluation process used to measure the performance or effectiveness of a system and its elements. Assessment can refer to any of the following: audits, performance evaluation, management system review, inspection or surveillance. An *audit* is a systematic and independent examination to determine whether quality activities comply with specifications and whether these specifications are implemented effectively. Quality system audits can be indicators of the adequacy and implementation of activities to achieve the objectives of a monitoring programme, the quality of the data generated, and compliance with the governing monitoring specifications.

Audits allow the laboratory to measure how well it is performing. Their purpose is to check that what is set out in the QAP is being done, and to identify any non-conformities or gaps in laboratory performance. It can also show whether the laboratory procedures and policies require revision or are not being followed. If the auditing process is not undertaken, laboratory operations can result in poor quality output. Effective auditing ensures continuous improvement and demonstrates that the laboratory is meeting accreditation, regulatory or customer requirements. An audit asks questions about the procedures and policies in place in the laboratory, how these procedures and policies are defined and whether they are being followed and comply with standards, regulations and requirements (**Box 7.2**).

Auditing is characterised by the reliance on a number of principles, namely confidentiality, fair representation, due professional care and independence. These principles should help to make the audit an effective and reliable tool in support of management policies and controls, by providing information on which an organisation can act in order to improve its performance.

Confidentiality relates to the security of information. Those undertaking audits should:

- Perform their work with honesty, diligence, and responsibility.

BOX 7.2 INFORMATION GATHERED DURING A LABORATORY AUDIT

Processes and operating procedures

Staff competence and training

Equipment

The laboratory environment

Handling of samples

Quality control and verification of results

Recording and reporting practices

- Observe and comply with any applicable legal requirements.
- Demonstrate their competence while performing their work.
- Perform their work in an impartial manner, i.e., remain fair and unbiased in all their dealings.
- Be sensitive to any influences that may be exerted on their judgement while conducting an audit.

Auditors should exercise discretion in the use and protection of information acquired in the course of their duties, including the proper handling of sensitive or confidential information. Audit information should not be used inappropriately for personal gain by the auditor or the audit client (auditee), or in a manner detrimental to the legitimate interests of the auditee.

Fair representation is the obligation to report *truthfully* and *accurately*, i.e., audit findings, conclusions and reports should reflect truthfully and accurately the audit activities. Significant obstacles encountered during the audit and unresolved diverging opinions between the audit team and the auditee should be reported. The communication should be truthful, accurate, objective, timely, clear and complete.

Due professional care is the application of diligence and judgement in auditing whereby auditors should exercise due care in accordance with the importance of the task they perform, and the confidence placed in them by the auditee and other interested parties. This may require the ability to make reasoned judgements in all audit situations. In some instances, auditors and those being audited may disagree on approaches to non-conformance resolution and the auditors should not expect unreasonable action to be taken unless this is critical to maintaining quality.

Independence is the basis for the impartiality and objectivity of the audit conclusions. Auditors should be independent of the activity being audited wherever practicable, and should in all cases act in a manner that is free from bias and conflict of interest. For internal audits, auditors should be independent of the operating managers of the function being audited but, for small organisations, it may not be possible for internal auditors to be fully independent of the activity being audited. Nevertheless, every effort should be made to remove bias and encourage objectivity. Auditors should maintain objectivity throughout the audit process to ensure that the audit findings and conclusions are based *only on the audit evidence* and not influenced by personal opinion. An evidence-based approach is the rational method for reaching reliable and reproducible audit conclusions in a systematic audit process. Audit evidence should be verifiable. It will, in general, be based on samples of all the information available, because an audit is conducted during a finite period of time and with finite resources.

7.2 Internal and external audits

Internal audits, conducted by laboratory staff themselves, are used to verify that the laboratory complies with ISO 17025 (ISO 2017), and with internal technical and quality procedures. This serves as excellent preparation for external assessments. They should be scheduled to occur regularly, and at least on an annual basis. They can either cover the whole laboratory and all elements of the quality system at one specific period of time, or they can be divided into several subsections. It is more common to examine selected elements, for example on a monthly basis, rather than the whole programme. The audit programme should be managed by the Quality Manager with findings related to the quality of test and calibration results being reported to customers. Follow-up activities should include relevant corrective and preventive action.

External audits can be conducted by groups or agencies from outside the laboratory, which may include accreditation bodies, licensing bodies, regulatory bodies, or clients in some cases. They are designed to identify whether the procedures and processes are being followed in the laboratory and whether they are in compliance with current documentation. They also identify whether written policies and procedures comply with standards, regulations, and requirements. They need rigorous planning but might occasionally occur without prior notification by the accreditation body!

The audit process can be broken down into the following seven steps. Regardless of whether the audit is internal or external, the steps remain the same:

1. Planning
2. Review Documentation
3. Opening Meeting
4. Inspections and Interviews
5. Summary Report
6. Closing Meeting
7. Follow-up

Proper planning is required to ensure the availability of staff and resources. All relevant documentation should be available for inspection and review. For internal audits the level of formality may be reduced, such as in the opening and closing meetings, but good communication between the auditor(s) and the project management personnel is essential at all stages. The opening meeting will set out the inspection plan. Following assessment of planned activities and documents, the auditor(s) should prepare a summary report setting out their findings and the evidence for any non-conformances. These are generally classified into three categories: (i) observations, (ii) minor, (iii) major. At the closing meeting, the auditors and quality staff should discuss the audit findings and agree on a timeframe for their remediation. All non-conformances should be addressed and the relevant processes re-audited were necessary to assess the effectiveness of corrective actions.

7.3 Proficiency testing

Proficiency testing is specifically designed to assess the competence of a laboratory and uses inter-laboratory comparison to do this. The key objectives of proficiency testing can be summarised as:

- To provide an impartial evaluation of individual laboratory performance.
- To identify problems with test methods.
- To assess the comparability of test methods.
- To facilitate the validation of measurement uncertainty claims.

The ISO standard for proficiency testing is ISO/IEC 17043 (ISO 2010b), and many national schemes are modelled on this standard. The most common approach to testing is to use a split-sample design whereby each laboratory receives an aliquot of the same bulk sample for analysis and reports their data to the coordinating authority. The performance of the laboratory is generally reported as a "Z-score" which ranks the variation of their reported data relative to the reference or expected value. The closer to zero, the better. The statistical approaches are outlined in ISO 13528 (ISO 2022). The results of proficiency testing are used to identify poor accuracy, describe consistent bias, and to verify the effectiveness of corrective actions.

CHAPTER 8

QUALITY ASSURANCE OF BIOLOGICAL AND MICROBIOLOGICAL WATER QUALITY MONITORING

Quality assurance in water quality monitoring is usually assumed to be an exercise applied to laboratory analysis of water samples. The same principles of quality assurance should be considered and applied to biological approaches to water quality monitoring, particularly ecological methods that involve the collection of biota and their identification and enumeration. Biological approaches to monitoring water quality are described in detail in a companion guidebook on *Freshwater Quality Monitoring with Biota*.

This chapter provides a brief summary of protocols for the quality assurance of the identification of stream macroinvertebrates and common aquatic microbiology. It focusses on identifying where quality issues may arise when using an ecological approach to monitoring water quality using benthic macroinvertebrates. It then goes on to suggest ways in which quality may be assured for the collection and identification of biota and suggests some quality assurance procedures for sampling and analysis activities for microbiological assessment of water quality.

8.1 Biological field work: stream sampling

Benthic macroinvertebrates are visible to the naked eye and are usually defined as organisms that are retained by a net or sieve of mesh size 0.6 mm. They fulfil quite well the criteria for use in biological monitoring of water quality, and the equipment needed is usually simple and inexpensive. Sampling of benthic macroinvertebrates is generally done in shallow streams, or riffle sections of rivers and streams, using a kick-sampling technique just upstream of the sampling net. Sampling and riverbank assessment at a single site can generally be completed in about 20–30 minutes. Organisms caught in the net must be sorted from the debris, such as leaf and woody fragments (**Fig. 8.1**). Guidance on sampling is available from ISO (2012b) and national guidance documents such as Cavanagh, Nordin and Warrington, (1997). The time and effort for the kick sampling should be standardized between sites and sampling occasions; 2–3 minutes is generally adequate. Note

Figure 8.1 **Unsorted kick sample collected from a river. © Patrick Cross**

that the kick effort, and therefore the number and variety of organisms dislodged, could vary from one sampling technician to another.

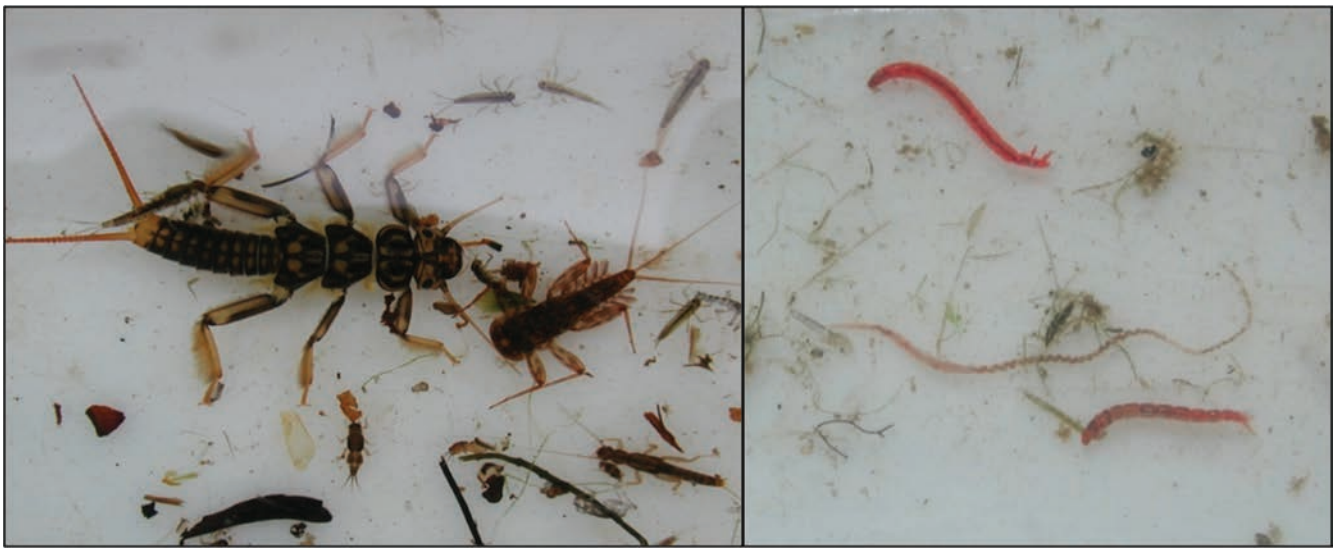
Accuracy and comparability of biological sampling must be ensured by thorough training of all field personnel or, preferably, by using trained biologists to collect the samples. Without thorough training, it is easy to miss certain types of organisms, such as the cased caddis larvae which can resemble woody debris. This could bias the result of the biological assessment method. Ideally, an experienced invertebrate biologist should sort, identify and count the organisms present. **Fig. 8.2** shows two sorted samples. The predominance of two and three tailed larvae indicate good water quality, whereas in the sample from a poor quality river they are notably absent and the dominant species are worms.

The success of any biological monitoring programme depends on the use of taxonomic keys for species identification. There are a range of these available, some of which are specific for a country or region, such as Pawley *et al*. (2011), and it is important that the most relevant keys are used.

8.2 Auditing a biological monitoring programme

Quality control is not widely implemented for biological monitoring programmes because of the complexity and high cost associated with the intercalibration of biological systems. Identifying errors is often not as straightforward for biomonitoring as it is for

Figure 8.2 **Macroinvertebrates in two different samples from rivers. The left shows species associated with good water quality and the right shows species from poor water quality. © Patrick Cross**



other types of analyses, such as chemical analyses, where statistical techniques are used to measure objective values. Applying quality control to biological monitoring based on species collection, identification and counts is challenging. It is difficult to evaluate whether differences in the results obtained are significant, because identifying species is often subjective and depends on the individual technician processing the sample. Despite the difficulties, it is valuable to know how operator-dependent differences in sample collection and identification can affect the outcomes of a stream water quality assessment.

Human error potentially affects all stages of freshwater biomonitoring, including site selection, field sampling, sorting, identification, data entry, analyses and interpretation (Clarke and Hering, 2006). Currently, there is little understanding of the impact of human error on a monitoring programme because few suitable auditing schemes exist. Haase *et al*. (2010) audited the macroinvertebrate samples collected from the official European Union Water Framework Directive (EU WFD) monitoring sites in Germany. There were two main components to the audit performed by Haase *et al.* (2010):

i. *Sorting Audit*. This aimed to detect specimens that remained in the sample residue. Field personnel were instructed to remove all individuals from sample material and to retain the residue separately. Auditors then re-sorted the whole sample residue, removing any individuals found and placing them in a new labelled vial. The auditors then counted these species and identified them to taxonomic level and compared the number of individuals, number of taxa and assessment results for the original list generated by field staff with the results after the sorting audit. The sorting audit was assumed to represent 100 per cent of the organisms in the sample.

ii. *Identification audit*. Field personnel were instructed to retain up to five specimens for each identified taxon, which is common practice in Germany. Voucher specimens were then re-identified by the auditor.

The study found that during biological monitoring, human error was more prominent than previously thought. The study suggested that the high level of error in sorting and identification is a general pattern that is not only applicable to a river system, but that such errors would also occur during the sampling of lake ecosystems.

8.3 Microbiological quality control

Microbiological analysis is another key element of water quality assessment, both for public health protection and for pollution assessment. This section provides an insight into the complexity of this type of analysis and recommends some procedures to ensure that high quality results are obtained. Hands-on training and further reading are, nevertheless, an absolute necessity.

Microbiological analysis is, in many ways, more complex than chemical analysis because it relies extensively on eliminating any cross contamination. Most laboratory staff can potentially undertake this type of analysis, but it requires specific equipment and specialist training. Necessary supplies for sampling are listed in **Box 8.1**. Sampling procedures must be clearly documented, and country-specific guidance may be available from national water agencies or regulators to assist with this. For rivers, streams, lakes,

BOX 8.1 NECESSARY SUPPLIES FOR MICROBIOLOGICAL SAMPLING

General requirements:

- Disposable rubber gloves.
- Sterile sample bottles.
- Cool box with gel ice packs for transporting samples.
- Water resistant marker to label bottles.

Additional requirements for the collection of drinking water samples from distribution pipes:

- Alcohol wipes or a wash bottle of 70% Industrial Methylated Spirits (IMS).
- Sodium thiosulphate for water samples taken from chlorinated supplies.

reservoirs, springs or shallow wells, the following procedures should normally be followed:

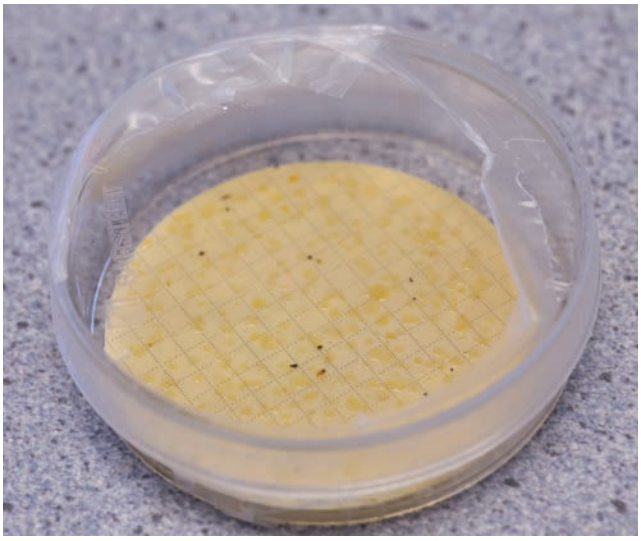
- An unopened sample container is labelled with date, sample location and sample time before proceeding to the sample area.
- Before the sample container is opened, the technician should put on sterile gloves and they should only open the sample container just before taking the water sample.
- The sample bottle should be held by its base and plunged into the water source with the neck facing down. The inside of the bottle cap should not be touched or the cap left on the ground. Samples should be representative of the water body and, therefore, it is undesirable to take samples too near the bank or too close to the sediment.
- The bottle is then turned until the neck is pointing slightly upward and the mouth is directed towards the current (if any), allowing the bottle to fill. After filling, the bottle is removed from the water with the neck pointing up (there should be no air gap).
- The cap is secured tightly on the bottle, which is then placed immediately into a cool box with sufficient ice packs to keep the sample cold during transport to the laboratory within 24 hours.

When sampling drinking water, the bottle is filled directly from the water source, i.e., distribution network, storage tank, reservoir, or pump. The tap should be disinfected with an alcohol wipe or by spraying with IMS and allowed to run, for at least 3 minutes or for sufficient time to ensure all of the water standing in the pipe is flushed through, before filling the sample bottle.

8.3.1 Microbiological analysis methods

The test methods used for microbiological analysis depend on the microorganisms being examined and any stipulations for the method provided by regulatory agencies. The three main techniques are:

Figure 8.3 **Colonies of *E. coli* (yellow patches) filtered from a water sample and growing on a filter membrane on a culture plate. © Patrick Cross**



Membrane filtration and culture on selective media. A portion of the well-shaken sample (typically 100 ml) is filtered through a sterile 0.45 µm pore size membrane, which is then placed onto selective media in a petri dish and incubated. Colonies which grow and are typical of the test organism are counted, giving results as CFU/100 ml (Colony Forming Units) or Number/100 ml (**Fig. 8.3**).

Multiple tube assays. Replicate serial dilutions of the sample, typically 1:1, 1:10, 1:100, are prepared and incubated in a selective media solution. Those showing growth at each dilution are counted and the Most Probable Number (MPN) of organisms determined from special tables of MPN.

Commercial multiple well assays. These are an adaptation of the MPN approach, providing MPN from counts of wells showing growth. Some use biochemical reactions to aid enumeration for specific species, e.g., for *E. coli*. (**Fig. 8.4**).

Methods and procedures must be followed precisely because even small deviations from the method

Figure 8.4 **An example of a commercially available multiple well assay tray for *E. coli* © Jean O'Dwyer**

could affect results, such as incorrect incubation temperature or inappropriate culture medium.

8.3.2 Equipment

Equipment such as autoclaves, ovens, water baths and refrigerators are common to many microbiological methods. Thermal equipment such as ovens, incubators and autoclaves need to be calibrated for temperature variations in accordance with international standards. It is important to check for any potential hot or cold spots that could affect the microbiological assay or sterilization process. For temperature stepping ovens, it is particularly important to verify that the temperature ramping programme is operating correctly. Temperature-critical pieces of equipment should be monitored using temperature probes, data loggers, or temperature sensitive indicator strips (as in the case of autoclaves) to ensure correct operation.

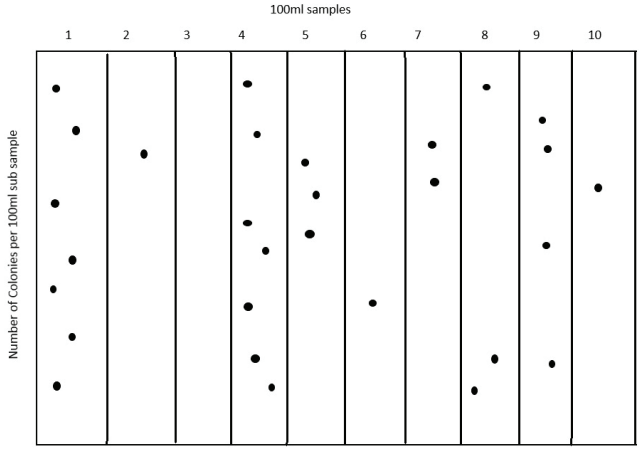
More general equipment such as micropipettes should be checked for accuracy and potential for cross contamination (see section 4.3). Sterility checks should be made on work surfaces to minimize the likelihood of contamination.

There are many different types of culture media and the instructions for their preparation provided by commercial suppliers must be followed very carefully. Some substances that are used can be harmful to health, so suitable safety precautions should always be taken. Batches of media should be checked for suitability before use, using both the test organism for which the media is suited (positive controls) and also to ensure that no competing organisms can grow in it (negative controls). Storage conditions vary but all media, once prepared, should have their "use-by" date recorded.

If glassware is washed in a laboratory glass washer that has acid/alkali rinse cycles, the pH may need to be checked to ensure it is compatible with the growth conditions for the microorganisms being cultured.

In addition to providing microbiological media, filters and other equipment, many commercial suppliers also provide easy-to-use materials for quality control purposes. Among the range of products available are quality control CRMs. The use of these materials allows laboratories to ensure the quality of their analyses in much the same way as for chemical analyses. Although they are quite expensive, they are manufactured to ISO standards to ensure their quality. Their use is a key requirement of laboratory accreditation under ISO 17025 (ISO 2017) (see Chapter 7).

Figure 8.5 **Theoretical distribution of the number of colonies of microorganisms (y-axis) in 10 different samples of 100 ml each (x-axis). The average number is 3 colonies per 100 ml.**



8.3.2 Quality control

One of the fundamental differences between the quality control of chemical analysis and that of microbiology methods, is that microorganisms are seldom, if ever, uniformly distributed in the sample. In the example shown in **Fig. 8.5** where each division is a discrete 100 ml sample, the overall average number of colonies is 3 per 100 ml, but the range is from 0 to 7 per 100 ml.

The assessment of characteristics, such as limit of detection, repeatability etc. are just as important in microbiological analysis as for chemical analysis and, as with chemical analysis, Shewhart charts can be used for microbiological process control (see Chapter 5). In these cases, the warning and action lines are more generally referred to as "response lines" because many more microbiological samples are required to demonstrate "out of control" situations. Jarvis (2016) provides examples of how control charts can be used in microbiological analyses.

CHAPTER 9

METHOD DEVELOPMENT AND PERFORMANCE TESTING

This chapter introduces commonly applied approaches to selecting and developing analytical methods and for evaluation of the performance of the methods and laboratory procedures. It describes some potential sources of misunderstanding that might influence method selection and recommends considerations for analytical method selection in order to achieve high quality results. The steps and associated activities involved in method validation are described.

9.1 Defining the parameter and method selection

It is often the case that a decision is made to monitor a substance that can occur in different forms. If it is unclear exactly what is needed, and which form is to be monitored, the laboratory may choose an inappropriate method. This problem arises from the day-to-day use of "common" terms for substances which, although broadly understood, are not sufficiently detailed for laboratories to interpret client requirements (see **Box 9.1**). As many of these tests are method-specific, it is important to check with the person requesting the analyses precisely what they mean and what they expect.

Several standard methods are available for most of the analytical determinations involved in water quality monitoring. These standard methods frequently include extensive validation data that allow them to be easily evaluated, and many are sanctioned by appropriate national and international organisations. The choice of method is critical to ensure that the results of the analyses meet the requirements of the monitoring programme objectives, because different methods have different precisions and sensitivities, and are subject to different potential interferences. Before selecting an analytical method, there are a number of key questions that must be answered and that assist with making the right choice (**Box 9.2**).

For most of the commonly analysed substances in water, there is a wide choice of suitable published methods available, such as those published by the ISO and the European Committee for Standardization (CEN). Many ISO standards have been adopted by national, state or provincial standards bodies which publish country, state or province-specific standards. In addition, there are organisational methods, such as those of the American Public Health Association (APHA), the American Water Works Association (AWWA), and the Water Environment Federation (WEF) (Rice, Baird and Eaton 2017). Regulatory bodies usually require the use of approved "standard methods" which they specify, or for which they produce in-house documented procedures.

Whereas client or customer requirements may be specified for particular substances, laboratory-developed ("in-house") methods may be required for novel substances, such as trace organics. In-house methods or "modified" standard methods should be used with caution and should be fully assessed for their performance before use.

BOX 9.1 EXAMPLES OF POTENTIALLY CONFUSING TERMINOLOGY FOR WATER QUALITY SAMPLE ANALYSIS

Physical parameters: e.g., conductivity – but at what reference temperature 20 °C or 25 °C? There can be a variation in values of about 10% between the two temperatures.

Empirical parameters: e.g., chemical oxygen demand and biochemical oxygen demand – these are defined by the analytical methodology used.

Grouped parameters: e.g., "Total Phenols", "Total Pesticides" – what is meant by the term "Total" in this context?

Method specific parameters: e.g., volatile organic compounds (VOCs), diesel range organics (DRO), etc.

Cyanide Does this term mean free or readily dissociated such as potassium cyanide (KCN), combined such as potassium thiocyanate (KSCN), or complexed such as potassium ferricyanide $K_4Fe(CN)_6$? "Free" dissociated cyanide can be measured by colorimetry but distillation from acid solution is required to measure the parameter "total cyanide" in the sample. Non-distillation methods can be subject to interference from other species in the sample e.g., sulphide?

Phenols Does this mean phenol (C_6H_5OH) itself, "Phenol Index" (as measured by extraction 4-AminoAntiPyrene), or groups of possible substances defined by test procedures such as the US EPA methods (US EPA 1984, US EPA 2000)?

Total heavy metals The International Union for Pure and Applied Chemistry (IUPAC) has no specific definition of a "heavy metal". It can be characterised by molecular weight, but it is more widely deemed to include substances such as molybdenum, tungsten and uranium. It is essential to confirm the metals and metalloids that are required to be analysed in the sample.

Hydrocarbons These are often characterised by very loose terminology which is linked to the number of carbon atoms in their structure, e.g., PRO – Petrol Range Organics are typically defined as those from C_6 to C_{12}; DRO – Diesel Range Organics as C_{12} to C_{20}, and Mineral Oils as $>C_{20}$. These can vary from one laboratory to the next and therefore need to be explicitly defined.

Detergents (anionic, cationic, non-ionic) These are defined by the procedure used, for example anionic detergents are usually measured by reaction with Methylene Blue dye, hence the term Methylene Blue Active Substances (MBAS). There can also be a choice of reference substances, such as sodium dodecyl sulphate (lauryl sulphate). Other reference substances will give a different numeric output and so the results are seldom comparable.

BOX 9.2 KEY QUESTIONS THAT ASSIST A LABORATORY ANALYST IN SELECTING THE CORRECT ANALYTICAL METHODS

What actual substances are to be measured?

Will they be measured as 'Total' or as 'Dissolved / Filtered'?

Will samples need to be preserved and, if so, how and with what?

What is the expected concentration range of the variable?

How are they to be measured and reported, e.g., mg l^{-1} Nitrate or as mg l^{-1} N?

What limit of detection is required? (Usually a tenth of any regulatory value).

What is the accuracy required for the measurements?

What is the expected frequency of the testing?

Will the analysis be done on-site or in a laboratory?

What range of analytical instrumentation is available?

Will there be just a few samples or enough to enable grouping them into batches for instrumental analysis?

Do results need to be accredited or not?

What turnaround time is required?

What is the overall cost per sample?

9.1.1 Test kits

An increasing number of water variables can now be measured using commercial test kits. These offer quick assessment of water quality but can lack sensitivity. Examples of standards for commercial kits are given in ISO (2003b) and BSI (2009). Test kits can be convenient for use by non-scientific staff, e.g., citizen scientists, and very useful for pollution investigation sampling. They may require no (or little) reagent preparation and can also be very cost-effective for *in situ*, or even laboratory analyses, depending on the degree of sensitivity required. There are limitations that should be taken into consideration before deciding to use them in a monitoring programme and the quality of the instructions provided varies significantly between different kits. Some kits may suffer from interferences due to colour or turbidity in the water samples, and potential interferences in the analyses may not always be well documented. In some cases, the degree of resolution between concentration values given in discrete steps may be too wide for specific uses. The limitations of the kits need to be appreciated and understood by both the user and the "client" who has requested the measurements.

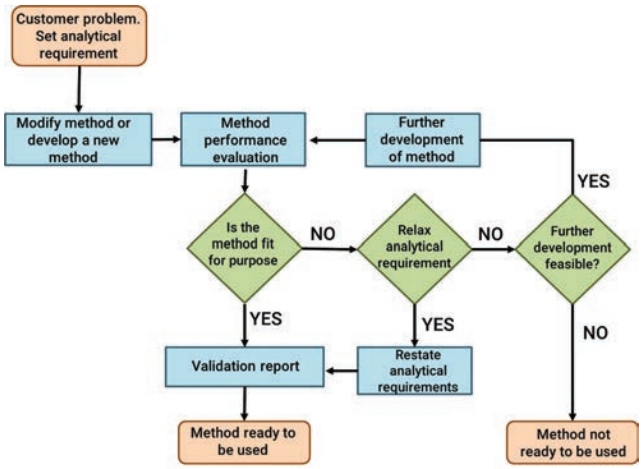
9.2 Method validation

Validation is a process through which it can be determined whether the chosen method is fit for purpose, whether it does what is expected, and provides information on its reliability. Fitness for purpose can be described as the degree to which data produced by a measurement process enables the user to make technically and administratively correct decisions for a stated purpose. In order to be fit for purpose, the performance characteristics should be adequate to meet the needs of the "client" or the Regulator, as appropriate, the method should be specific to the determinand of interest, and the interferences should be known. The performance characteristics and interferences should be determined by evaluating the actual performance of the method in the laboratory rather than by using those stated by the supplier or manufacturer.

The method validation process from the customer problem to the laboratory decision on whether or not the customer request can be carried out with an identified method is shown in **Fig. 9.1** (Magnusson and Örnemark 2014). The key steps for method validation and development are:

- Review existing literature and published standard methods for the determinand.
- Evaluate the applicable or anticipated concentration range.
- Select the performance criteria needed and write a validation protocol of how the method will be tested.
- Test the method for linearity, accuracy and bias, precision, per cent recovery, sensitivity, selectivity, and other specific criteria such as peak shape or chromatographic resolution etc., as required.
- Calculate the limit of detection (LoD), which is often referred to as the method detection limit (MDL) and also the limit of quantitation (LoQ), which is a multiple of the LoD (see section 9.2.3 below).

Figure 9.1 **The analytical method validation process. (after Magnusson and Örnemark 2014)**



If the method meets the validation performance criteria, quality control criteria can be established using stock standards or other suitable reference materials, and the measurement uncertainty and other performance criteria can be assessed (see Chapter 5). If the method does not meet the validation criteria, the criteria that do not meet the required targets, together with the intended targets, have to be reviewed. If possible, this should be done in consultation with the "client". If the method is still unsuitable, it will be necessary to choose an alternative approach.

Microbiology poses its own separate issues when it comes to validation of test methods. The validation methods are geared towards ensuring that the test method cultures and identifies the test organism and no other related microorganisms. Validation should assess the impacts of factors such as media (both media choice and the reagent supplier), pH, incubation temperatures, and recovery, using controlled counts of test organisms, etc. Further information is available in Association of Official Analytical Collaboration (AOAC) International (2016).

9.2.1 Linearity

Linearity is checked using a series of stock standards (in distilled or de-ionised water) covering the expected analytical range. It is preferable to undertake a specific calibration response for each batch of analysis. For very stable systems it may be sufficient to check the linear response and sensitivity using both low and high range standards e.g., 20 per cent and 80 per cent of the linear range and reviewing their response against predetermined criteria, e.g., peak area counts.

The ideal calibration curve is one which is linear within the most useful range, with a regression coefficient of 0.95 or better. It is not unusual, however, to see some curvilinear response at both low and high concentrations. Data should generally be reported only if they are in the linear range of the calibration. If the calibrations do not show a linear response, $\log_{10}$ transformation of the data may improve the linearity but, if not, it will be better to select an alternative analytical method that does show a linear response over the range of values of interest. **Box 9.3** shows an example of a linearity check for the colorimetric determination of Ammonia in wastewaters.

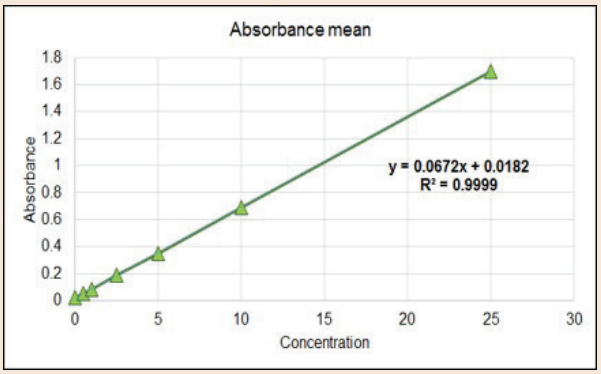
BOX 9.3 AN EXAMPLE OF A LINEARITY CHECK FOR THE COLORIMETRIC DETERMINATION OF AMMONIA IN WASTEWATERS (0.5–25 MG L-1 N RANGE)

Table 9.A Results from the analyses of solutions with a range of concentrations of ammonia

Concent-ration	Absorbance Reading 1	Absorbance Reading 2	Absorbance Mean
0	0.02	0.028	0.024
0.5	0.054	0.053	0.0535
1	0.085	0.085	0.085
2.5	0.186	0.187	0.1865
5	0.345	0.347	0.346
10	0.688	0.692	0.69
25	1.69	1.712	1.701

Figure 9.A Graph showing the linearity check between the concentration of ammonia (mg l^{-1}) N and the mean of two absorbance readings per concentration

In this example, the method is clearly linear up to the top standard of 25 mg l^{-1} N.



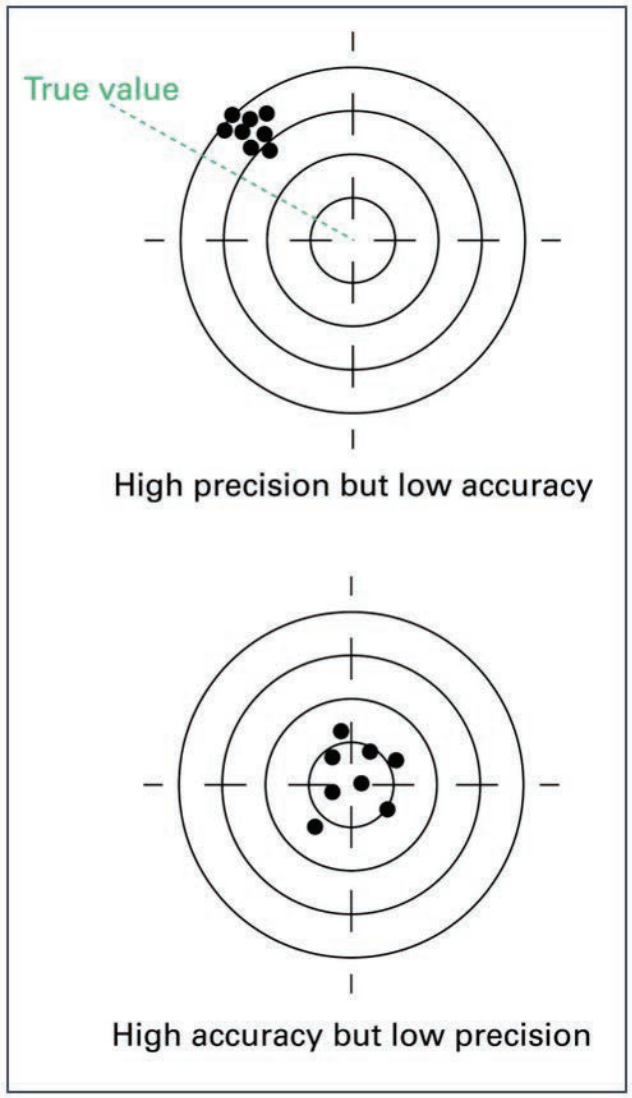
Source: P. Webster

9.2.2 Accuracy and precision

Accuracy is defined as the closeness of the measurement to the "true" concentration (ISO 1994). When the term is applied to sets of measurements of the same analyte, it involves a component of random error and a component of systematic error. In this case, trueness is the closeness of the mean of a set

of measurement results to the actual (true) value and precision is the closeness of agreement among a set of results (**Fig. 9.2**). Accuracy can be determined by the analysis of reference materials with known concentrations of the variable, or by comparing analyses with existing methods in other laboratories where possible (see section 7.3). Deviation from the "true" measurement is referred to as bias. Accuracy and bias can be determined by analysis of upper and lower standards and assessing them against the validation targets. In the example of bias check in **Table 9.1** the bias should be <10%, and therefore the method shows no significant bias, although at lower concentrations some improvement may be desirable.

Figure 9.2 **Accuracy is the closeness of the measured values to the true value (centre). It is possible to have high accuracy but low precision where there is great variability between measured values**



Precision is a measure of the variability associated with the measurement (Fig. 9.2). The more closely grouped, the more precise the method. Precision is about identifying whether the test method can be reliably repeated by different analysts at different times. Within-day, and between-day, coefficients of variation should be performed at three concentration levels.

Precision is generally assessed by the replicate analysis of a blank, low standard (about 10% of range), high standard (ca. 80% of range), low concentration real sample and spiked real sample (to about 50% range). These are analysed in up to 10 batches and the performance evaluated.

9.2.3 Limit of Detection and Limit of Quantitation

The Limit of Detection (LoD) is the lowest concentration of the variable that can be distinguished from zero with 95% confidence (ISO 2009). This is defined as:

$$LoD = 2\sqrt{2}.t\,(df, \alpha = 0.05).s_w$$

Table 9.1 **Checking for bias in an analytical method using results from two standard solutions**

Concentration mg l^{-1}	Mean Result mg l^{-1}	Absolute bias	Bias %	Outcome
5	4.809	-0.191	-3.962	Pass
20	20.1	0.1	0.498	Pass

Source: Data from P. Webster

Where t denotes the Student's t-test value and s_w denotes the within-batch standard deviation.

For a one-sided test with 11 degrees of freedom, $t = 1.796$ and this transposes as LoD = $5.08s_w$. In practice, the LoD is generally taken to be a minimum of three times the standard deviation of repeated blank sample outputs as in ISO 13530 (ISO 2009). In the Total Nitrogen assay worksheet (Appendix C), the blank standard deviation (S_t) was 0.046. On this basis, the LoD would be defined as 0.15 (to the nearest two figures).

The Limit of Quantitation (LoQ) is the practical reporting limit and is generally taken to be the lowest concentration of the variable that can be reliably quantified subject to defined accuracy and precision criteria. A factor of $10s_w$ is commonly used. In the Total N example in Appendix C, this would equate to $0.046 \times 10 = 0.46$. This would be rounded to 0.5 mg l^{-1}. Measured values less than 0.5 should be reported as <0.5 mg l^{-1}. In practice, it is common for laboratories to apply and report values using the LoQ as the limit. They may (if requested) also report the measured value but this should not be considered reliable and should not be used for summation purposes as it has a high degree of uncertainty.

9.3 Treatment of non-detects

One of the more challenging issues in data management is how to treat results which are below the LoD or LoQ. Data below a detection limit (DL) are referred as left-censored data. There could also be multiple detection limits involved if an instrument is upgraded during the project period or data are combined from multiple sources or laboratories. Although values below the detection limits complicate the analysis of the water quality data, they can still be of importance because, for example, they may be used in assessing a potential health hazard. Various strategies have been developed to analyse data that fall below detection limits. Simple deletion of data is not desirable in any situation and, in general, the most common practice is to substitute the <LoD value with a fixed value, such as zero, ½ DL, 1/√2DL or the DL itself. The European Commission Directive 2009/90/EC (European Commission 2009) on the technical specifications for chemical analysis and monitoring of water status advocates the use of ½ DL.

Simple substitution methods are easy to apply but inevitably cause some degree of bias particularly at high censoring levels such as >20%. When the results strongly depend on the values being substituted, particularly for data with multiple detection limits, then substitution methods are not generally suitable and, if used, the resulting outcome should be interpreted with caution. A more complex approach is to use an interpolative method such as Regression on Ordered Statistics (ROS). This treatment uses probability modelling to assign quantitative values to non-detects. An extensive review of this and other methodologies is set out in Helsel (2012).

9.4 Per cent recovery

For those test methods which rely on conversion of one form of a substance to another for analysis (e.g., in the determination of Total Nitrogen or Total Phosphorus), recovery is a measure of the degree of conversion. It is generally expressed as a percentage value and is based on the assessment of substances that need to be converted before analysis, e.g., organic phosphorus compounds to ortho-P. In general, recovery should be > 90%. The approach can also be applied to methods such as solid phase extraction techniques to assess the extent to which a determinand can be extracted from the test sample.

It is important when assessing recovery that potential interferences are taken into account in the assessment, such as in the determination of Total N by persulphate digestion, because per cent recovery is seriously reduced in the presence of high concentrations of COD.

9.5 Resolution, peak asymmetry and selectivity

For most chemical analyses the approaches to validation outlined above would meet the requirements of ISO 17025 accreditation (see Chapter 7). However, for chromatographic separation

techniques such as GC or HPLC, additional criteria such as resolution, peak shape (asymmetry), and selectivity are important. Acceptable limits should be set for all these metrics in the analysis protocol.

In chromatographic methods, *resolution* is the degree to which one analyte can be distinguished from another. The resolution between the two chromatographic peaks shown in **Fig. 9.3** is R_{AB}, and is a quantitative measure of their separation. It is defined as 2 × the time difference between the peaks (Δt) divided by the sum of the peak widths ($w_A + w_B$). As R_{AB} increases, the separation of the peaks improves (Fig. 9.3). Because resolution is a quantitative measure of the success of a separation, it is a useful way to determine if a change in experimental conditions leads to a better separation. This may require adjustment of chromatographic conditions such as temperature, column polarity, solvent ratio, etc.

Peak asymmetry (shape) is determined by setting criteria for peak asymmetry (peak tailing) and retention time (how long the component takes to elute). This is usually assessed relative to a reference substance. Where peaks are significantly mis-shapen, an alternative column or elution matrix may need to be applied. Ideally the peak shape metric A_s should be as close to 1 as possible (**Fig. 9.4**).

Figure 9.3 **Schematic representation of the relationship between resolution and the separation of a two-component mixture. The green peak and the red peak are the elution profiles for the two components, A and B. The chromatographic peak, which is the sum of the two elution profiles, is shown by the solid black line.**

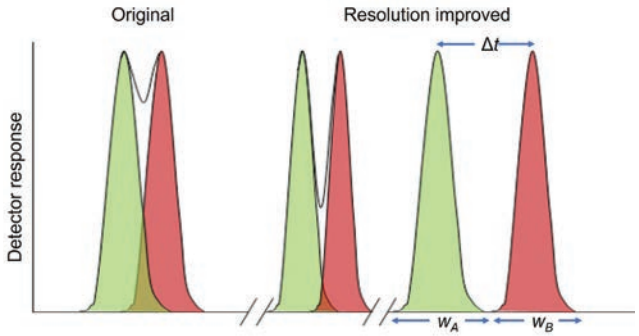
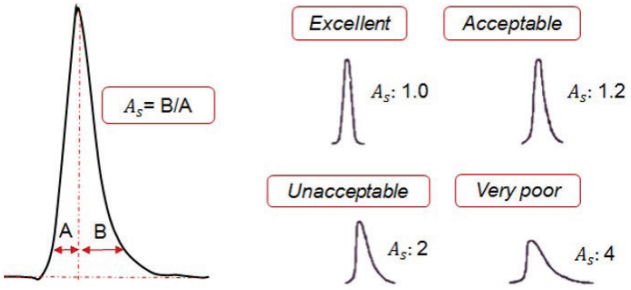
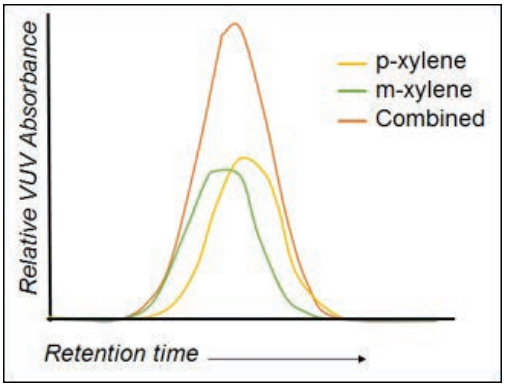


Figure 9.4 **Different values for peak asymmetry (A_s), from excellent (1.0) to very poor (4).**



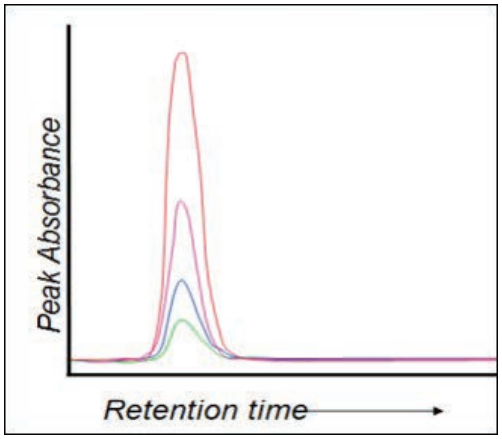
An analytical method is specific if its signal responds only to the analyte. Although specificity is the ideal situation, few analytical methods are completely free from the influence of interfering species. The *selectivity*, or separation, is the ability of a chromatographic system to "chemically" distinguish between sample components (**Fig. 9.5**). Selectivity is a measure of the freedom of a method from such interferences.

Figure 9.5 **In chromatographic analysis, m-Xylene and p-Xylene often coelute making it difficult to quantify each separately.**



Sensitivity is generally defined by setting a minimum peak area (or peak height), or count, for a control standard. This response factor (i.e., the sensitivity) is the ability to distinguish between small increments in concentration. The larger the signal to concentration ratio, the more sensitive the method (**Fig.9.6**). As with conventional colorimetric methods, the method used should display linearity across the selected measurement range.

Figure 9.6 An example of good sensitivity in a method. Different increments of concentration are clearly distinguishable from each other.



CHAPTER 10

TROUBLESHOOTING

No analytical methodology is going to be entirely free of problems while in use. When quality issues emerge, it is important to have a systematic approach to identifying the issue and resolving it. This chapter provides an overview of the key steps involved in troubleshooting analytical quality issues. It introduces the fundamental principles of, and need for, Root Cause Analysis (RCA). For further detail of RCA see Vanden Heuvel *et al*. (2008).

Root Cause Analysis is a method of problem solving that tries to identify the primary causes of faults or problems by attempting to identify and correct the root causes of events, rather than simply addressing their symptoms (see **Box 10.1** for an example). By focusing corrective actions on root causes, recurrence of the problem can often be prevented or, at least, minimized.

There are many potential drivers of the need for RCA within a water quality laboratory which can necessitate some form of systematic assessment of their origin, occurrence, and significance. While some procedural issues may be relatively straightforward, with a clear and obvious cause (often personnel or policy related), it is often not quite so clear for analytical issues. Some of the more common types of problems which can occur include:

- Complaints from data users (i.e., clients), e.g., clients not receiving test results within the expected or contracted timeframe.
- Quality control failures, e.g., QC sample analyses being outside of action or warning limit values.
- Proficiency testing failures, e.g., poor performance for certain variables during inter-laboratory calibration exercises.

BOX 10.1 EXAMPLE OF APPLYING THE CONCEPT OF ROOT CAUSE ANALYSIS TO AN ANALYTICAL METHOD THAT WAS OUT OF CONTROL

Quality control results for Ammonia analyses indicated that the method was "out of control".

The cause of the problem was identified as incorrect dilution. Investigation revealed the immediate cause was an air displacement pipette that was used to make the dilutions and it was not functioning correctly.

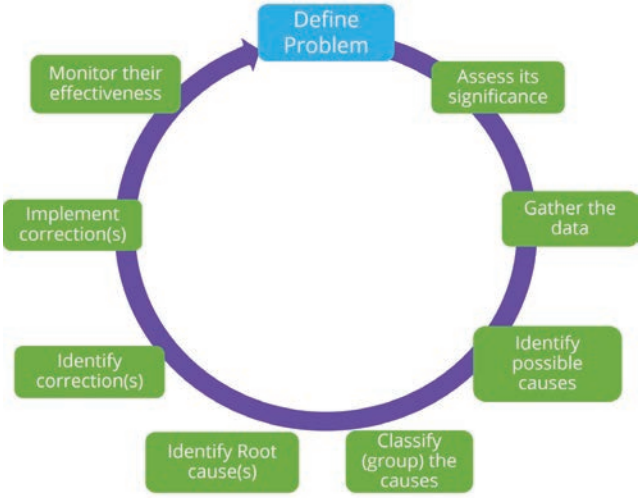
Further investigation and recalibration identified that *the root cause* was likely to have been due to the pipette having been overfilled and was no longer dispensing the correct volumes, due to a sticky plunger.

Corrective action was to service the pipette and increase its calibration frequency.

- Statistical process control anomalies, e.g., a rising trend in the results for a particular variable over time.
- Quality System Audits, e.g., audits conducted to support an accreditation application that identify non-conformances.
- Management reviews or annual planning exercises, e.g., a management review that requires a reduction in the rate of non-conformancies.

The RCA process is a cyclical procedure as outlined in **Fig. 10.1**, which represents the various steps in the

Figure 10.1 The steps to be taken in the Root Cause Analysis procedure



RCA. It starts with defining the problem and ends with monitoring the effectiveness of the corrective actions.

10.1 Defining the problem

When a problem in the analytical processes has been identified as requiring corrective action, the first step in the RCA is to develop a problem statement.

The problem statement must clearly identify the problem before any data and information can be collected (see example in **Box 10.2**). The problem statement is like a mission statement, it states what will try to be resolved through the RCA process. Having a clear problem statement will help keep the technical team on track as they go through the steps of the RCA. It reduces the temptation to attempt to resolve all the problems that may be encountered and encourages sticking to the most important goal (which is to resolve the problem statement). The problem statement should answer the following questions:

- What exactly is the problem? The non-conformance or "problem" must be clearly identified.
- When did the problem occur? The timeframe must be identified and measurement descriptors used to describe the impact of the problem, e.g., how many staff, clients or tests are involved, etc.
- What requirement did the laboratory fail to comply with?

As a general rule, an RCA should be conducted when the event or problem is deemed to be "significant". For example, if the validity of results is called into question; if there is a personal injury or breach of confidentiality, such as sending test results to the wrong client; if there is damage to equipment from fire,

BOX 10.2 EXAMPLES OF WELL-DEFINED AND POORLY DEFINED PROBLEM STATEMENTS

A well-defined problem statement lists descriptors and states the deviation from procedure which will help during the data and information collection phase of the RCA.

Well-defined problem statement: "Our Inductively Coupled Plasma (ICP) Spectrophotometer has not been working for three out of the last five days of this week causing us to miss due dates on 15 sets of river samples. This resulted in a 10% penalty on the total invoice cost for these samples."

This statement clearly defines:

- The problem.
- The consequence of the problem.
- Where the laboratory failed to meet obligations.

Poorly defined problem statement: "Our Laboratory missed the contracted due dates on a number of occasions over the last week."

The above statement does not identify:

- The frequency of the issue.
- Which clients were affected, and how.
- Which data were incorrect?

The poorly defined problem statement must be much more precise in order to help guide the investigation. It needs to contain much more information, and that information needs to be more specific.

malfunction or another cause; and any event where the risk of recurrence is high and will be costly for the laboratory.

Several factors should be considered before commencing the RCA. These include:

- The severity of the event.
- The "cost", both financial and reputational, associated with conducting or not conducting a thorough RCA.
- Damage to the laboratory's reputation and its external credibility as a testing facility.
- Possible staff resentment if the situation is not taken seriously enough.

When the decision is made to conduct an RCA, it should be commenced immediately, in order to reduce any further impacts from the problem.

10.2 Data gathering

When the decision to proceed with an RCA has been made, an individual (or a small team if required) should be appointed to investigate the problem and lead the RCA process. Others may be included on an *ad hoc* basis as necessary. Data are required to inform the process and gathering the relevant data can be the most time-consuming activity. Data can come in many forms and can include, for example, control charts, test reports, instrument logs, calibration records, etc. The data gathering exercise should include the following activities:

1. Review of documentation – policies, procedures and working instructions, as well as all available records and all relevant QC data and records.
2. Interview all people involved in the problem.
3. Repeat the test procedure – exactly and with modifications, and then observe the outcomes.

10.3 Identifying and classifying causes and effects

Where the problem is not particularly obvious, brainstorming (getting multiple inputs regardless of how unlikely they may be) can be used to generate hypotheses of the potential cause(s) of the problem. The brainstorming team should include members that represent all aspects of the non-conforming process and could include an impartial facilitator. Three different methods of identifying and classifying possible root causes are commonly used. These are cause and effect diagrams (Fishbone or Ishikawa diagrams (Ishikawa 1976)), Pareto Charts and the Five Whys Technique.

10.3.1 Cause and effect diagrams

Ishikawa diagrams (sometimes known as cause-and-effect diagrams, herringbone diagrams, fishbone diagrams, or Fishikawa) (Ishikawa 1976) show the potential causes of a specific event. Each cause or reason for imperfection is a source of variation. Causes are usually grouped into major categories to identify and classify these sources of variation.

The categories related to the problem illustrated in **Fig. 10.2** are along the spine. The major categories are equipment, process, people, materials and supplies, environment, and the management and systems. An assessment of the impact of each should lead to a successful outcome. In this instance, poor maintenance of the plasma torch was identified as the primary cause of operational failure.

10.3.2 Pareto Charts

Pareto charts are particularly helpful when analysing data about potential problems in processes, or the frequency of problems. They are also useful when dealing with several possible causes, but when needing to focus on the most significant ones. They help with analysing wide-reaching causes by focussing on their individual components and they convey information about the data easily to others.

Figure 10.2 **Fishbone diagram for the example of an Inductively Coupled Plasma (ICP) Spectrophotometer that has not been working for three out of the last five days, with the different categories of influence and potential causes in each category.**

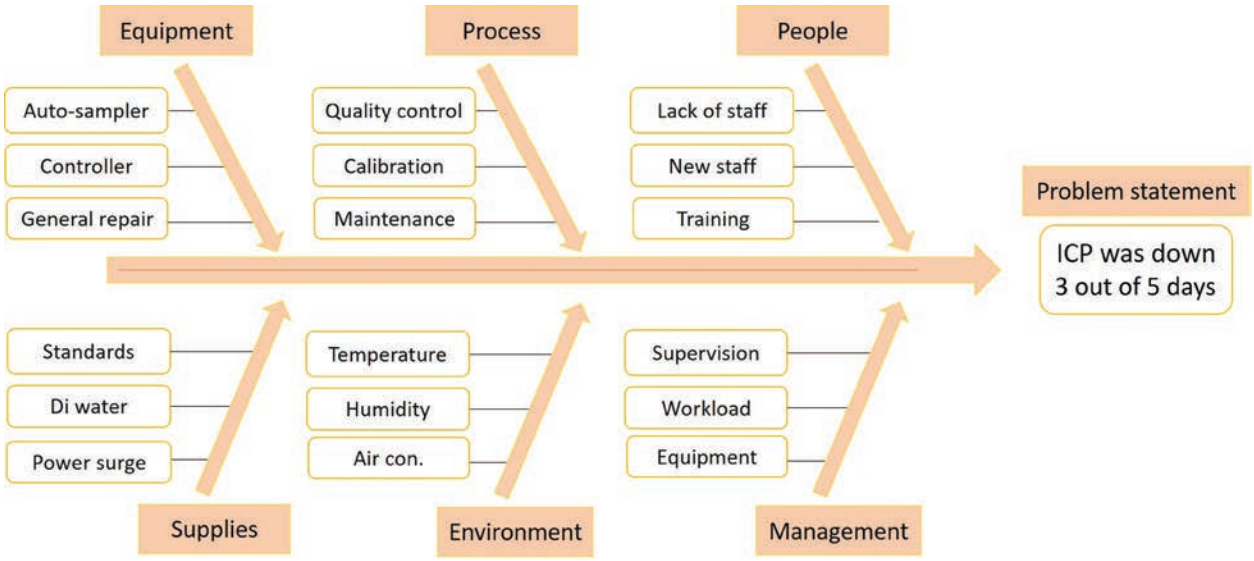
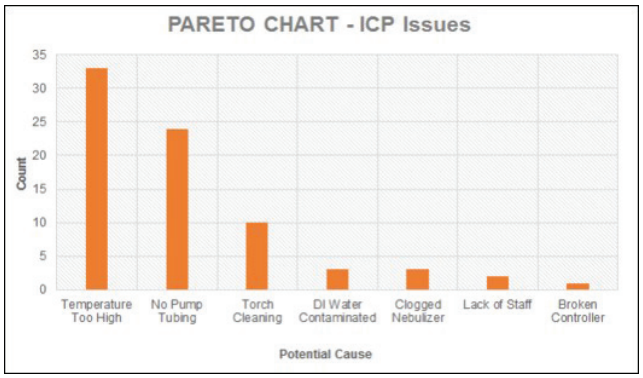


Table 10.1 and **Fig. 10.3** show a Pareto table and associated chart of the values for the same ICP Spectrophotometer that has not been working for three out of the last five days (see Fig.10.2). The plasma torch temperature seems to be the most prevalent issue (occurring 33 times).

Table 10.1 **Data for the various faults with an out of service ICP spectrophotometer**

Potential cause	Count	Cumulative per cent
Temperature too high	33	43%
No pump tubing	24	75%
Torch cleaning	10	88%
Deionised water contaminated	3	92%
Clogged nebulizer	3	96%
Lack of staff	2	99%
Broken controller	1	100%
Total	76	

Figure 10.3 **Pareto chart for ICP failure causes given in Table 10.1**



10.3.3 Five Whys approach

The Five Whys approach is a simple, but very effective, technique for problem-solving that generally helps to get to the potential root of a problem fairly quickly. The problem is examined by asking: "Why?" or "What caused the observed impacts?". The answer to the first "why?" should prompt another question "why?", and the answer to the second "why?" will

prompt another, and so on. The Five Whys problem-solving approach can be used in conjunction with the Fishbone diagram to identify relatively simple problems. After identifying a few of the critical factors that need to be addressed using a Fishbone Diagram, the Five Whys technique can be applied to those factors. This technique can be used to explore the cause-and-effect relationship around the problem. The theory is that by asking "why?" five times, it is possible to get to the real root cause of the problem. For example, "Why is the plasma torch temperature so much of a problem?"

1. The temperature too high. Why?

2. Possible thermostat problem. Why?

3. Incorrectly located. Why?

and so on.

When applying the Five Whys technique, it is generally better not to ask simply "Why" five times, but to use other more probing questions, such as:

- What actually occurred?
- When did it occur?
- Who was involved?
- Who else needs to be involved?
- Were standard practices used?
- Are there any inconsistencies between what the systems are supposed to do and what they are actually doing?
- What assumptions need to be tested?
- What unintended consequences can result from implementing effects?

Human error or personal factors should never be left as the root cause because it is the process and not the people that generally fail. If prevention steps are too costly, monitoring to detect the issue can be implemented at far less cost. It is better not to waste time fixing the problem when it could be prevented from happening in the first place by focusing on quality as a first step.

10.4 The final steps

The ultimate goal of the cause-and-effect analysis is to identify and eliminate the root cause of the problem. The relative significance of each cause needs to be assessed and appropriate corrective actions selected. There are three potential approaches:

1. Applying a "Quick Fix". This provides an immediate resolution but may not be sustainable in the longer term.

2. Remedial action(s). These are determined by the RCA procedure discussions.

3. Corrective action(s). Examples are recalling all suspect data, taking remedial action to correct data and re-issue the results or test report(s), or discontinuing the analysis if a recurrence is likely or if remedial action fails to rectify the problem.

When implementing corrective actions, it is necessary to record the corrective action(s), to assign authority and responsibility to appropriate staff and to ensure the required resources, including any financial resources, are available. Appropriate documents must be updated and relevant practical skills training and opportunities for increased knowledge (if required) should be arranged.

The effectiveness of the RCA process and corrective actions must be monitored to ensure that they have eliminated the problem. This is done by keeping data and detailed records, agreeing on a future review, and deciding whether the fix was effective. If the fix was not effective, the cycle should be started again from the beginning, with a fresh set of questions and interviews.

Appendix A

SAMPLE DATA FOR QUALITY CONTROL CHECKS USING CUSUM CHARTS (DATA FROM MONTGOMERY 2009)

Sample with $\mu_0 = 10$, $\sigma = 1$, hence $k = 0.5$

					μ	**10**			
					k	**0.5**			
				Positive deviations			Negative deviations		
Sample	x_i	$x_i - \mu$	**CuSum**	$x_i - (\mu + k)$	**C_i^+**	**N+**	$(\mu - k) - x_i$	**C_i^-**	**N-**
1	9.45	-0.55	-0.55	-1.05	0	0	0.05	0.05	1
2	7.99	-2.01	-2.56	-2.51	0	0	1.51	1.56	2
3	9.29	-0.71	-2.72	-1.21	0	0	0.21	1.77	3
4	11.66	1.66	0.95	1.16	1.16	1	-2.16	0	0
5	12.16	2.16	3.82	1.66	2.82	2	-2.66	0	0
6	10.18	0.18	2.34	-0.32	2.5	3	-0.68	0	0
7	8.04	-1.96	-1.78	-2.46	0.04	4	1.46	1.46	1
8	11.46	1.46	-0.5	0.96	1	5	-1.96	0	0
9	9.2	-0.8	0.66	-1.3	0	0	0.3	0.3	1
10	10.34	0.34	-0.46	-0.16	0	0	-0.84	0	0
11	9.03	-0.97	-0.63	-1.47	0	0	0.47	0.47	1
12	11.47	1.47	0.5	0.97	0.97	1	-1.97	0	0
13	10.51	0.51	1.98	0.01	0.98	2	-1.01	0	0
14	9.4	-0.6	-0.09	-1.1	0	0	0.1	0.1	1
15	10.08	0.08	-0.52	-0.42	0	0	-0.58	0	0
16	9.37	-0.63	-0.55	-1.13	0	0	0.13	0.13	1
17	10.62	0.62	-0.01	0.12	0.12	1	-1.12	0	0
18	10.31	0.31	0.93	-0.19	0	0	-0.81	0	0
19	8.52	-1.48	-1.17	-1.98	0	0	0.98	0.98	1
20	10.84	0.84	-0.64	0.34	0.34	1	-1.34	0	0
21	10.9	0.9	1.74	0.4	0.74	2	-1.4	0	0
22	9.33	-0.67	0.23	-1.17	0	0	0.17	0.17	1
23	12.29	2.29	1.62	1.79	1.79	1	-2.79	0	0
24	11.5	1.5	3.79	1	2.79	2	-2	0	0
25	10.6	0.6	2.1	0.1	2.89	3	-1.1	0	0
26	11.08	1.08	1.68	0.58	3.47	4	-1.58	0	0
27	10.38	0.38	1.46	-0.12	3.35	5	-0.88	0	0
28	11.62	1.62	2	1.12	4.47	6	-2.12	0	0
29	11.31	1.31	2.93	0.81	5.28	7	-1.81	0	0
30	10.52	0.52	1.83	0.02	5.3	8	-1.02	0	0

Appendix B

EXAMPLE CALCULATION OF MEASUREMENT UNCERTAINTY FOR TOTAL SOLIDS

Calculation of Practical Reporting Limit and Combined Uncertainty

The test procedure uses a standard solution of 1000 mg l^{-1} potassium chloride as the quality control standard. A 100 ml volume (100 mg) is used each time the test is carried out and a standard deviation (*s*) of 2.2 mg (2.2%) was calculated from routine AQC measurements covering February 1997 to June 1999 (n = 20). Based on this, the standard deviation of the AQC control solution at 1,000 mg l^{-1} was calculated as 22 mg l^{-1}.

1. The calculated *Combined Uncertainty* (U_c) for the procedure at 1,000 mg l^{-1} is as follows:

   U (balance) = 0.16 mg (obtained from the balance calibration certificate)

   U (observed) = 2.2 mg (obtained from the AQC data)

   $U_c = 2 \times \sqrt{0.16^2 + 2.2^2} = 2 \times 2.21$ mg => 4.4 mg (based on a mass of 100 mg residue)

   The 95% confidence interval, applying a coverage factor of 2 (k = 2) for a 1,000 mg l^{-1} solution, is thus ± 44 mg l^{-1}.

2. The internal *Relative Standard Deviation* (RSD – standard deviation / mean) associated with this procedure is thus (2.2 / 100) = 0.022. Applying a coverage factor of k = 2 (95%) gives an expanded *Relative Uncertainty* of 0.044.

   Thus, a measurement at, for example, 300 mg l^{-1} based on a 100 ml sample (30 mg as a dried residue) would have a confidence interval (k = 2) of:

   0.044 × 30 mg = ± 1.3 mg => ± 13 mg l^{-1}.

3. At concentrations closer to the *Practical Reporting Limit* of 10 mg l^{-1} (i.e., where there is very little mass) the combined uncertainty is calculated as:

   $2 \times \sqrt{0.16^2 + 0.15^2} = 2 \times 0.22$ mg => 0.44 mg

   where 0.15 is the standard deviation of repeated measurements of a low mass check weight used solely for this purpose, and reweighed each time the weigh balance calibration certificate is renewed.

   Thus, the calculated minimum uncertainty is ± 0.44 mg, which could be taken as ± 0.5 mg.

   Therefore, for a 50 mg l^{-1} sample (100 ml = 5 mg residue) the expanded uncertainty (k = 2, 95%) would be 0.5 mg. The measurement plus *Confidence Interval* would be 50 ± 5 mg l^{-1}.

4. Assessment of the method *Limit of Detection*

   The observed standard deviation of the low check weight for this balance, covering the period January to May 1999, was 0.15 mg. This is assumed to be representative of the repeatability of procedural blanks.

The method Limit of Detection is implied as 4.65 times this value = 0.70 mg. This has been suitably rounded to 1 mg l^{-1} for ease of interpretation, thereby equating to a Practical Reporting Limit of 10 mg l^{-1}

This is comparable with a factor of twice the Combined Uncertainty at low concentrations (as defined in Section 1 above).

Note: The LoD is generally calculated from the expression $LOD = \sqrt{2}.t.sw$ where t is the one-sided Student's t-test statistic (95% confidence level) and sw is the within-batch standard deviation of results from samples ideally containing zero concentration of the determinand of interest (see section 9.2.3). In practice, it is common to apply a limit of 3 times the blank standard deviation as set out in ISO 13530:2009 (ISO 2009). However, at infinite degrees of freedom, the value of t ($\alpha = 0.05$) becomes 1.645 and LOD = $4.65sw$.

Appendix C

A WORKED EXAMPLE OF VALIDATION

This Appendix outlines the approaches generally used for multiple batch verification of method performance. Data can be readily set up in Excel or other spreadsheet formats using the formulae provided.

Step 1 Calibration linearity

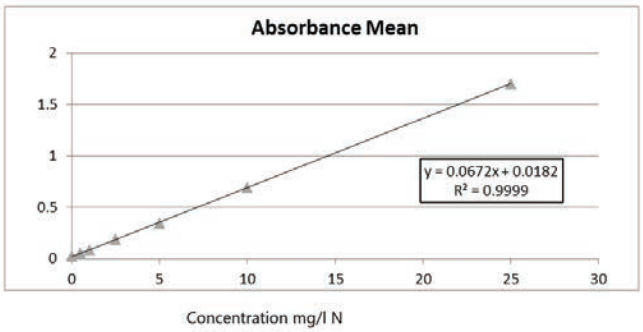
The linearity of the test method should be determined using at least five calibration standards plus a reagent grade water blank (**Table C.1**). The response should be linear across the selected range with a minimum R^2 of 0.95 (**Fig. C.1**).

Table C.1 **Results from the analysis of calibration standards for Ammonia as N for a linearity check in the range 0.5–25 mg l^{-1}**

Concentration	Absorbance Reading 1	Absorbance Reading 2	Absorbance Mean
0.000	0.020	0.028	0.024
0.500	0.054	0.053	0.054
1.000	0.085	0.085	0.085
2.500	0.186	0.187	0.187
5.000	0.345	0.347	0.346
10.000	0.688	0.692	0.690
25.000	1.690	1.712	1.701

Source: Data from P. Webster

Figure C.1 **Checking linearity of measurements for calibration standards in Table C.1**



Step 2 Bias assessment

The bias associated with the test method can be assessed using a low and high standard (typically 20% and 80% of the calibration range). In the example in Table C.1, the 5 mg l^{-1} and 20 mg l^{-1} standards were analysed in nine batches and the mean responses used (**Table C.2**).

It was intended that the test method should cover a potential measurement range of up to 2,000 mg l^{-1} (with appropriate dilution) and that the method should have a precision of better than 0.1 mg l^{-1} or 5% (whichever was greater). Using standards at 0.5 mg l^{-1} (considered to be close to the practical Limit of Quantitation) and a surface water spiked to +2,000 mg $l^{-1}$the bias was 0.069 mg l^{-1} at the lower concentration and – 0.74% at the higher concentration. The natural concentration

Table C.2 Tabulated check for bias assessment

Bias Check				
Standard Concentration (mg l^{-1})	**Mean result**	**Absolute bias**	**Bias %**	**Result**
5.000	4.809	-0.191	-3.962	**Pass**
20.000	20.100	0.100	0.498	**Pass**
Limit of Quantitation / Measurement Range				
	0.500	**2,000.000**		
Replicate 1	0.563	1970.000		
Replicate 2	0.574	1980.000		
Replicate 2	0.564	1990.000		
Replicate 4	0.574	1996.000		
Replicate 5	0.571	1990.000		
Mean	0.569	1985.200		
Precision (*s*)	**0.005**	10.257		
Precision (%)	1.071	**0.513**		
Bias (abs)	**0.069**	-14.800		
Bias (%)	13.840	**-0.740**		
	Pass	**Pass**		

Source: P. Webster

in the spiked surface water was very low and not considered to be of significance given that the test sample required at least x100 dilution to be within the linear range. Both precision and bias requirements were met.

Step 3 Recovery assessment

Using the data in **Table C.3** for surface water and spiked surface water samples, the per cent recovery can be determined from the spiked surface water data. The spike was added to a partially filled 300 ml volumetric flask and the volume made up to the mark. The expected (theoretical) recovered concentration was 9.98 mg l^{-1}. The expected recovered concentration is calculated as:

$$\frac{\text{Spike volume x (spike concentration} - \text{sample mean concentration)}}{\text{Spike volume + sample volume}}$$

$$= \frac{2 \times (1{,}000 - 2.056)}{2 + 198}$$

The actual mean recovery of the nine batches is shown in **Table C.4** and was 9.61 mg l^{-1} giving a recovery of 96.3%.

A recovery of 90–105 is generally acceptable. If recovery is less than 85%, or more than 115%, then the method is unsatisfactory and data should be checked for errors. This approach can also be applied to methods such as solid phase extraction techniques to assess the extent to which a determinand can be extracted from the test sample.

Table C.3 Raw data from the analysis of surface water samples and spiked samples used in precision and recovery assessment

Batch	Replicate	Surface water	Spiked surface water (+ 10 mg l^{-1})	Surface water recovery (mg l^{-1})	Surface water recovery (%)
1	1	1.920	11.300	9.38	
1	2	2.000	11.600	9.60	
	Batch mean	1.960	11.450	9.49	94.9
	Batch SD	0.057	0.212	0.16	
	Variance	0.003	0.045	0.02	
2	1	1.990	11.400	9.41	
2	2	2.050	11.400	9.35	
	Batch mean	2.020	11.400	9.38	93.8
	Batch SD	0.042	0.000	0.04	
	Variance	0.002	0.000	0.00	
3	1	2.090	12.900	10.81	
3	2	2.230	13.000	10.77	
	Batch mean	2.160	12.950	10.79	107.9
	Batch SD	0.099	0.071	0.03	
	Variance	0.010	0.005	0.00	
4	1	2.060	11.500	9.44	
4	2	2.090	11.600	9.51	
	Batch mean	2.075	11.550	9.48	94.8
	Batch SD	0.021	0.071	0.05	
	Variance	0.000	0.005	0.00	
5	1	2.030	11.800	9.77	
5	2	2.230	12.100	9.87	
	Batch mean	2.130	11.950	9.82	98.2
	Batch SD	0.141	0.212	0.07	
	Variance	0.020	0.045	0.00	
6	1	2.070	11.600	9.53	
6	2	2.130	11.700	9.57	
	Batch mean	2.100	11.650	9.55	95.5
	Batch SD	0.042	0.071	0.03	
	Variance	0.002	0.005	0.00	
7	1	2.050	11.500	9.45	
7	2	2.090	11.600	9.51	
	Batch mean	2.070	11.550	9.48	94.8
	Batch SD	0.028	0.071	0.04	
	Variance	0.001	0.005	0.00	
8	1	2.030	11.200	9.17	
8	2	2.050	11.300	9.25	
	Batch mean	2.040	11.250	9.21	92.1
	Batch SD	0.014	0.071	0.06	
	Variance	0.000	0.005	0.00	
9	1	1.960	11.100	9.14	
9	2	1.940	11.400	9.46	
	Batch mean	1.950	11.250	9.30	93.0
	Batch SD	0.014	0.212	0.23	
	Variance	0.000	0.045	0.05	
Overall mean		2.056	11.667		
Overall Mean Recovery				9.61	96.1

Note: SD – standard deviation. Source: P. Webster

Table C.4 Recovery data for analyses from Table C.3

Matrix	Surface Water
Sample Volume (V)	198.000
Mean Sample Conc (U)	2.056
Mean Spiked concentration (S)	11.667
Spike Concentration (c)	1000.000
Spike Volume (v) OR (w)	2.000
Expected Recovery Conc.	9.979
Mean recovery	9.611
Overall Mean Recovery %	96.304
Assessment	**Pass**

Step 4 Calculation of performance statistics for raw data

The assessment of overall performance involves a number of stages and calculations as outlined below to determine various statistical outputs. A good explanation of the terms and equations used is given in Environment Agency (2018).

Table C.3 and **C.5** show the actual raw data for the blank, high and low standard solutions and also the freshwater samples. From these data, a series of statistical parameters can be calculated as follows for both sets of the data.

Within batch standard deviation (s_w)

1. Sum the variances of all nine batches

2. Enter the number of batches (m = 9) and the number of replicates per batch (n = 2)

3. Calculate the "within batch mean square" function Mo as the (Variance sum / No. of batches)

4. Calculate Mo^2

5. Calculate s_w as the square root of Mo

Between batch standard deviation (s_b)

1. Calculate the standard deviation of all the batch mean values (s_{bm}) using the Excel function STDEV

2. Calculate the variance of batch means as the square of s_{bm}

3. Calculate the "between batch mean square" function ($M1$) as the product of s_{bm}^2 (above) and the number of replicates per batch (n) … $M1 = n \times s_{bm}^2$

4. Calculate the square of $M1$

5. Calculate the between batch standard deviation as

$$S_b = \sqrt{\frac{M1 - Mo}{n}}$$

Total standard deviation calculation

1. Calculate the numerator as:
   $M1 + (n - 1)Mo$

2. The denominator is the number of replicates per batch (n)

3. Calculate the division product as: $\frac{M1 + (n - 1)Mo}{n}$

4. Calculate the total standard deviation (s_t) as the square root of the division product:

$$S_t = \sqrt{\frac{M1 + (n - 1)Mo}{n}}$$

5. The total standard deviation for the blank was calculated as 0.046. The LoD would therefore be 0.14 mg l^{-1} ($3s_t$) and the LoQ = 0.5 mg l^{-1} ($10s_t$ rounded upwards).

Degrees of freedom calculation (Df)

1. Calculate the numerator using the formula:
   $m(m - 1)\ (M1 + (n - 1)Mo)^2$

2. Calculate the value of $(m - 1) \times (n - 1)$ … in this case the values are $(9 - 1) \times (2 - 1)$

Table C.5 Raw data for analyses of standard solutions used in precision and recovery assessment

Batch	Replicate	Blank	5 mg l^{-1} standard	20 mg l^{-1} standard
1	1	0.085	4.710	19.800
1	2	0.148	4.780	20.100
	Batch mean	0.117	4.745	19.950
	Batch SD	0.045	0.049	0.212
	Variance	0.002	0.002	0.045
2	1	0.165	4.840	20.000
2	2	0.203	4.770	19.900
	Batch mean	0.184	4.805	19.950
	Batch SD	0.027	0.049	0.071
	Variance	0.001	0.002	0.005
3	1	0.160	4.780	20.200
3	2	0.185	4.830	20.200
	Batch mean	0.173	4.805	20.200
	Batch SD	0.018	0.035	0.000
	Variance	0.000	0.001	0.000
4	1	0.239	4.840	20.000
4	2	0.240	4.880	20.000
	Batch mean	0.240	4.860	20.000
	Batch SD	0.001	0.028	0.000
	Variance	0.000	0.001	0.000
5	1	0.137	4.770	20.600
5	2	0.241	4.980	20.800
	Batch mean	0.189	4.875	20.700
	Batch SD	0.074	0.148	0.141
	Variance	0.005	0.022	0.020
6	1	0.135	4.830	20.100
6	2	0.184	4.840	20.200
	Batch mean	0.160	4.835	20.150
	Batch SD	0.035	0.008	0.071
	Variance	0.001	0.000	0.005
7	1	0.118	4.760	19.900
7	2	0.193	4.810	20.100
	Batch mean	0.156	4.785	20.000
	Batch SD	0.053	0.035	0.141
	Variance	0.003	0.001	0.020
8	1	0.126	4.800	19.900
8	2	0.177	4.850	19.900
	Batch mean	0.152	4.825	19.900
	Batch SD	0.036	0.035	0.000
	Variance	0.0013005	0.00125	0.000
9	1	0.113	4.730	20.000
9	2	0.146	4.770	20.100
	Batch mean	0.130	4.750	20.050
	Batch SD	0.023	0.028	0.071
	Variance	0.001	0.001	0.005
	Overall mean	0.166	4.809	20.100

Notes: SD – standard deviation. Source: P. Webster

3. Calculate the *Df* denominator as:
   $m\,M1^2 + ((m - 1)(n - 1)\,Mo^2)$

4. Calculate the *Df* estimate as the numerator / denominator and express this as an integer value

Analysis of Variance (ANOVA) precision test

1. Determine the mean value of all of the test batches

2. Recall the within, between and total batch standard deviations (s_w, s_b, s_t)

3. Determine the relative standard deviation as: total standard deviation / sample mean concentration

4. Record the target standard deviation. This will generally be in the format of an absolute value and a percentage value whichever is the greater, e.g., *"0.1 mg l^{-1} or 5% whichever is greater"*. Expressed in this way, the lower value applies to concentrations below the crossover point (2 mg l^{-1} in this example), thereafter the percentage figure applies. In the case of the blank there is no specific target and in Table C.5 the value of 0.05 (5%) has been applied for convenience of spreadsheet calculations.

5. Multiplying the sample concentration by the target standard deviation % (expressed as a numeric value) gives the target standard deviation equivalent (target SD equivalent)

6. From the table of F-values (**Table C.6**) determine the value of $F_{(0.05)}$ that is associated with the integer value of the degrees of freedom estimate (*Df* est)

7. Determine the calculated F-value as: $(s_t$ / target SD equivalent$)^2$

8. If the calculated F-value (F calc) is less than the tabulated F-value (F tab) then the performance criteria are met, which equals a **PASS**

9. Repeat the above process for all of the test solutions

Table C.6 Table of F-values ($F_{0.05}$)

Df	1	2	3	4	5	6
f(0.05)	1.04	3	2.61	2.37	2.21	2.1
Df	7	8	9	10	11	12
f(0.05)	2.01	1.94	1.88	1.83	1.79	1.75
Df	13	14	15	16	17	18
f(0.05)	1.72	1.69	1.67	1.64	1.62	1.61
Df	13	14	15	16	17	18
f(0.05)	1.72	1.69	1.67	1.64	1.62	1.61

REFERENCES

Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27(1), 17–21. https://doi.org/10.1080/00031305.1973.10478966

Association of Official Analytical Collaboration (AOAC) International (2016). AOAC International Methods Committee Guidelines for Validation of Microbiological Methods for Food and Environmental Surfaces. In: Latimer, G.W. (ed.) *Official Methods of Analysis of AOAC International*. Appendix J. Available at: https://www.yumpu.com/en/document/read/65659513/free-download-pdf-official-methods-of-analysis-of-aoac-international-20th-aoac-2016-gigapaper

Bell, S. (2001). *Measurement good practice guide - A Beginners Guide to Uncertainty of Measurement*. Issue 2. Teddington: National Physical Laboratory. Available at: https://www.esscolab.com/uploads/files/measurement-guide.pdf

British Standards Institute (2009). *British Standard BS 1427:2009 Guide to on-site test methods for the analysis of waters*. London. Available at: https://knowledge.bsigroup.com/products/guide-to-on-site-test-methods-for-the-analysis-of-waters/standard

Cavanagh, N., Nordin, R.N. and Warrington, P.D. (1997). *Freshwater biological sampling manual*. BC Ministry of Environment, Lands and Parks - Water Management Branch. Available at: https://www2.gov.bc.ca/assets/gov/environment/natural-resource-stewardship/nr-laws-policy/risc/freshwaterbio.pdf

Clarke, R. T. and Hering, D. (2006). Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. *Hydrobiologia*, 566(1), 433–439. https://doi.org/10.1007/S10750-006-0079-2

Crumbling, D.M., Lynch, K., Howe, R., Groenjes, C., Shockley, J., Keith, L. *et al*. (2001). Applying the concept of effective data to contaminated sites could reduce costs and improve cleanups. *Environmental Science and Technology*, 35(19), 404A-409A. https://doi.org/10.1021/ES012490G

European Commission (2009). Commission Directive 2009/90/EC of 31 July 2009 laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status (Text with EEA relevance). *Official Journal of the European Union*, L. 201/36. Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:201:0036:0038:EN:PDF

Ellison, S.L.R. and Williams, A. (eds) (2012). *Eurachem/CITAC guide: Quantifying uncertainty in analytical measurement* (3rd ed.). ISBN 978-0-948926-30-3. Available at: https://www.eurachem.org/index.php/publications/guides/quam

Environmental Protection Agency (undated) Gender strategy. Available at: https://www.epa.ie/our-services/research/epa-research-2030/strategies-and-policies/gender-strategy/ [Accessed 06 October 2022]

Finnish Meteorological Institute (2002). *Mann-Kendall Test and Sen's Slope Estimates for the Trend of Annual Data. Version 1.0 Freeware*. Available at: https://en.ilmatieteenlaitos.fi/makesens

Haase, P., Pauls, S.U., Schindehütte, K. and Sundermann, A. (2010). First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: Human error greatly lowers precision of assessment results. *Journal of the North American Benthological Society*, 29(4), 1279–1291. https://doi.org/10.1899/09-183.1

Helsel, D.R. (2012). *Statistics for Censored Environmental Data Using Minitab® and R: Second Edition*. Hoboken NJ: John Wiley and Sons, Inc . https://doi.org/10.1002/9781118162729

Irish National Accreditation Board (2019). *PS15 Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories Guide to Method Validation for Quantitative Analysis in Chemical Testing Laboratories (ISO 17025)*. Dublin. Available at: https://www.inab.ie/inab-documents/mandatory-policy/guide-to-method-validation-for-quantitative-analysis-in-chemical-testing-laboratories-iso-17025-.pdf

International Accreditation New Zealand (2021). *Technical Guide. Chemical and Microbiological Testing Laboratories. Measurement Uncertainty, Precision and Limits of Detection* AS TG 5. Auckland. Available at: https://www.ianz.govt.nz/resources/technical-guides

Ishikawa, K. (1976). *Guide to Quality Control*. Tokyo: Asian Productivity Organization. Available at: https://archive.org/details/guidetoqualityco00ishi

International Organization for Standardization (1994). *ISO 5725-1:1994 - Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*. Geneva. Available at: https://www.iso.org/standard/11833.html

International Organization for Standardization (2003a). *ISO 15839:2003 - Water quality — On-line sensors/analysing equipment for water — Specifications and performance tests*. Geneva. Available at: https://www.iso.org/standard/28740.html

International Organization for Standardization (2003b). *ISO 17381:2003 - Water quality — Selection and application of ready-to-use test kit methods in water analysis*. Geneva. Available at: https://www.iso.org/standard/30626.html

International Organization for Standardization (2009). *ISO/TS 13530:2009 - Water quality — Guidance on analytical quality control for chemical and physicochemical water analysis*. Geneva. Available at: https://www.iso.org/standard/52910.html

International Organization for Standardization (2010a). *ISO 16269-4:2010 - Statistical interpretation of data — Part 4: Detection and treatment of outliers*. Geneva. Available at: https://www.iso.org/standard/44396.html

International Organization for Standardization (2010b). *ISO/IEC 17043:2010 - Conformity assessment — General requirements for proficiency testing*. Geneva. Available at: https://www.iso.org/standard/29366.html

International Organization for Standardization (2012a). *ISO 11352:2012 - Water quality — Estimation of measurement uncertainty based on validation and quality control data*. Geneva. Available at: https://www.iso.org/standard/50399.html

International Organization for Standardization (2012b). *ISO 10870:2012 Water quality — Guidelines for the selection of sampling methods and devices for benthic macroinvertebrates in fresh waters.* Geneva. Available at: https://www.iso.org/standard/46251.html

International Organization for Standardization (2013). *ISO 7870-2:2013 - Control charts — Part 2: Shewhart control charts*. Geneva. Available at: https://www.iso.org/standard/40174.html

International Organization for Standardization (2014). *ISO 5667-14:2014 - Water quality — Sampling — Part 14: Guidance on quality assurance and quality control of environmental water sampling and handling*. Geneva. Available at: https://www.iso.org/standard/55452.html

International Organization for Standardization (2017). *ISO/IEC 17025:2017 - General requirements for the competence of testing and calibration laboratories*. Geneva. Available at: https://www.iso.org/standard/66912.html

International Organization for Standardization (2018). *ISO 5667-3:2018 - Water quality — Sampling — Part 3: Preservation and handling of water samples*. Geneva. Available at: https://www.iso.org/standard/72370.html

International Organization for Standardization (2020). *ISO 5667-1:2020 - Water quality — Sampling — Part 1: Guidance on the design of sampling programmes and sampling techniques*. Geneva. Available at: https://www.iso.org/standard/72369.html

International Organization for Standardization (2021). *ISO 7870-4:2021 - Control charts — Part 4: Cumulative sum charts*. Geneva. Available at: https://www.iso.org/standard/74101.html

International Organization for Standardization (2022). *ISO 13528:2022 Statistical methods for use in proficiency testing by interlaboratory comparison.* Geneva. Available at: https://www.iso.org/standard/78879.html

Jarvis, B. (2016). *Statistical aspects of the microbiological examination of foods*. Third edition. London. Academic Press.

Magnusson, B., Naeykki, T., Hovind, H., Krysell, M. and Sahlin, E. (2017). *Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report TR 537* (4th ed.). Taastrup: Nordtest. Available at: http://www.nordtest.info/wp/

Magnusson, B. and Örnemark, U. (eds) (2014). *Eurachem Guide: The Fitness for Purpose of Analytical Methods - A Laboratory Guide to Method Validation and Related Topics*. 2nd edition. Gembloux: Eurachem. ISBN 978-91-87461-59-0. Available at: https://www.eurachem.org/index.php/publications/guides/mv

Montgomery, D. C. (2019). *Introduction to statistical quality control*. 8th edition. Hoboken, NJ: John Wiley and Sons, Inc. Available at: https://www.wiley.com/en-us/Introduction+to+Statistical+Quality+Control%2C+8th+Edition-p-9781119635444

Nelson, L. S. (1984). The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology*, 16(4), 237–239. https://doi.org/10.1080/00224065.1984.11978921

New Zealand National Institute of Water and Atmospheric Research (n.d.). *Time Trends. Analysis of water quality trends*. Aukland. Available at: https://niwa.co.nz/freshwater/management-tools/water-quality-tools/analysis-of-water-quality-trends (Accessed 24 September 2022).

Pawley, S., Dobson, M., Fletcher, M. and Powell, A. (2011). *Guide to British freshwater macroinvertebrates for biotic*

assessment. SP67. Newby Bridge: Freshwater Biological Association. Available at: https://www.fba.org.uk/shop/p/book-41

R Core Team (2018). *R: a language and environment for statistical computing*. Vienna: Foundation for Statistical Computing. Available at: https://www.R-project.org/

Rice, E.W., Baird, R.B., Eaton, A.D. (eds) (2017). *Standard Methods for the Examination of Water and Wastewater*. 23rd edition. Prepared and published jointly by the American Public Health Association (APHA), American Water Works Association (AWWA) and the Water Environment Federation (WEF). Available at: https://www.wef.org/resources/publications/books/StandardMethods/

Singapore Accreditation Council (2008). *Technical guide 2. A guide on measurement uncertainty in chemical and microbiological analysis* 2nd edition. Standards, Productivity and Innovation Board. Available at: https://www.sac-accreditation.gov.sg/files/documents/laboratory-accreditation/Technical-Guide-2.pdf

Standing Committee of Analysts (2018). *Estimation of Uncertainty of Measurement for Chemical and Physio-chemical Determinants in Drinking Water 2018 - Methods for the Examination of Waters and Associated Materials*. Nottingham.

UNISTAT Ltd (undated). Variable Control Charts in Excel with UNISTAT. Available at: https://www.unistat.com/931/ [Accessed 28 November 2022].

United States Environmental Protection Agency (1984). *Appendix A to Part 136. Methods for Organic Chemical Analysis of Municipal and Industrial Wastewater. Method 604 - Phenols*. Office of Water. Washington, DC. Available at: https://www.epa.gov/sites/default/files/2015-08/documents/method_604_1984.pdf

United States Environmental Protection Agency (2000). *Method 528 - Determination of Phenols in Drinking Water by Solid Phase Extraction and Capillary Column Gas Chromatography/Mass Spectrometry (GC/MS)*. National Exposure Research Laboratory. Cincinnati, OH. Available at: https://www.epa.gov/sites/default/files/2015-09/documents/m_528.pdf

United States Environmental Protection Agency (2002). *Guidance for Quality Assurance Project Plans EPA QA/G-5*. Washington, DC. Office of Environmental Information. Available at: https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf

United States Environmental Protection Agency (2006). *Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4.* Washington, DC. Office of Environmental Information. Available at: https://www.epa.gov/sites/default/files/documents/guidance_systematic_planning_dqo_process.pdf

Vanden Heuvel, L. N., Lorenzo, D. K., Jackson, L. O., Hanson, W. E., Rooney, J. J. and Walker, D. A. (2008). *Root Cause Analysis Handbook: A Guide to Efficient and Effective Incident Investigation*. 3rd edition. Produced by ABS Consulting. Brookfield, CT: Rothstein Associates Inc.

Vetter, T. (2001). Quantifying Measurement Uncertainty in Analytical Chemistry, a Simplified Practical Approach. In: *Measurement Science Conference 2001*, Session V-B, Anaheim, CA, Available at: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=904890. [Accessed 24 September 2022]

Wheeler, D. J. and Chambers, D. S. (2010). *Understanding statistical process control*. 3rd edition. Knoxville, TN: SPC Press. Available at: https://www.spcpress.com/book_understanding_statistical_process_control.php

UNEP GEMS/Water Capacity Development Centre
Environmental Research Institute
University College Cork
Lee Road
CORK
Ireland
e-mail: gemscdcadmin@ucc.ie
Tel: +353 21 4205276