Chapter 3
Planning your monitoring programme.
Sampling and measurements

This chapter addresses how to design research studies and establish monitoring programmes, with an emphasis on quality assurance, quality control, and the collection of representative samples.

The contents in this chapter are applicable to both treatment plant monitoring and water quality monitoring.

CHAPTER CONTENTS
3.1 Types of Monitoring Programmes and Studies .......................................................... 40
3.2 Quality Assurance and Quality Control ................................................................. 41
3.3 Sample Collection ................................................................................................. 50
3.4 Sample Size, Containers, and Holding Times ....................................................... 62
3.5 Statistical Power and Number of Samples ............................................................. 63
3.6 Check-List for Your Report ..................................................................................... 67

© 2020 The Authors. This is an Open Access book chapter distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (https://creativecommons.org/licenses/by-nc-nd/4.0/). This does not affect the rights licensed or assigned from any third party in this book. The chapter is from the book Assessment of Treatment Plant Performance and Water Quality Data: A Guide for Students, Researchers and Practitioners, Marcos von Sperling, Matthew E. Verbyla and Silvia M. A. C. Oliveira (Authors).
doi: 10.2166/9781780409320_0039
3.1 TYPES OF MONITORING PROGRAMMES AND STUDIES

Whether you are in charge of monitoring the performance of a treatment plant, monitoring an outfall for compliance with regulations, or completing a special study such as a thesis project, there are several important considerations that must be taken regarding:

- What is a sample?
- Where should I collect samples?
- When should I collect samples?
- How should I collect samples?
- How many samples should I collect?
- What measurements should I take in the field and in the laboratory?

A robust operational monitoring programme is essential for any water or wastewater treatment plant to evaluate the efficacy of the treatment system or a water body to assess the quality of its water. Monitoring programmes at some facilities (especially large facilities serving urban centres) or for some research projects or special studies might include continuous and real-time measurements by probes, sensors, and/or data loggers, or remote-operated controls to make operational changes to the system based on incoming data or system alarms. However, at a minimum, monitoring programmes and studies should include the following elements:

- Visually inspect different components of the treatment system periodically
- Measure flow rates in the system
- Collect and analyse liquid and/or solid samples for the concentrations of relevant contaminants
- Implement quality assurance and quality control measures and document them in a quality assurance project plan (QAPP) report

Compliance monitoring refers to monitoring activities that are intended to ensure compliance with laws and regulations, such as the maximum contaminant levels (MCLs) for drinking water or effluent discharge limits on the concentrations of certain pollutants for wastewater treatment facilities. These laws and regulations are often established by governmental environmental agencies and public health authorities, operating either at the national level and/or the regional (state, department, and province) level. Many of the specifics of compliance monitoring programmes are often driven by local laws and regulations. The contaminants or pollutants of interest for the study are often specified by the legislation. They may vary from site to site, depending on the water body in question, its beneficial use category, the characteristics of its watershed (land use and industrial activity), or the results obtained in previous monitoring efforts. You should conform with any specific details specified in the legislation, such as monitoring frequencies, detection limits for reporting, sample location, and sample type.

For research projects or special studies, the types of monitoring activities and elements must be chosen by the researcher or the director of the study. It is typically up to the researcher to decide which method to use for analysing flow rates (e.g., manual measurements or automated measurements) and water quality (e.g., collecting and analysing samples in the laboratory versus the use of probes and sensors with data loggers). If you are engaged in a study like this, often you must balance the precision and accuracy of the different methods with the cost of acquiring the equipment, materials, and supplies needed to use a particular method.

Emergency studies are often triggered by specific environmental accidents, public health emergencies, or hazardous weather events, such as chemical spills, disease outbreaks, algal blooms, hurricanes, wildfires, or other natural disasters. The parameters to be studied are typically associated with the nature and type of
disaster or accident, and the duration of the study is typically short and intensive, in order to obtain answers as quickly as possible.

3.2 QUALITY ASSURANCE AND QUALITY CONTROL

3.2.1 Introductory concepts

Quality assurance (QA) and quality control (QC) are essential in order for people to have trust in the results of your study. Furthermore, starting with a good quality assurance plan can help you ensure that you are collecting the right samples, processing and analysing them using appropriate methods, and getting enough results to make appropriate findings based on the project’s goals. We recommend that you take the following five-step process, which is inspired by the California State Water Resources Control Board’s Surface Water Ambient Monitoring Program Quality Assurance Plan, to document your monitoring assignment, project, or research study. You should summarize the following items in a quality assurance project plan (QAPP), which can be part of your overall report or a separate report from your main study report:

- Scope of the study
- Samples and populations
- Measurements and anticipated use of data
- Standard assessment thresholds and operating procedures
- Quality control samples
- Data management and analysis

3.2.2 Scope of the study

First, you must determine what you plan to study or monitor. Start by asking yourself the following three questions that define what will be addressed, as well as where and when the study or monitoring programme will happen.

- What question do you hope your study will answer?
- What are the boundaries and limits of the study in terms of its location?
- What is the general length and time frame of the study?

Let us discuss each of the questions individually.

**What question do you hope your study will answer?**

You should develop a guiding question (or a set of questions) for your study or programme. Your question(s) should be specific and measurable. The study or programme you are proposing should also be **FIRE** – that is, it should be Feasible, Interesting, Relevant, and Ethical; and, if you are a thesis student or a research scientist, the research question should also be novel and have intellectual merit (Farrugia *et al.*, 2010).

A feasible study is one that (a) includes a sufficient number of samples, (b) utilizes methods that are standardized, recognized, or rigorously tested, and (c) can be completed for a cost that fits within the project or programme’s budget. An interesting study is one that will be read and/or referenced by others. A relevant study is one that has direct and practical application to practice or policy. An ethical study is one that protects the rights, welfare, and well-being of participants or beneficiaries, ensures compliance with local, national, and international laws and regulations, and adheres to the principles...
outlined in the Belmont report, specifically: respect for persons (treating people as autonomous agents and protecting individuals with diminished autonomy), beneficence (securing the well-being of people, doing no harm or maximizing possible benefits while minimizing possible harms), and justice (selecting participants equitably in terms of who receives the benefits of research studies and who bears their burden).

In the case of water and wastewater treatment plants, ethics could pertain to the treatment plants selected for study and their beneficiary populations, as well as the people or organizations responsible for operating such facilities.

After stating your question(s), write down a brief background or context of the problem(s) being addressed. Even if there are currently no problems, write down a summary of the problem(s) that you are trying to avoid by executing the study. During the planning stage of a monitoring and sampling programme, it is often helpful to determine if there are synergistic or overlapping monitoring and evaluation efforts or other studies that have been previously completed or are currently in progress to avoid duplication of efforts. When planning a research project, for example, this can be done by conducting a literature review on the topic.

**What are the boundaries and limits of the study in terms of its location?**

**Describe** the water body(ies) or treatment plant(s) that will be the focus of the study. Provide a brief description of the study location(s) with a map, if available. Show a schematic of the water body or the treatment plant, with a visual indication of the location of all samples collected and analysed. For more information about where to collect samples, refer to Section 3.3.

You should also take note of any obstacles that may interfere with collecting samples or obtaining a complete data set, for example, is the site bounded by fences, is access limited to daytime hours, are there safety concerns with going to the sampling site, is there a potential for dangerous weather conditions, is a permit required for accessing the sampling site, etc.

In terms of ethical considerations, think about who is potentially benefitting or putting themselves at risk as a result of the study being carried out at the chosen location. For example, if you are conducting a research study that documents the performance of a wastewater treatment plant at removing pathogens, what if the findings indicate that pathogens are not very effectively removed from the treatment plant. If these findings are made public and linked to the facility, will it put the manager or operator of the facility at risk of losing his or her job? Will there be a potential for public fear or outrage due to the findings? Will those findings benefit or harm the public in the long run? Are there certain populations who will gain economic or health benefits as a result of the knowledge being produced, and if so, are these populations the frequent recipients of such benefits (e.g., due to the treatment plant being located close to the university), and are there other more remote communities that will fail to benefit from the data being produced by the research? These are important considerations when choosing the location for a study.

**What is the general length and time frame of the study?**

Here, you should indicate if the project is intended to be short term or ongoing. If it is short term, indicate the day(s), month(s), and/or year(s) during which the study will take place. The length and time of a study may be determined based on the research question (sufficient to obtain a large enough sample size to answer the question or address the hypothesis), the legislative authorities (which may specify the length, frequency, or nature of sampling and measurements), and budgetary limitations. More samples, more information, and more data points will always be desirable and helpful to answer a particular research question, but this comes at a cost, and it is the researcher’s job to determine cost-benefit trade-offs and establish a length and time frame that are appropriate for making cost-effective decisions or findings.

Planning an appropriate time frame for a study is an important consideration, especially if the research question addresses variations with seasonality or time of day. For example, if temperatures during winter...
seasons cause lower efficiencies in terms of treatment plant performance, or the pollutant levels in a water body are of greatest concern during the rainy season, then the timeframe for the study should take place during the season of concern. For more information about the temporal aspects of sampling and sample collection, see Section 3.3.

Table 3.1 shows three hypothetical sampling and monitoring programmes and their respective scopes of work, including a study/research question, a description for the general timeframe and length of the study, and whether or not the study is connected with another project.

### 3.2.3 Environmental samples, statistical samples, and populations

Your study will produce data and information that result from samples that are collected and analysed and measurements that are made in the field. Our purpose for collecting these samples and taking these measurements is to understand the quality of the system, the efficiency of the process, or the quality of the liquid, solid, or gas products emitted by our system. We often deal with different matrices, including liquids, solids, and gases. For the purpose of the explanation that follows below, we will talk about the mass of some constituent in a liquid volume of water. However, understand that this same concept applies to the mass, number, or amount of constituent in any matrix (e.g., mass of solid, volume of air, etc.).

Therefore, we want to know the quantity of some constituent in our system. For example, we might want to quantify the mass of suspended solids in treated wastewater effluent, or the mass of total nitrogen in drinking water, or the amount of dissolved oxygen in a river. However, these systems are often ‘turned on’ 24 h/d, 7 d/week, 52 weeks/year, so it is impossible for us to know the true total quantity of solids in all of the water discharged, the true total mass of nitrogen in all of the water at the drinking water plant, or the true total amount of dissolved oxygen in the entire river.

You can consider these true total amounts to be the population of the pollutants or constituents of interest. The population is sort of like if you were able to collect an infinite number of samples from the system. But, because we cannot collect an infinite number of samples, we will never know the true amount of a constituent in the system. Therefore, we collect samples, for instance, of a manageable volume of water, and we measure the quantity of the constituent (say, nitrogen) contained in these samples. We then use those measurements to make inferences (i.e., draw conclusions) about the amount of the constituent likely contained in the rest of the volume of water that we were not able to sample and analyse. The more volumes of water sampled, the more confidence we have about the true amount of nitrogen (or any other constituent) in our system.

Therefore, in summary, with respect to monitoring programmes, when we talk about the total quantity of pollutants or constituents in our system, we are referring to the population.

---

**The population** of the concentration of any given pollutant or constituent comprises the amounts of that constituent (e.g., mg, moles, colony-forming units, etc.) contained in many individual volumes of water—in fact, the amounts contained in so many volumes of water that they account for every single drop of water in your system.

**A sample** is the amount of constituent contained in a limited number of smaller volumes of water collected as a subset of the total amount of water in your system. For instance, if our system is a lake, then the population is all of the water in the entire lake and our sample is the small volume of water taken back to our laboratory.
<table>
<thead>
<tr>
<th>Study/research question</th>
<th>Non-Point-Source Contamination Study</th>
<th>Evaluation of Treatment Plant Performance</th>
<th>Drinking Water Source Compliance Programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>During a rain event, how does the concentration of total suspended solids (TSS) in the New River change with respect to time, at locations upstream and downstream of the New River Condominiums construction site?</td>
<td>In a pilot-scale upflow anaerobic sludge blanket (UASB) reactor treating wastewater from a small city with low industrial activity, how does the reduction of chemical oxygen demand (COD) vary throughout the year and with respect to highly fluctuating ambient temperatures and variable influent organic and hydraulic loading rates?</td>
<td>In the Unionville Reservoir, which is a proposed source of raw water for a new drinking water treatment plant, are the concentrations of general physical contaminants, inorganic contaminants, nutrients, regulated synthetic organic compounds, and volatile organic compounds below the current maximum contaminant levels?</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>The New River</td>
<td>A pilot-scale UASB reactor in a small city (with highly fluctuating ambient temperatures, variable organic and hydraulic loading rates)</td>
<td>The Unionville Reservoir</td>
</tr>
<tr>
<td>Background or context of the problem</td>
<td>Concentrations of TSS in a local river are not in compliance with regulatory standards, and the water management authority is suspecting that a construction site may be one source of the pollution. Therefore, the authority has arranged to carry out a special study to measure the construction site’s impact on water quality during rain events</td>
<td>A thesis student is evaluating the performance of UASB reactors in a region with highly fluctuating seasonal temperatures and variable water usage throughout the year, affecting organic and hydraulic loading at the treatment plant. The goal is for this research to inform the design and operation of a proposed upgrade to the treatment plant</td>
<td>A county water authority is designing and planning a new drinking water treatment plant, and as such needs to evaluate the quality of a proposed source of raw surface water for the treatment facility. There is no inherent problem to be solved, but the monitoring programme is intended to avoid health problems associated with contaminants</td>
</tr>
<tr>
<td>Length and timeframe</td>
<td>This project will start in October 2019 (at the beginning of the rainy season) and is anticipated to end in April 2020 (at the end of the rainy season)</td>
<td>This project will begin in September 2019 and is anticipated to end in May 2021, with sampling carried out throughout the year, with samples collected during different seasons and for different water use scenarios</td>
<td>This project originally spanned from January until December 2007. However, the water body is now the main raw water source for the treatment facility, and the programme has continued as a long-term ongoing programme with no established end-date</td>
</tr>
<tr>
<td>Synergistic studies and programmes</td>
<td>Monthly historical TSS data at a downstream location is available since 2001 from an ongoing compliance programme. The construction firm also has a stormwater pollution prevention plan report, dated March 2019, which includes a peak runoff analysis and information about pollution prevention measures. Inspection reports are available from the local government authority</td>
<td>A previous student published a thesis on a study of the performance of UASB reactors at high temperatures and steady loading rates. There are also numerous academic journal articles focussed on the performance of these reactors under various conditions. A thorough literature review should be conducted by the student</td>
<td>There are several inputs to the water body from local rivers and streams, some of which are monitored regularly as part of a water quality compliance programme. There is also land use data for the watershed, available as geographical information system shapefiles</td>
</tr>
</tbody>
</table>
Finally, it is worth noting some additional details about the semantics of the word ‘sample’ that may cause confusion for some readers. In our discipline, and in all disciplines that deal with water quality and treatment processes, the word ‘sample’ refers to the physical smaller volume or mass of water (or another liquid, or a solid or gas, etc.) that is collected from a larger body of water (and typically analysed, for instance, in a laboratory).

However, in the field of statistics, the word ‘sample’ refers to a smaller set of data collected from a larger population. Therefore, our statistical sample would consist of the data obtained from several water samples collected from a larger body of water. A good way to avoid confusion in the terminologies is to distinguish a statistical sample from the environmental samples (e.g., water samples, sludge samples, biogas samples, etc.). In this chapter and in Chapter 4, we will discuss best practices for collecting and analysing environmental samples. You will learn more about statistical samples and distributions in Chapter 5.

### 3.2.4 Measurements and anticipated use of data

Characterize what type of data will be collected, determine what measurements and/or observations will be made (Table 3.2), and specify what type of matrix will be sampled, observed, or probed. It may include one or more of the following:

- Water (drinking water, environmental water, polluted (waste)water)
- Sludge/biosolids
- Soil/sediments
- Animal tissue/collection of organisms (e.g., for a bioassessment)

It is helpful to define the anticipated use of the data prior to commencing the monitoring programme or study. This will typically depend on the type of data collection activity being conducted (see Section 3.1 – e.g., operational monitoring, compliance monitoring, emergency assessment, research project, etc.). If the programme’s intent is to monitor ambient water quality, then the data might be used to characterize watershed health, support water quality control plans, develop policies, or address impacts to human and animal health (e.g., fishing, swimming, or drinking advisories). If the purpose of the study is purely to advance science, then the data might be used for a peer-reviewed journal article to elucidate a mechanism associated with a treatment process, to evaluate cutting edge methodologies, or to pilot-test innovative technologies. In some cases, the data might be used for regulatory purposes (e.g., issuing permits, investigative orders, waivers, or establishing maximum daily loads).

Then, determine what kinds of decisions will be made from the study’s results and identify possible actions that may be taken, depending on the results obtained. For example, will a fine be applied if a discharge point to a water body is found to be not in compliance with regulations? Will a treatment process be implemented in full scale if it achieves a certain per cent removal of a contaminant at a pilot scale?

### 3.2.5 Standard assessment thresholds and operating procedures

It is important to document and communicate any assessment thresholds needed for your project to ensure that the analytical results are fully supportive of your decision. Assessment thresholds may include any of the following:

- A total maximum daily load (TMDL) is defined as the maximum amount of a pollutant allowed to enter a waterbody in order to meet water quality standards. In the United States, the TMDL determines the target pollutant reduction and allocates load reductions necessary for any
Table 3.2 Examples of measurements commonly used for monitoring programmes and research studies.

<table>
<thead>
<tr>
<th>Type of measurement</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Field measurements  | • Dimensions of the treatment unit process  
|                     | • Temperature  
|                     | • Wind speed  
|                     | • Water depth |
| Bioassessment       | • Benthic macroinvertebrate survey  
|                     | • Periphyton survey  
|                     | • Fish survey |
| Continuous data     | • Flow rate  
|                     | • Turbidity  
|                     | • Temperature  
|                     | • Dissolved oxygen  
|                     | • Conductivity  
|                     | • Ammonia nitrogen  
|                     | • Nitrate  
|                     | • pH  
|                     | • Dissolved organic carbon (DOC) |
| Chemistry           | • Conventional  
|                     |   • Alkalinity  
|                     |   • Hardness  
|                     |   • Biochemical oxygen demand (BOD)  
|                     |   • Chemical oxygen demand (COD)  
|                     | • Nutrients  
|                     |   • Organic nitrogen  
|                     |   • Ammonia nitrogen  
|                     |   • Nitrate  
|                     |   • Nitrite  
|                     |   • Phosphate and total phosphorus  
|                     | • Inorganics  
|                     |   • Trace metals  
|                     |   • Mercury  
|                     | • Organics  
|                     |   • Pesticides  
|                     |   • Fuels  
|                     |   • Surfactants  
|                     |   • Solvents  
| Microbiology        | • Total heterotrophic count  
|                     | • Microscopic evaluation (e.g., of mixed liquor suspended solids)  
|                     | • Faecal indicator bacteria  
|                     | • Microbial source tracking markers  
|                     | • Pathogenic microorganisms |
| Solids              | • Total solids  
|                     | • Volatile solids  
|                     | • Total suspended solids |

(Continued)
source(s) of the pollutant. It is equal to the sum of all waste load allocations from point sources of pollution, plus the sum of all load allocations from non-point sources of pollution, plus a margin of safety to account for the uncertainty associated with predicting pollutant reductions (US EPA, 2018).

- A **maximum contaminant level goal** (MCLG) or **public health goal** (PHG) is defined as the level of a contaminant in drinking water that does not pose a significant risk to health (OEHHA, 2019). MCLGs and PHGs are not regulatory standards but instead are used to trigger risk communication activities. For example, in some jurisdictions, if the MCLG or PHG for a public water system is exceeded, a public notice must be distributed to all users of the water system, but no fine or penalty is imposed to the water authority. MCLGs and PHGs are established using rigorous methods. It starts with a compilation of relevant information about a contaminant from the scientific literature (e.g., studies of the contaminant’s effects on laboratory animals and humans who have been exposed to the contaminant). The data from these studies are then used to perform a chemical or **microbial risk assessment** to determine the levels of the contaminant that could be associated with various adverse health effects. Certain thresholds have to be set in order to establish the MCLG or PHG – for example, in California, PHGs are calculated assuming a maximum one in 1,000,000 probability of adverse health effects for people who drink water every day for 70 years. This means that, on average, not more than one person in a population of $1 \times 10^6$ would be expected to develop cancer as a result of exposure to the particular pollutant. For microbial risk assessments, lower thresholds are often adopted, such as one in 10,000 or even as low as one in 100 in some countries.
- A **maximum contaminant level** (MCL) is the maximum permissible level of a contaminant in water delivered to any user of a public water system in the United States (U.S. Code, 1974). These levels are set as close to the MCLG or PHG as feasible. Other countries have adopted similar terminologies for such levels.

You should also define what **standard operating procedures** (SOPs) will be used for sample collection and field measurements. In many cases, if the programme is for compliance, the SOPs will be specified by the regulations. For research projects, the SOPs must be based on protocols recognized in the scientific literature or must be thoroughly tested against other standard methods for quality control.

### Table 3.2 Examples of measurements commonly used for monitoring programmes and research studies (Continued).

<table>
<thead>
<tr>
<th>Type of measurement</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspended sediment concentrations</td>
<td></td>
</tr>
<tr>
<td>Total dissolved solids</td>
<td></td>
</tr>
<tr>
<td>Algal bloom response</td>
<td></td>
</tr>
<tr>
<td>Toxins</td>
<td></td>
</tr>
<tr>
<td>Microscopy</td>
<td></td>
</tr>
<tr>
<td>Chlorophyll-a</td>
<td></td>
</tr>
<tr>
<td>Toxicity</td>
<td></td>
</tr>
<tr>
<td>Acute</td>
<td></td>
</tr>
<tr>
<td>Chronic</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Satellite imagery</td>
<td></td>
</tr>
<tr>
<td>Remotely sensed data</td>
<td></td>
</tr>
<tr>
<td>Aerial drones</td>
<td></td>
</tr>
<tr>
<td>Cutting edge research methodology</td>
<td></td>
</tr>
</tbody>
</table>
3.2.6 Quality control samples

Establishing and maintaining a quality control is essential for any project or programme. A quality control programme should consist of an initial demonstration of capability, ongoing demonstration of capability, method detection limit determination, and quality control sampling, which consists of control and background samples (analysed to isolate background conditions and site-specific effects) and an assortment of variability controls, including sample spikes, sample blanks, inhibition controls, and field or laboratory replicates (APHA, 2017; US EPA, 2017).

First, the laboratory must demonstrate its initial capability of implementing each method. This typically requires analysing the following control samples, which are used to determine the precision (standard deviation) and accuracy (per cent recovery):

- Any required calibration standards
- At least one reagent blank (negative control), which should be free of the contaminant of concern
- ≥4 spiked controls (positive controls), which are like reagent blanks that are spiked with known concentrations of the contaminant of concern
  - Accuracy: From these samples, calculate the per cent recovery and ensure that it is within the specified acceptance criteria. In the absence of acceptance criteria, aim for a per cent recovery between 80% and 120% as a starting point (APHA, 2017).
  - Precision: As a measure of sample precision, calculate the coefficient of variation (CV) from the replicate spiked controls, which is equal to the standard deviation divided by the mean value. Ensure that the CV is within the specified acceptance criteria, but if none are provided, then aim to achieve a CV of ≤20% as a starting point (APHA, 2017).

The method detection limit (MDL) is defined as the concentration that produces a signal that is different from the blank with a probability of 99%. At a minimum, at least seven replicates of a process blank (also known as a method blank or a reagent blank) should be analysed. A process blank is a sample blank (typically reagent water), that is free from the contaminant of interest, and that is analysed and processed exactly the same way as the samples, using the same methods, and coming into contact with all other reagents in the complete procedure. This is distinct from an instrument blank, which is a sample blank that is only analysed in the instrument (but not processed). For more information about how to calculate the MDL and other detection and quantitation limits, see Chapter 4.

After demonstrating initial capabilities, the laboratory should continue to demonstrate ongoing capabilities by analysing process blanks and spiked controls periodically and evaluating them to ensure continued precision and accuracy. The frequency of ongoing demonstration of capability should be as specified in the protocol or standard operating procedure but at a minimum should be conducted quarterly. If process blanks are reading concentrations below the MDL, then no qualification is needed in the results. If process blanks are above the MDL but below the limit of quantification (see Chapter 4), then a qualifying statement should be provided with the sample results to indicate a positive process blank. If the process blank is detected at a concentration above the limit of quantification, then corrective action is needed (APHA, 2017).

A background control sample is one that is collected from a site that is not impacted by pollution or from a time when the level of the pollutant is at a stable ‘background’ level. This type of control sample is especially useful if you are trying to identify a source of contamination. Specifically, you should compare concentrations in this sample with concentrations in samples collected at sites suspected to be impacted from the pollution source to give you more confidence that the levels you detect in the sample are indeed elevated by the suspected source of pollution.
A field blank is a sample of reagent water that is taken out to the field during sample collection, stored along with the samples, transported to the laboratory along with the samples, and analysed along with the samples. The purpose of the field blank is to test for contamination that may have occurred during sample collection, storage, or transportation. If a contamination event is detected in the field blank, it can be compared with the process blank and the instrument blank to determine where the contamination happened (Table 3.3).

Other important variability controls include field replicates and laboratory replicates, which can be used to calculate coefficients of variation for losses of precision resulting from variation in the field or in the laboratory. These coefficients of variation can be compared to the coefficient of variation calculated for spiked laboratory controls. Field or laboratory replicates might be analysed for every one out of 10 or 20 samples.

Inhibition controls are a normal part of quality control sampling for certain protocols. Essentially, some environmental constituents may inhibit certain reactions that are necessary to produce a signal. Tests for inhibition can be done either by diluting the sample and measuring the resulting signal, which should be proportional to the dilution factor. Otherwise, samples can be spiked with a known concentration of the contaminant and the measured to see if the amount added corresponds to the increase in the signal (Table 3.4).

### Table 3.3 Method for analysing field blanks, process blanks, and instrument blanks to determine the source of contamination.

<table>
<thead>
<tr>
<th>Field blank result</th>
<th>Process blank result</th>
<th>Instrument blank result</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>No contamination occurred</td>
</tr>
<tr>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td>Contamination occurred at the instrument or the instrument needs to be recalibrated</td>
</tr>
<tr>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
<td>Contamination likely occurred during sample processing and may also have occurred at the instrument</td>
</tr>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Contamination likely occurred during sample collection and/or transportation and storage. It may also have occurred during sample processing or at the instrument</td>
</tr>
</tbody>
</table>

### Table 3.4 Method for interpreting dilution or spike controls for inhibition.

<table>
<thead>
<tr>
<th>Type of inhibition test</th>
<th>Result</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilution (1:10)</td>
<td>Sample concentration in the dilution control is 10% of the undiluted sample</td>
<td>No evidence of inhibition</td>
</tr>
<tr>
<td></td>
<td>Sample concentration in the dilution control is much greater than 10% of the undiluted sample</td>
<td>Inhibition likely occurred</td>
</tr>
<tr>
<td>Spiked sample (three times the sample concentration was spiked into a replicate sample and analysed)</td>
<td>Spiked sample concentration is four times as high as the un-spiked sample</td>
<td>No evidence of inhibition</td>
</tr>
<tr>
<td></td>
<td>Spiked sample concentration is less than, equal to, or only slightly greater than the un-spiked sample concentration</td>
<td>Inhibition likely occurred</td>
</tr>
</tbody>
</table>
3.2.7 Data management and analysis

Data should be managed in a way that allows for it to be archived in a common format to data from other studies, and with appropriate metadata to describe the data set. More information about data management is provided in Chapter 4. In addition to primary data (e.g., data collected from laboratory analysis on collected samples), list what other sources of data, if any, will be used in the study or monitoring programme. This might include data provided by another agency or entity, complimentary data gathered from a weather station, or qualitative data gathered through surveys, interviews, observation, or a mixed methods approach.

Before even starting to sample and collect data, you should determine which statistical method(s) will be used to analyse the data and what the acceptable level of error will be for the statistical test(s) being used. A common level of acceptable error for research studies is an alpha error of 0.05 and a beta error level of 0.20. This is the probability you are willing to accept for making a type I error (false positive) or a type II error (false negative), respectively. Further description of these types of errors is presented in Chapter 10.

Table 3.5 shows a summary of different types of studies that are commonly completed for research projects and the corresponding statistical test(s) that should be used for such studies. It should be noted that this is not necessarily a comprehensive list, as there are many more statistical tests that are outside the scope of this book. Chapters 5 and 6 of this book cover descriptive statistics; however, an in-depth coverage of some of the more advanced statistical tests highlighted in Table 3.5 is beyond the scope of this book. There are many excellent text resources which cover these methods and others (e.g., Sokal & Rohlf, 2012).

For a useful analogy on understanding the meaning of alpha and beta errors, consider the penal system of a country. Suspects are considered innocent until proven guilty, just as two samples are considered equal until proven to be significantly different from each other. Accepting an alpha level of 0.05 (5%) is like accepting that you may erroneously convict an innocent person to be guilty 1 out of 20 times (5%) on average. Accepting a beta level of 0.20 (20%) is like accepting that you may fail to convict a guilty person (for lack of sufficient evidence) 1 out of every 5 times (20%) on average.

3.3 SAMPLE COLLECTION

3.3.1 Spatial aspects of sampling

To evaluate and monitor the efficacy of a treatment system, samples should be collected at the influent and effluent of the system. For compliance programmes, at a minimum, samples should be collected at the final effluent location (to demonstrate compliance with MCLs and effluent discharge limits). However, it is also useful to monitor the performance of a particular unit process. This requires collecting samples at the influent and effluent point of that unit process. In some cases, it might even be desirable to collect samples at various locations within the unit process (e.g., in a reed bed or horizontal constructed wetland, you might want to collect samples at intermediate points spatially distributed within the wetland bed). Also, environmental variables (e.g., temperature, dissolved oxygen, etc.) and control variables (e.g., mixed liquor suspended solids, sludge blanket levels) may need to be collected or measured inside the treatment unit. In many treatment systems, it is useful to collect data from samples collected in side streams or waste streams associated with the process.

Many researchers place a greater emphasis on water samples, but for many treatment systems, it is useful to collect sludge samples as well. For some processes, collecting gas might be necessary or beneficial. It might be useful to think of sampling locations as being essential, important, and potentially useful. Essential sampling locations are the most important locations that you shall not do without. The final effluent point of water and wastewater treatment systems is an essential sampling point, because it allows...
Table 3.5 A guide for choosing the appropriate statistical test or procedure based on the purpose of the experiment, the number of sample groups, and notes about the type of data that can be used.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Samples</th>
<th>Statistical tests</th>
<th>Types of data</th>
<th>Chapters</th>
</tr>
</thead>
</table>
| **Descriptive:** Describing the central tendency and variation of data (such as the mean, median, standard deviation, percentiles, confidence interval around the mean, etc.) | 1 or more | • Confidence intervals  
  o Real numbers  
  o Proportions (percentages)  
  o Binary parameters  
  o Positive integers | • Continuous numbers  
  • Proportions and percentages (binomial data from binary measurements)  
  • Poisson counts (positive integers) | 5 and 6 |
| **Inferential comparative (one sample):** Comparing sample(s) against a threshold (e.g., regulatory limit or target value) | 1       | • One-sample \( t \)-test  
  o One-sided or two-sided | • Continuous numbers  
  • Proportions and percentages (binomial data from binary measurements)  
  • Poisson counts (works best with large samples; often necessary to transform data) | 9 and 10 |
| **Inferential comparative (two samples):** Comparing treatment versus control differences  
Comparing two samples with different treatments applied | 2       | • Two-sample \( t \)-test  
  o One-sided or two-sided  
  o Independent or paired samples  
  o Pooled or not pooled variance (homoscedastic/heteroscedastic)  
  • Rank tests (non-parametric) such as Mann–Whitney–Wilcoxon for independent samples or Wilcoxon signed-rank test for dependent samples may be better for small data sets or with non-normally distributed data | • Continuous numbers  
  • Proportions (check the normality assumption for parametric methods) | 10 |
| **Inferential comparative (two samples):** Comparing proportions and percentages between a treatment and a control (2 × 2 contingency tables) | 2       | • Chi-square test (expected frequencies in contingency table must be >5)  
  • Fisher’s exact test (can be used when expected frequencies are <5) | • Proportions and percentages | Outside scope of this book |
| **Inferential comparative (three+ samples):** Multiple comparisons with one or more treatment factor | 3 or more | • Analysis of variance (ANOVA)  
  o One-way, two-way, 2\(^2\) factorial  
  o Balanced versus unbalanced  
  o Blocking factor versus no blocking | • Continuous numbers  
  • Proportions (check the normality assumption for parametric methods) | 10 |

(Continued)
Table 3.5 A guide for choosing the appropriate statistical test or procedure based on the purpose of the experiment, the number of sample groups, and notes about the type of data that can be used (Continued).

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Samples</th>
<th>Statistical tests</th>
<th>Types of data</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>○ Multiple comparisons to control or multiple comparisons of all</td>
<td>2</td>
<td>• Correlation</td>
<td>• Continuous numbers (regression)</td>
<td>11</td>
</tr>
<tr>
<td>○ Post hoc comparisons with adjustments to achieve the desired familywise error rate or control the false discovery rate (FDR)</td>
<td>2</td>
<td>• (Multiple) linear regression</td>
<td>• Binomial proportions (GLM, use link function)</td>
<td></td>
</tr>
<tr>
<td>○ Least significant difference</td>
<td>2</td>
<td>• Generalized linear models (GLMs)</td>
<td>• Poisson counts (GLM, use link function)</td>
<td></td>
</tr>
<tr>
<td>○ Dunnett</td>
<td>2</td>
<td>○ Gaussian (continuous numbers; conventional linear regression – identity link)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Tukey–Kramer</td>
<td>2</td>
<td>○ Logistic Regression (binomial – logit link)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Benjamini–Hochberg (FDR)</td>
<td>2</td>
<td>○ Gamma (positive continuous numbers – inverse link)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Storey (FDR)</td>
<td>2</td>
<td>○ Poisson counts (positive integers – log link)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>○ Kruskal–Wallis non-parametric test followed by the post hoc Dunn test</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Descriptive trends:
Determine the relationship between explanatory variable(s) and a response variable

1 | • Correlation
• (Multiple) linear regression
• Generalized linear models (GLMs)
○ Gaussian (continuous numbers; conventional linear regression – identity link)
○ Logistic Regression (binomial – logit link)
○ Gamma (positive continuous numbers – inverse link)
○ Poisson counts (positive integers – log link)

Inferential trends (two samples):
Compare two regression slopes

2 | • Regression or GLM with F-test

Inferential trends (three or more samples):
Comparing more than two regression slopes

3 or more | • Regression or GLM with analysis of covariance

Note: Common underlying assumptions for the above-mentioned statistical tests include the following: homoscedasticity (constant variance of errors); non-stochastic explanatory variables (explanatory variables are accurately measured); normal distribution of residual errors; linearity (randomness of residuals with respect to the explanatory variables); no multi-collinearity (no significant correlation between explanatory variables); and independence of observations.
you to assess compliance and risk by comparing your measurements against some regulatory limit or desired level. However, the nature of essential sampling locations also depends on the aim or goals for the study. For example, if the purpose of the project is to evaluate the performance of a particular treatment technology, then samples should obligatorily be collected at both the influent and effluent points at a minimum. If you want to study the performance of each unit comprising the treatment plant, you need to collect samples upstream and downstream each unit (see Figure 3.1). In some cases, especially when evaluating wastewater treatment reactors using a mass balance approach, it is also important to collect samples of sludge and sometimes also gas emissions.

For studies related to water quality in rivers and streams, samples should be collected immediately upstream and downstream of the suspected point source of pollution. In addition, the point source of pollution should be sampled, and if the point source originates from a treatment plant, ideally samples should also be collected at the influent of the treatment plant, in order to evaluate the efficacy of the treatment process at eliminating the pollutant of concern. Finally, it may be desirable to collect several additional samples further downstream of the treatment plant, at different distances, to evaluate the degradation or further dilution of the pollutant in the water body (Figure 3.2).

You should avoid sampling in areas where water is stagnant or where reverse flow patterns occur. In addition, areas near the inner edge of curves in a river may not be representative due to the patterns of flow and turbulence at those locations. Samples are best collected below the surface to avoid the influence of surface boundary effects. Samples should also not be collected too close to the bottom of a river. However, if collected and analysed separately, samples collected at the bottom sediment of a river or another water body surface may help understand the evolution of pollution over time and the potential for accumulation of possible chemical substances in macrobiota. The sampling points should be representative, avoiding areas affected by atypical habitats, such as those under bridges (ABNT, 1987).

**Figure 3.1** Recommended sampling points for different types of studies in a treatment plant.
Table 3.6 shows some example sampling locations and timeframes for our three hypothetical studies described in Table 3.1. Note that the frequency of sample collection should be determined after conducting a power analysis with the proposed alpha and beta error levels and desired effect size (see Section 3.5 for power analysis).

### 3.3.2 Types of samples

Figure 3.3 illustrates different types of samples that we may collect, depending on the objective of our study and the available resources:

- **Instantaneous conditions:** grab sample
- **Approximation of average conditions:** composite sample (fixed volume and flow-proportional volume)
- **Concentration profiles over time:** sequence of grab samples or measurements by a sensor

#### Grab sample

A **grab sample** (Figure 3.3a) consists of a single sample of water collected at a given instant of time. It is the easiest type of sample to collect, but it may not be the most representative at locations where the quality of water changes throughout the day. This type of sample does not take into account the potential variability of concentrations with respect to time, and it may lead to the underestimation or overestimation of the true mean concentration, unless concentrations are relatively constant with respect to time. If you need to know the variation in the concentrations over time, several sequential grab samples must be collected individually and analysed separately (Figure 3.3d).

Some types of analysis require the use of grab samples, since the samples cannot be stored for the period of time required for a composite sample (see below), rather they must be analysed or measured immediately after collection. Some examples include pH, temperature, and dissolved oxygen. If using grab samples over a long period of time, it is important to ensure that samples are collected at approximately the same time of day for consistency. Grab samples are
### Table 3.6 Example sampling locations and types for three hypothetical studies.

<table>
<thead>
<tr>
<th>Study/research question</th>
<th>Non-point-source contamination Study</th>
<th>Evaluation of treatment plant performance</th>
<th>Drinking water source compliance programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>During a rain event, how does the concentration of TSS in the New River change with respect to time, at locations upstream and downstream of the New River Condominiums construction site?</td>
<td>In a pilot-scale UASB reactor treating wastewater from a small city with low industrial activity, how does the reduction of COD vary throughout the year and with respect to highly fluctuating ambient temperatures and variable influent organic and hydraulic loading rates?</td>
<td>In the Unionville Reservoir, which is a proposed source of raw water for a new drinking water treatment plant, are the concentrations of general physical contaminants, inorganic contaminants, nutrients, regulated synthetic organic compounds, and volatile organic compounds below the current maximum contaminant levels?</td>
<td></td>
</tr>
</tbody>
</table>
| This project will start in October 2019 (at the beginning of the rainy season) and is anticipated to end in April 2020 (at the end of the rainy season) Flow-proportional spatial and temporal composite samples will be collected in the New River during up to 10 rain events, at locations immediately upstream and immediately downstream from the construction site. In addition, grab samples will be collected prior to the start of each rain event to understand background concentrations during dry weather. Grab samples will also be collected after the hydrograph recedes back to background flowrates | This project will begin in September 2019 and is anticipated to end in May 2021, with sampling carried out throughout the year, with samples collected during different seasons and for different water use scenarios. Samples will be collected at the following locations:  
- Influent wastewater  
- Effluent treated water  
- Waste sludge  
- Sludge blanket | This is a long-term ongoing programme with no established end-date. Spatial composite samples are collected from the reservoir on a monthly basis, with aliquots at five different depths that correspond to the intake depths used by the water treatment facility. In addition, the influent flowrate will be measured daily, and the volume of waste sludge will be measured each time the reactor is de-sludged. |
appropriate for the assessment of an effluent stream that does not discharge on a continuous basis and to provide information about the concentration of a contaminant at a particular time of day. Certain parameters, including pH, temperature, dissolved oxygen, and residual chlorine, cannot be analysed with composite samples due to short holding times (US EPA, 2017). In most other cases, composite samples are the most appropriate, especially when calculating loading rates as a mass per unit time.

Figure 3.3 Different types of samples to be collected and analysed.
Temporal composite sample

A **temporal composite sample** is a mixture of smaller sub-samples (called aliquots), collected periodically throughout the day. This type of sample is more representative at locations where water quality changes throughout the day, as the composition of the sample helps minimize the effects of variability in the concentrations over time, giving a better representation of the true average concentration. It is especially useful at wastewater treatment plants, where the flow rate and quality of influent sewage can vary considerably throughout the day. Usually, a composite sample is collected over a 24-h period, and autosamplers can be programmed to collect composite samples for a period of 24 h. However, in some cases, 12-h or 8-h composite samples are used when autosamplers are not available, for convenience purposes (e.g., to avoid having to collect samples during the night time or during non-working hours). The frequency of collection for each aliquot is usually every 1 h, but may be higher or lower, depending on the expectation of variability of concentrations. You should ensure that the aliquots collected at the beginning of sampling are well preserved to prevent internal reactions that may affect the concentration of the pollutant of concern. For this reason, it is recommended to use a cooler to store samples or to use automatic samplers with ice space. When preparing the composite sample once all aliquots are collected, each container containing the aliquots should be thoroughly mixed, as sedimentation may have occurred.

There are different types of temporal composite samples: the two most common are fixed-volume composite samples (aliquots each have equal volumes) (Figure 3.3b) and flow-proportional composite samples (aliquots have volumes that are proportional to the flow measured at the time they are collected – higher flow → higher volume; lower flow → lower volume) (Figure 3.3c). Flow-proportional samples are more representative of changing water quality conditions throughout the day, which is common with wastewater plants. Example 3.1 illustrates the differences between a fixed-volume composite sample and a flow-proportional composite sample. The associated Excel spreadsheet contains a worksheet that allows you to calculate aliquot volumes for your own flow-proportional composite sample.

Spatial composite sample

A **spatial composite sample** refers to the combination of individual samples collected at different geographical or physical positions. This type of sample is especially important to get representative estimates of water and solid matrices in systems with poor mixing. For example, when collecting samples from a mid-size or large river, it is recommended to collect samples at various points in the cross section of the river and mix them into a single sample. This way, you get an idea of the average concentration in the water passing through all points of the river. Spatial composite samples are also commonly used when collecting sludge or sediments.

When collecting a spatial composite sample, it is important to note that the aliquots should be collected within a short time interval to minimize the influence of temporal variations. In rivers, concentrations of the constituents are rarely homogeneous throughout their cross section. In fact, the river cross section may have several stagnant zones, in which the concentrations may vary greatly. Furthermore, there may be differences with respect to depth. The Brazilian National Standards Organization (ABNT, 1987) recommends sampling aliquots at different locations along the cross section and at different depths, depending on the width and depth of the river (Figure 3.4). Sample aliquots will compose a spatial composite sample that accounts for variation throughout the cross section. In general, if the river or stream width is greater than 5 m, then your composite sample should include spatial variations; if the river or stream is deeper than 2 m, then a composite sample with aliquots collected at different depths should be collected.
Sensors

Sensors are used to collect real-time measurements of certain parameters, or surrogate measurements that correlate with the concentrations of certain pollutants. Sensors are commonly used in treatment plants, because they provide real-time information to operators, who may make operational changes based on the sensor readings. Sensors may collect single measurements at a time (e.g., if the sensor is manually inserted into the water body) or multiple measurements throughout the course of a day (e.g., if the sensor is installed in-line or connected to a data logger) (Figure 3.3e). There are sensors for various parameters of interest in water quality, such as temperature, pH, dissolved oxygen, and electrical conductivity.

3.3.3 Need for a time delay to collect the downstream sample?

There is some debate whether a downstream sample should be collected after a time interval from the collection of the upstream sample, with this time interval being equivalent to the hydraulic retention time (HRT) of the unit or system. Let us analyse Figure 3.5 and the following possibilities:

- **Sampling in a river receiving a point-source pollution.** This first case (top figure) is slightly simpler. The river flows approximately like a plug flow (see Chapter 14 for the concept of plug flow), and then the time spent for the water to reach the downstream sampling location is approximately the travelling time dictated by the distance between the points and the mean flow velocity. You could then take this into account to collect a sample with a delay equivalent to the travelling time, in an effort to collect the same plug of water that received the discharge.

- **Sampling in a treatment unit.** The second case (middle and bottom figures) is a much debated one. Several people argue that we should collect the downstream sample with a delay equivalent to the hydraulic retention time of the treatment unit, treatment plant, or water body. The expectation is that we would be able to collect the same water that entered the unit, underwent treatment, and then left the unit. However, this will depend essentially on the hydrodynamic behaviour of the
unit. If our unit approaches plug flow, then the same considerations made above for a river would apply, but to a lesser extent. In this case, the travelling time through the unit could be close to HRT, if there is little dispersion in the unit. However, if the unit has some degree of mixing (as most units do), the contaminant is dispersed in the reactor volume, and any peak value in the influent would bring a response in the effluent at a faster time compared with HRT. The higher the degree of mixing, the faster the response in the outlet. In this case, implementing a delay equal to HRT does not assist us in obtaining the same fluid elements, before and after the unit.

An overall comment is that our monitoring programme should be established on a practical basis, according to the frequently difficult logistics on site. If HRT is 12 h and you collect the influent sample at 9:00 am, you would need to collect the effluent sample at 9:00 pm if you believe that the strategy of the delay equivalent to the HRT should be implemented. Ok, you could have an automatic sampler and solve this problem. But what if the HRT of the unit is 5, 10, 30, or 60 days, as some units in natural treatment processes have? Would you wait that long? Would it be meaningful? Would you still believe that you are sampling the same fluid elements, before and after treatment?

We believe not, and we think you should be practical in your monitoring programme and collect as many samples as possible from the influent and effluent locations (preferably composite samples). By analysing the time series of data, you will be able to draw conclusions about the performance of the unit. If you want to make more advanced analyses between the upstream and downstream data sets, you could study the cross-correlation between them (correlation with one of the series subjected to a lag – see comments in Chapter 11).

**EXAMPLE 3.1 CALCULATE ALIQUOT VOLUMES FOR TEMPORAL COMPOSITE SAMPLES**

Develop a plan to collect (a) a 1-L fixed-volume temporal composite sample and (b) a 1-L flow-proportional composite sample of wastewater at a treatment plant.
The flow rate measured at two-hourly intervals is

<table>
<thead>
<tr>
<th>Time of Day</th>
<th>Flow (L/s)</th>
<th>Time of Day</th>
<th>Flow (L/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>1.2</td>
<td>12:00</td>
<td>3.1</td>
</tr>
<tr>
<td>02:00</td>
<td>2.4</td>
<td>14:00</td>
<td>3.7</td>
</tr>
<tr>
<td>04:00</td>
<td>3.7</td>
<td>16:00</td>
<td>3.2</td>
</tr>
<tr>
<td>06:00</td>
<td>4.3</td>
<td>18:00</td>
<td>2.5</td>
</tr>
<tr>
<td>08:00</td>
<td>3.8</td>
<td>20:00</td>
<td>2.0</td>
</tr>
<tr>
<td>10:00</td>
<td>3.3</td>
<td>22:00</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Note: This example is also available as an Excel spreadsheet.

Solution:

(a) **Fixed-volume composite sample**

For fixed-volume composite samples, you can collect sub-samples every hour for 24 h. However, you can also collect sub-samples at any time interval (e.g., every 30 min, every 2 h, or every 3 h), as long as you continue for a period of 24 h. The representativeness of the sample increases for smaller intervals.

Sometimes, composite samples are only collected during daytime hours (for convenience); however, it should be noted that this may introduce a bias in the measurement of water quality. Flow rates are typically much lower during evening hours, and influent wastewater quality may be quite different during late evening and early morning hours (when users are sleeping) than it is during the day (when users are awake). Wastewater quality may also change drastically throughout the day as users engage in different activities (e.g., using the toilet versus showering and washing dishes). Industrial activities (which often only take place during working daytime hours) can also drastically change the quality of wastewater.

Suppose you choose to collect a fixed-volume composite sample of 1 L at intervals of 3 h, for a total of eight sub-samples \((24/3 = 8)\). The sub-samples, with volumes of 125 mL \((1000/8 = 125)\), could be collected as shown in the following table, then mixed to form a composite sample with a volume of 1 L.

Fixed-volume composite sample collection plan (eight aliquots) is shown in the following table:

<table>
<thead>
<tr>
<th>Time of the Day</th>
<th>Volume of Sub-sample (aliquot) (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:00</td>
<td>125</td>
</tr>
<tr>
<td>09:00</td>
<td>125</td>
</tr>
<tr>
<td>12:00</td>
<td>125</td>
</tr>
<tr>
<td>15:00</td>
<td>125</td>
</tr>
<tr>
<td>18:00</td>
<td>125</td>
</tr>
<tr>
<td>21:00</td>
<td>125</td>
</tr>
<tr>
<td>00:00</td>
<td>125</td>
</tr>
<tr>
<td>03:00</td>
<td>125</td>
</tr>
</tbody>
</table>

If you had chosen to collect hourly sub-samples, the number of aliquots in a day would be 24, and the volume of each aliquot would be \(1000/24 = 42\) mL.

(b) **Flow-proportional composite sample**

Like fixed-volume composite samples, for flow-proportional composite samples, the sampling interval can be anything (e.g., every 30 min, every 2 h, or every 3 h). The representativeness of
the sample likewise increases for smaller sub-sample collection intervals. Suppose you choose to collect a 1-L flow-proportional composite sample with 2-h intervals. A total of 12 sub-samples (24/2 = 12) may be collected as shown in the following table and then mixed together to form a composite sample with a volume of at least 1 L.

To determine the sub-sample volume, you need to
- assume an average daily flow rate;
- divide the desired sample volume by the number of sub-samples to get the sub-sample volume for a fixed-volume composite sample;
- measure the flow rate each time a sub-sample is collected;
- calculate a multiplier ratio by dividing the measured flow rate by half of the assumed average daily flow rate;
- multiply the multiplier ratio by the average sub-sample volume.

For the example shown below in the following table, assume that the average daily flow rate is expected to be approximately 2.9 L/s. As a matter of fact, we adopted here the average of the 12 flow measurements. However, in practice, you cannot anticipate the average flow you will have when collecting the sub-samples.

Later, determine the sub-sample volume for a fixed-volume composite sample (1000 mL/12 = 83.3 mL). Then, calculate the multiplier ratio by dividing the measured flow rates by the assumed average daily flow rate. Finally, calculate the sub-sample volume by multiplying the multiplier ratio by 83.3 mL. With these elements, you can construct the following table.

Flow-proportional composite samples (12 aliquots) are shown in the following table.

<table>
<thead>
<tr>
<th>Aliquot number</th>
<th>Measured flow rate (L/s)</th>
<th>Ratio of the flow rate to the average flow rate</th>
<th>Volume of each aliquot (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td>0.416</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>2.4</td>
<td>0.832</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>3.7</td>
<td>1.283</td>
<td>107</td>
</tr>
<tr>
<td>4</td>
<td>4.3</td>
<td>1.491</td>
<td>124</td>
</tr>
<tr>
<td>5</td>
<td>3.8</td>
<td>1.318</td>
<td>110</td>
</tr>
<tr>
<td>6</td>
<td>3.3</td>
<td>1.145</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>3.1</td>
<td>1.075</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>3.7</td>
<td>1.283</td>
<td>107</td>
</tr>
<tr>
<td>9</td>
<td>3.2</td>
<td>1.110</td>
<td>92</td>
</tr>
<tr>
<td>10</td>
<td>2.5</td>
<td>0.867</td>
<td>72</td>
</tr>
<tr>
<td>11</td>
<td>2.0</td>
<td>0.694</td>
<td>58</td>
</tr>
<tr>
<td>12</td>
<td>1.4</td>
<td>0.486</td>
<td>40</td>
</tr>
<tr>
<td>Average</td>
<td>2.9</td>
<td>Total volume</td>
<td>1,000</td>
</tr>
</tbody>
</table>

The profiles of flows and aliquot volumes over time is shown in the chart below. You can clearly see the relationship between flow rate and aliquot volume.
It is important to note that collecting a flow-proportional composite sample requires ‘guessing’ what the average flow rate will be before you start collecting the first sub-sample. If you overestimate the flow rate, the total volume of sample collected will be less than you originally anticipated. If you underestimate the flow rate, your sample volume will be more than you anticipated. It is smart to aim to collect a larger volume than you actually need for your laboratory analysis. For example, you could also have half of the assumed average daily flow rate (i.e., 2.9 L/s/2 = 1.45 L/s) for your sub-sample volume calculations, and this would have resulted in collecting a sample with a total volume of 2000 mL, more than you may have actually needed. You can always discard extra sample, but once you start the flow-proportional composite sampling you cannot go back and collect more volume if you come up short, in case the actual flow rate is less than the anticipated flow rate.

### 3.4 SAMPLE SIZE, CONTAINERS, AND HOLDING TIMES

The size of a sample (its volume or mass), the type of container used, the length of time between sample collection and analysis, and the methods used to preserve the sample prior to analysis all depend on the type(s) of analysis that will be conducted and are generally specified in the standard operating procedure or the standardized method. Every project quality assurance plan should include a table like that in Table 3.7, which outlines the parameters, methods, containers, preservation, and holding times for each analysis.

**Table 3.7** Methods, containers, preservation, and holding times for a selection of analytical and field measurement parameters (adapted from US EPA, 2005).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method Number/Reference</th>
<th>Maximum Holding Time</th>
<th>Container(s)</th>
<th>Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminium, arsenic, calcium, chromium, copper, iron, lead, manganese, magnesium, and zinc</td>
<td>EPA 200.7</td>
<td>6 months</td>
<td>1 × 1-L polyethylene bottle</td>
<td>HNO₃ to pH &lt;2</td>
</tr>
<tr>
<td>Antimony, cadmium, and selenium</td>
<td>EPA 200.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercury</td>
<td>EPA 245.1</td>
<td>28 days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anions (Cl, NO₃, NO₂, PO₄, and SO₄)</td>
<td>EPA 300.0</td>
<td>48 h</td>
<td>1 × 1-L polyethylene bottle</td>
<td>Chill to 4°C</td>
</tr>
<tr>
<td>Total dissolved solids (TDS)</td>
<td>EPA 160.1</td>
<td>7 days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkalinity</td>
<td>SM 2320B</td>
<td>14 days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total coliforms/E. coli</td>
<td>IDEXX Colilert</td>
<td>24 hours</td>
<td>1 × 500-mL polypropylene bottle, autoclaved</td>
<td>Chill to 4°C</td>
</tr>
<tr>
<td>Temperature, pH, and conductivity</td>
<td>Field probe</td>
<td>Immediate</td>
<td>1 × 250-mL mid-mouth glass bottle</td>
<td>None</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>Field probe</td>
<td>Immediate</td>
<td>None, <em>in situ</em> measurement</td>
<td></td>
</tr>
</tbody>
</table>
3.5 STATISTICAL POWER AND NUMBER OF SAMPLES

How many environmental samples should you collect (i.e., how many data points do you need for each statistical sample)? Often, practitioners, students, and scientists are not able to provide a good justification for their answer to this question. In some labs, it may be common to collect a sample size of \( n = 3 \) or \( n = 10 \) as a rule of thumb, but this may not always be the most appropriate sample size for every study. Furthermore, it might make more sense to spread out the samples temporally or spatially, depending on the study objectives, the project budget, and the desired statistical power.

If you want to conduct scientifically sound experiments and make the most use of limited time and funding, you need to use **power calculations** to determine the appropriate sample size. However, before a power test can be performed, you first need to define what type of comparison you want to make and what question you want to answer with your study. Only then can you determine which statistical test is the most appropriate to evaluate that comparison. The type of power calculation needed depends on the statistical test that will be used once you finish collecting your data. Table 3.8 shows three common

<table>
<thead>
<tr>
<th>Study type</th>
<th>Description</th>
<th>Examples</th>
<th>Statistical Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance monitoring</td>
<td>Comparing the average contaminant concentration with a target regulatory compliance limit</td>
<td>A wastewater treatment facility needs to evaluate if the average concentration of ( \text{BOD}_5 ) in the treated effluent is below a regulatory threshold of 30 mg/L. The regulatory guidelines state that the monthly average concentration should be significantly lower than the regulatory threshold</td>
<td>One-sample ( t )-test &lt;br&gt;Sign test &lt;br&gt;Wilcoxon signed-rank test &lt;br&gt;Z-test for proportions &lt;br&gt;Poisson probability of failure/success &lt;br&gt;Frequency analysis and reliability analysis</td>
</tr>
<tr>
<td>Evaluate alternative treatment processes</td>
<td>Comparing two parallel treatment trains (e.g., with different processes or operating conditions) to determine if one performs significantly better than the other with respect to the removal of some contaminant</td>
<td>An advanced water treatment facility utilizes a biological activated carbon filter followed by ultrafiltration and reverse osmosis. You are evaluating the impact of seeding the filter with different water sources to see its effect on downstream fouling</td>
<td>Two-sample ( t )-test &lt;br&gt;Wilcoxon rank sum or Mann–Whitney test &lt;br&gt;Sign test</td>
</tr>
<tr>
<td>Evaluate performance with respect to different factors</td>
<td>This is the type of study you might perform if you want to see how different design, environmental, or other factors influence the performance of a treatment process with respect to the removal of some contaminant</td>
<td>You are developing a new treatment process for the removal of phosphorus, and you would like to better understand how well the system removes phosphorus at different temperatures, ( pH ) levels, and loading rates</td>
<td>ANOVA &lt;br&gt;Factorial analysis &lt;br&gt;Kruskal–Wallis test &lt;br&gt;Regression &lt;br&gt;Correlation</td>
</tr>
</tbody>
</table>
types of studies that are often performed for the assessment of treatment plant performance and describes the statistical test that should be used for each comparison.

**Power calculations** to determine the appropriate sample size for any test start by defining a **level of acceptable error**. Convention is to use **0.05 for the alpha error** and **0.20 for the beta error** (i.e., 80% power). Alpha and beta errors have been briefly mentioned in Section 3.2.6 and are further detailed in Chapter 10.

Next, it is necessary to define the **desired standardized effect size**, also known as **Cohen’s d** (Cohen, 1988). Cohen’s d is calculated as the **difference that you desire to be able to detect (with significance) divided by the standard deviation of the sample mean**. Note that this difference is standardized by the precision with which you can measure the effect (i.e., it is divided by the standard deviation). *The smaller the difference you want to be able to detect with significance, the more samples you will need to analyse (i.e., the more data points you will need for your statistical sample).*

Once you determine the desired standardized effect size for the experiment, the next step is to use a **non-central distribution** to calculate the **beta error** for a given sample size. For example, if you are doing a t-test to compare your samples, you will use a non-central t-distribution. **Central distributions describe the test statistic under the null hypothesis, but non-central distributions describe the test statistic when the null hypothesis is false.** To define a non-central t-distribution for a power analysis, use a **non-centrality parameter that is equal to Cohen’s d multiplied by the square root of the sample size**. Evaluate the non-central distribution at the **critical statistic for your desired alpha level**. The cumulative value of this distribution will be equal to the beta error. Thus, the **power of the test is equal to 1 minus the beta error**.

Power calculations can be **easily performed in several statistical software packages** such as R, Minitab, etc. For a t-test, in order to calculate the **required sample size**, you generally need to provide the following inputs:

- Cohen’s effect size
- desired alpha or type I error (typically 0.05)
- desired beta or type II error level (typically 0.20)
- type of test (one sample or two sample, paired or unpaired, one-sided or two-sided).

The **non-central t-distribution cannot be computed in Excel**, but the Excel spreadsheet for Examples 3.2 through 3.4 of this book contains a custom power calculator, which accesses the non-central t-distribution using a series of look-up tables. Practice using it to calculate statistical power for a given sample size:

- **Example 3.2.** To find the power associated with a particular sample size and a desired effect size.
- **Example 3.3.** To find the required number of samples to detect a desired effect size with a particular power (e.g., 80%).
- **Example 3.4.** To find the minimum effect size that can be detected with a particular power and a particular sample size.

The power calculation for the **two-sample t-test** is similar. The main parameter that changes is the effect size. Instead of being the difference between the sample mean and the regulatory limit, divided by the sample standard deviation, it is equal to the difference between the mean values of the two samples, divided by the pooled standard deviation.

We will show the examples here, but you will need to follow the calculations in the associated Excel spreadsheet, given their complexity and need to use look-up tables.
This topic is an advanced one and uses concepts that are further discussed and detailed in other parts of the book. We opted to keep it here, because it is associated with the planning of your work.

You might need to consult other sections in our book and come back here to have a full grasp of the concepts involved. In special, in Section 10.3.3, we show an iterative procedure for estimating the required sample size for your studies, based on the concepts of hypothesis testing using the t-test. Both procedures lead to the same results.

### EXAMPLE 3.2 DETERMINE POWER BASED ON EFFECT SIZE AND SAMPLE SIZE

The maximum contamination level goal (MCLG) for nitrate in drinking water is 10 mg/L. Suppose you measure the concentration of nitrate in a water source with \( n = 5 \) samples and record a mean concentration of 9.3 mg/L with a standard deviation of 0.5 mg/L.

Using a one-sample, two-sided t-test, a p-value of 0.035 is calculated. See Chapters 9 and 10 for more on how to do a t-test and why some people prefer to use a one-sided t-test. Chapter 9 presents several methods to analyse compliance with a regulatory standard, and Excel spreadsheet for Example 9.2 allows you to do a two-sided one-sample t-test and come to this value of \( p = 0.035 \). This p-value indicates that the measured mean concentration is significantly below the MCLG level (at the 0.05 significance level).

However, the p-value alone does not tell us anything about the beta (type II) error or the power of the analysis. Use the Excel spreadsheet for Example 3.2 to calculate the post hoc power of this statistical analysis.

Note: This example is also available as an Excel spreadsheet.

**Solution:**

The beta error is found to be equal to 33% in this case, meaning that the test only had a power of 67%.

For this particular experiment, you might consider yourself to be ‘lucky’ to have found a significant difference, despite the low power of the experimental set-up. Remember, having a statistical power of only 67% means that you have a two out of three chance of finding a significant difference at the given effect level. This is like being a prosecution attorney and acknowledging that you only collect enough evidence to convict two out of every three guilty people on average.

In the future, it might be more prudent to collect more evidence (i.e., increase your sample size), so that your ‘conviction success rate’ (i.e., your statistical power) is at least 80%.

### EXAMPLE 3.3 DETERMINE SAMPLE SIZE TO ACHIEVE A DESIRED POWER

A wastewater treatment facility needs to determine how many samples need to be collected to determine if the average biochemical oxygen demand (BOD\(_5\)) concentration in a treated effluent is significantly below the regulatory threshold of 30 mg/L. Use the Excel spreadsheet for Example 3.3 to determine the minimum number of samples to ensure that the BOD\(_5\) concentration is significantly below the regulatory threshold with 80% statistical power. Assume a significance level of 0.05, a standard deviation of 4.6 mg/L (this is the assumed standard deviation of repeated BOD\(_5\) measurements in your laboratory from past experiments), and assume that you want to detect an effect size of 2 mg/L. If your desired effect size is 2 mg/L and the standard is set at 30 mg/L, then
the highest mean BOD5 concentration you can measure in the sample and still detect a significant difference from the regulatory threshold is \(30 - 2 = 28\) mg/L.

Note: This example is also available as an Excel spreadsheet.

Solution:

First, Cohen's \(d\) is found to be equal to 0.43, calculated as the difference between the regulatory threshold and the mean BOD5 concentration \((30 - 28 = 2)\), divided by the standard deviation (4.6). Therefore, \(d = 2/4.6 = 0.43\).

This is a one-sample, two-sided \(t\)-test (we use a two-sided test because we assume that the BOD5 concentration could be greater or less than the regulatory value). See Chapters 9 and 10 about whether to use a one-sided test versus a two-sided test.

Determining sample size is generally a trial and error process. Let us start with a typical sample size of \(n = 10\), which is used by default in some labs. The non-centrality parameter (\(\delta\)) is calculated by multiplying Cohen’s \(d\) by the square root of the sample size. Therefore, \(\delta = 0.43 \times \sqrt{10} = 1.36\).

The type II (beta) error is calculated by looking up the value of the non-central \(t\)-distribution table for the critical value associated with the alpha level and sample size chosen, as well as the non-centrality parameter, equal to Cohen’s \(d\) multiplied by the square root of the sample size.

For a sample size of \(n = 10\), the statistical power is only 23%. In order to achieve a power of 80%, we need to increase the sample size to at least \(n = 43\) data points. Therefore, environmental samples should be collected approximately weekly in order to acquire at least 43 data points throughout the year. However, using the Excel spreadsheet, you can see that if you increase the number of samples, you will see that the power also increases (e.g., for a sample size of \(n = 100\), the power is 99%).

EXAMPLE 3.4 DETERMINE THE EFFECT LEVEL GIVEN A SAMPLE SIZE AND A DESIRED POWER

A stormwater authority is investigating the contribution of agricultural runoff to phosphate pollution in a stream during storm events. To do this, they plan to collect samples upstream and downstream of the agricultural field and measure the concentration of phosphate in the upstream and downstream locations. If the difference between paired phosphate concentrations upstream and downstream is significantly greater than zero, they will determine that the runoff from the agricultural site is contributing phosphate pollution to the stream.

Assume a standard deviation of 0.44 mg/L (i.e., the standard deviation of the differences between paired phosphate concentrations). What effect size can be detected at a significance level of 0.05 with 80% power if the stream is only sampled during three storm events (i.e., experiment done in triplicate)?

Note: This example is also available as an Excel spreadsheet.

Solution:

This is a two-sample paired \(t\)-test. We can assume it is one-sided because our hypothesis is that the agricultural site will contribute phosphate to the river, making the concentrations relatively higher downstream, rather than the opposite.

With these assumptions and our standard deviation of 0.44 mg/L, we can determine that a sample size of only \(n = 3\) only allows us to detect a difference of 1.0 mg/L between upstream and downstream concentrations with a power of 80%.
If we want to detect a smaller difference between the upstream and downstream concentrations, the power of the test will be lower. For instance, if we change the effect size to 0.5 mg/L, we see that the power goes down to only 40%. To maintain a power of 80% and detect a difference of 0.46 mg/L between upstream and downstream concentrations, we would need to sample at seven different storm events.

### 3.6 CHECK-LIST FOR YOUR REPORT

- ✓ Check that quality assurance and quality control measures are summarized in a chapter of your report or as a separate, stand-alone report. In particular, make sure that you address the scope of the study, the type and anticipated use of the data, any relevant assessment thresholds, standard operating procedures, quality control samples, and data storage and management protocols.
- ✓ Confirm that quality control is demonstrated as acceptable precision and accuracy through an initial demonstration of capability and through ongoing demonstrations of capability, performed quarterly at a minimum.
- ✓ Verify that sample locations and sample types (e.g., grab versus composite) are described in detail, with appropriate consideration for anticipated temporal and/or spatial variabilities.
- ✓ Check that sample matrix, sample volume or mass, sample analysis methods, sample containers, sample preservation, and maximum holding times are defined for each parameter to be analysed and summarized (preferably in a table).
- ✓ Verify that acceptable type I (alpha) error and type II (beta) error levels are established.
- ✓ Confirm that the desired effect size has been established as well as the anticipated standard deviation between samples.
- ✓ Verify that the sample size has been determined using a power analysis for the desired alpha and beta error levels.