Water Quality Monitoring and Analysis

A Short Course For Environmental Monitoring Officers

Environmental Impact Assessments Limited, 21 Tui Glen Road, North Shore City 0626 Auckland Mobile 021 1104814 email: brett@eia.net.nz

WATER QUALITY MONITORING AND ANALYSIS - A SHORT COURSE FOR ENVIRONMENTAL MONITORING OFFICERS

BRETT STANSFIELD

ENVIRONMENTAL IMPACT ASSESSMENTS
MAY 2026

ENVIRONMENTAL IMPACT ASSESSMENTS LIMITED
21 TUI GLEN ROAD
BIRKENHEAD
NORHT SHORE CITY
AUCKLAND 0626

Table of Contents

The Physical Properties of Water	1
The Water Cycle	**5**
Water Quality Defined	**5**
Why Monitor?	**6**
Water Quality Monitoring Program Checklist	**8**
Course Structure	**8**
Water Quality Variables To Measure	9
Water temperature	**9**
pH	**10**
Synergistic Effects of pH	10
Total Alkalinity	**11**
Dissolved Oxygen	**11**

Electrical Conductivity **12**

Turbidity **12**

Total suspended solids **12**

Water Clarity **13**

Nutrients **14**

Phosphorus Cycle **15**

Total Phosphorus **16**

Soluble Reactive Phosphorus **17**

The Nitrogen Cycle **17**

Ammonia **18**

Nitrate and nitrite nitrogen **19**

Dissolved Inorganic Nitrogen **19**

Total Kjeldahl Nitrogen **19**

Contaminants In The Urban Environment **20**

Lead **20**

Copper **20**

Zinc **21**

Polycylic Aromatic Hydtrocarbons **22**

Monitoring Water Microbiology **22**

Detecting the source of contamination **23**

Monitoring Stream or Lake Ecology **25**

Chlorophyll a **27**

Macroinvertebrates **27**

Fish **28**

Undertaking a Stream Gauging **29**

Estimating Pollution Loading, Yield and Final Concentrations 31
Water Quality Indices 34
Macroinvertebrate Water Quality Indices **34**
The Index of Biotic Integrity **36**
Lake Indices **40**
Trophic Level Index **41**
Percentage Average Change **41**
The Water Quality Index **43**
Site Selection and Survey Design 45
Survey Design. **48**
Sampling Error **48**
Randomised Sampling **49**
Systematic Sampling **49**
Preparing Data For Exploratory Data Analysis 51
Exploratory Data Analysis **51**
Common Notation In Statistics **53**
Transforming Data To Achieve Normality **53**
Testing Hypotheses **58**
Graphical Techniques To Demonstrate Water Quality 59
Site comparisons **59**
Correlations with other variables or sites **61**
Examining the similarity of biological communities **62**
Statistical Analysis And Tools To Determine Environmental Significance 67
Regression Analysis **67**
Residual Analysis **68**

Interpreting A Regression Analysis **71**

What can be done if the regression is poor **73**

Multiple Linear Regression **74**

Forward Stepwise **75**

Backward Elimination **77**

Conventional Stepwise Regression **80**

Analysis of variance **80**

Two Way ANOVA **82**

Calculating a percentile **85**

Estimating The Correct Sample Size For A Compliance Programme **89**

Determining Appropriate Sample Size For Resource Surveys **90**

Non Parametric Methods **91**

Two Dependent Samples **92**

Two Independent Samples **95**

Greater Than Two Independent Samples **98**

Correlation Analysis **103**

Time Series Analysis **106**

The Moving Average Model **110**

Exponential (Average) Smoothing **110**

Lowess **111**

The Mann-Kendall Trend Test **113**

The Sen Slope Estimator **114**

Seasonal Kendall Trend Test **114**

Seasonal Kendall Sen Slope Estimator **114**

Determining Environmental Significance **115**

The Physical Properties of Water

Our planet is dominated by water with more than 70% of its surface covered with this simple molecule. Scientists estimate that the hydrosphere contains about 1.36 billion cubic kilometers of this substance mostly in the form of a liquid (water) that occupies topographic depressions on the earth. The second most common form of the water molecule on our planet is ice. If all our planet's ice melted sea level would rise by about 70 meters.

Water is essential for life comprising the major constituent of almost all life forms. Most animals and plants contain more than 60% water by volume. Without water life would probably never have developed on our planet.

Water has a very simple atomic structure. This structure consists of two hydrogen atoms bonded to one oxygen atom. The nature of the atomic structure of water causes its molecules to have unique electrochemical properties. The hydrogen side of the water molecule has a slight positive charge while on the other side of the molecule has a negative charge. This molecular polarity causes water to be a powerful solvent and is responsible for its strong surface tension.

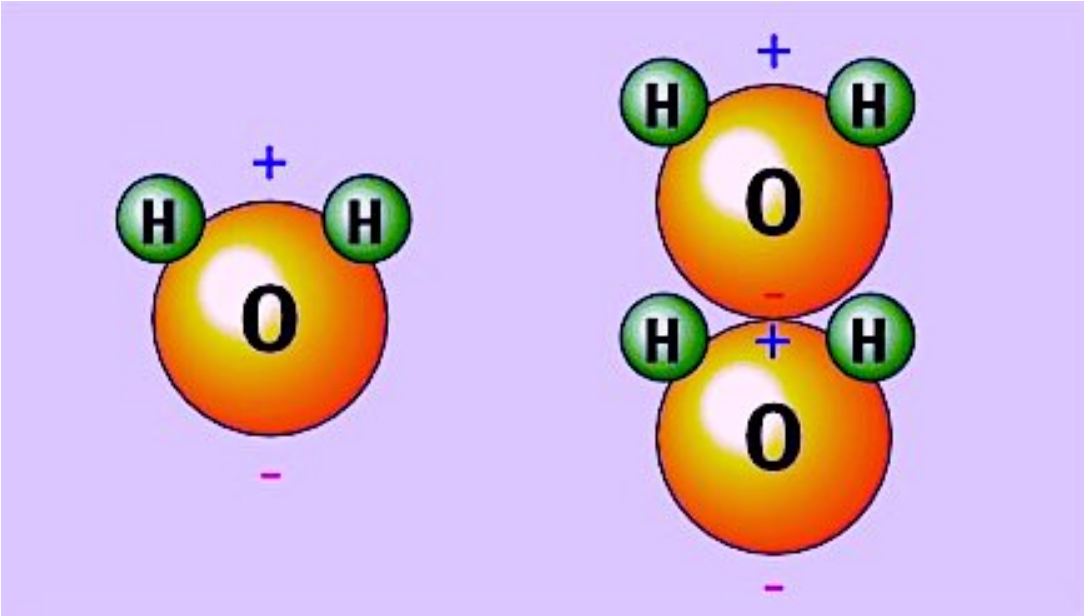


Figure 1: The atomic structure of a water (or dihydrogen monoxide) molecule

When the water molecule makes a physical change (e.g. from ice to water to vapour) its molecules arrange themselves in distinctly different patterns. The molecular arrangement taken by ice (the solid form of the water molecule) leads to an increase in volume and a decrease in density. Expansion of the water molecule at freezing allows ice to float on top of liquid water.

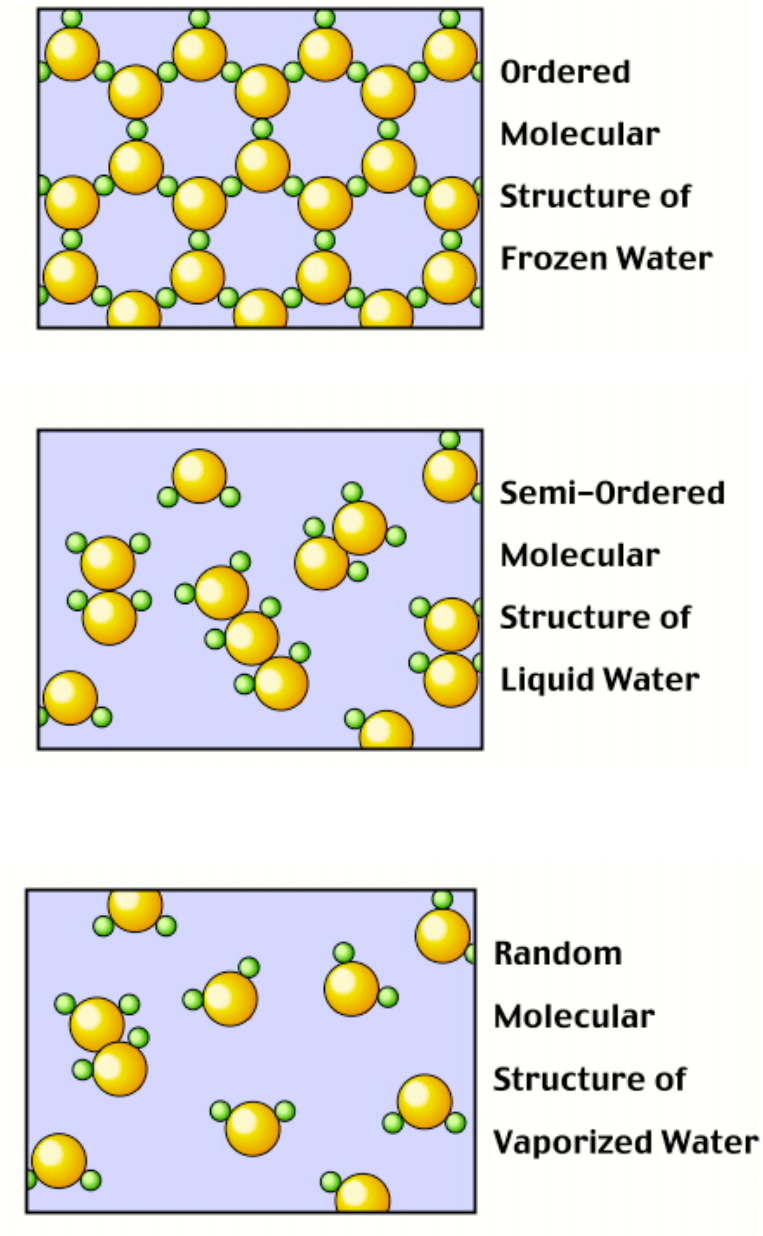


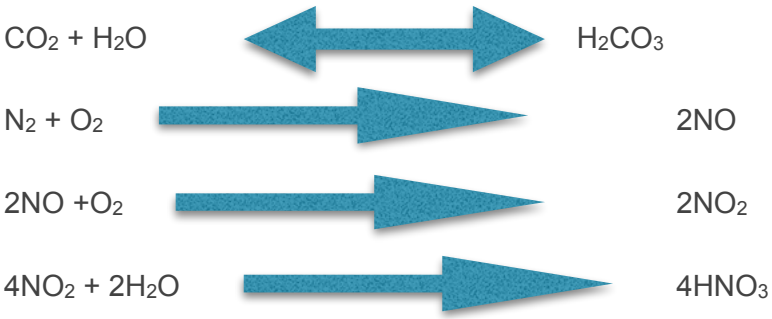
Figure 2: Molecular structure of water in the solid, liquid and gas phases.

The three diagrams illustrate the distinct arrangement patterns of water molecules as they change their physical state from ice to water to gas. Frozen water molecules arrange themselves in a particular highly organised rigid geometric pattern that causes the mass of water to expand and to decrease in density. The diagram above shows a slice through a mass of ice that is one molecule wide. In the liquid phase, water molecules arrange themselves into small groups of joined particles. The fact that these arrangements are small allows liquid water to move and flow. Water molecules in the form of a gas are highly charged with energy. This high energy state causes the molecules to be always moving reducing the likelihood of bonds between individual molecules from forming.

Water has several other unique physical properties. These properties are:

- Water has a high specific heat (the heat required to raise its temperature). Because water has a high specific heat, it can absorb large amounts of heat energy before it begins to get hot. It also means that water releases heat energy slowly when situations cause it to cool. This is why maximum water temperatures in rivers are often well after solar noon (usually around 3 p.m.) while for lakes it can be closer towards evening. Water's high specific heat allows for the moderation of the Earth's climate and helps organisms regulate their body temperature more effectively.

- Water in a pure state has a neutral pH (discussed in next chapter). As a result, pure water is neither acidic (low pH) or basic (high pH). Water changes its pH when substances are dissolved in it. Rain has a naturally acidic pH of about 5.6 because it contains natural derived carbon dioxide and sulfur dioxide. Rainfall events can often lead to depressed pH levels in small streams owing to the rainfall entering them.



- Water conducts heat more easily than any liquid except mercury.

- Water molecules exist in liquid form over an important range of temperature from 0 - 100° Celsius. This range allows water molecules to exist as a liquid in most places on our planet.

- Water is a universal solvent. It is able to dissolve a large number of different chemical compounds. This feature also enables water to carry solvent nutrients in runoff, infiltration, groundwater flow and living organisms.

- Water has a high surface tension. In other words, water is adhesive and elastic, and tends to aggregate in drops rather than spread out over a surface as a thin film. This phenomenon also causes water to stick to the sides of vertical structures despite gravity's downward pull. Water's high surface tension allows for the formation of water droplets and waves, allows plants to move water (and dissolved nutrients) from their roots to their leaves, and the movement of blood through tiny vessels in the bodies of some animals.

The following illustration shows how water molecules are attracted to each other to create high surface tension. This property can cause water to exist as an extensive thin film over solid surfaces. In the example overleaf, the film is two layers of water molecules thick.

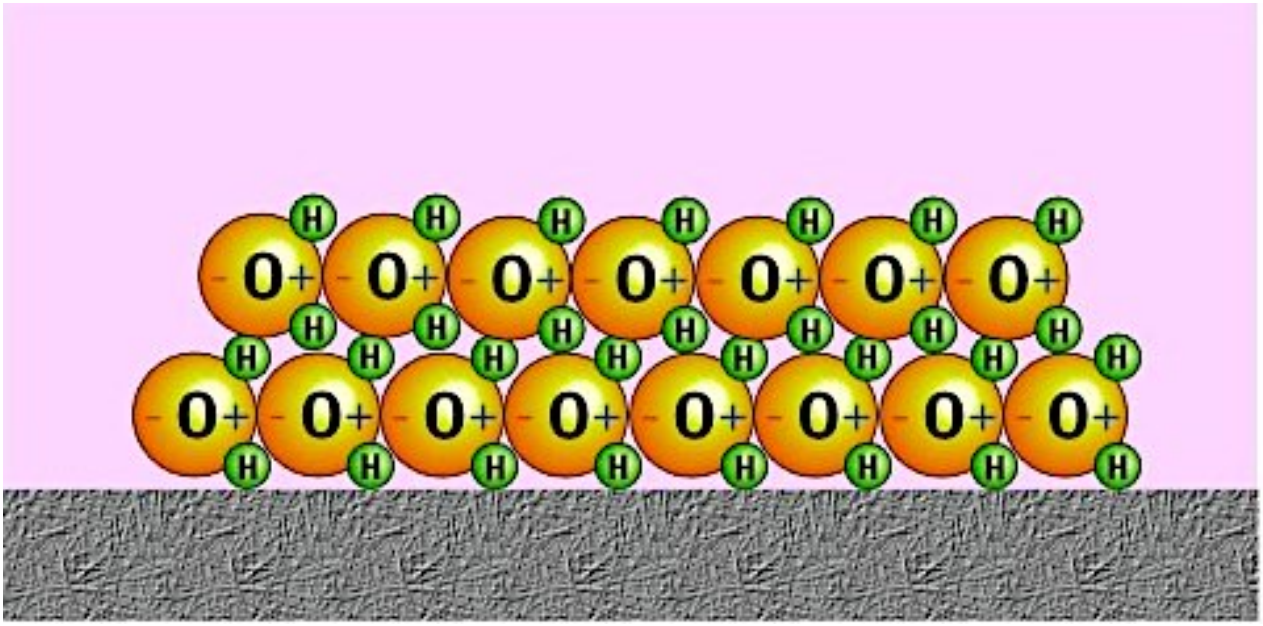


Figure 3: The aggregation of water molecules.

Figure 4: The adhesive bonding property of water molecules allows for the formation of water droplets (Photo © 2004 **Edward Tsang**).

Water molecules are the only substance on Earth that exist in all three physical states of matter: solid, liquid and gas. Incorporated in the changes of state are massive amounts of heat exchange. This feature plays an important role in the redistribution of heat energy in the Earth's atmosphere. In terms of heat being transferred into the atmosphere, approximately 3/4's of this process is accomplished by the evaporation and condensation of water.

The freezing of water molecules causes their mass to occupy a larger volume. When water freezes it expands rapidly adding about 9% by volume. Fresh water has a maximum density at around 4° Celsius. Water is the only substance on this planet where the maximum density of its mass does not occur when it becomes solidified.

All of these unique properties of water are fundamental for its cycling (the water cycle) to proceed.

The Water Cycle

Water is constantly being cycled between the atmosphere, the ocean and land. This cycling is a very important process that helps sustain life on Earth.

As the water evaporates, vapours rise and condense into clouds. The clouds move over the land, and precipitation falls in the form of rain, ice or snow. The water fills streams and rivers, and eventually flows back into the oceans where evaporation starts the process anew.

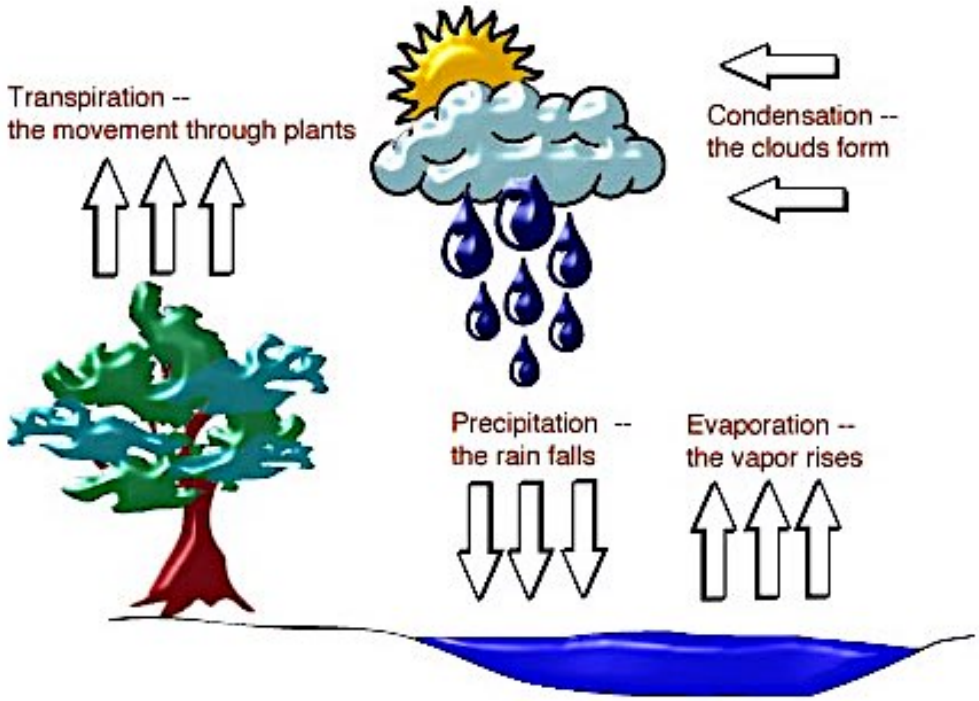


Figure 5: The water cycle

Water Quality Defined
What do we mean when we mention the words water quality?

To many people water quality can mean different things. To a trout fisherman, it could mean how good the fishing has been, the quality of spawning habitat for the fish or the natural character of the river. To a farmer it may mean how good the water is for his stock to drink. To iwi it may equate to the cultural value of a stream, its ‘mahinga kai’ (food gathering properties), its sinuosity, the sound of the water, the smell of the water, the taste of the water, the colour or its spiritual value. To a swimmer it could mean how safe the water is to swim in. To a tramper it could be the natural

character of the surrounding catchment. To a biologist it could mean the diversity of species that live within that water body.

Clearly defining water quality encompasses many values. When managing water it is first necessary to determine what values you are wanting to protect or enhance before you can proceed with any monitoring programme design or subsequent analysis.

The field of water quality is an evolving science which is broad and complex. It includes many fields including environmental decision making, aquatic ecology, microbiology, statistics, water chemistry, ecotoxicology (the toxicity of contaminants to our biota), data management, large scale hydrology and many other sciences. Because of its multidisciplinary manner, few practitioners have all the necessary expertise to develop laws or implement successful monitoring programmes. In New Zealand a large degree of collaboration has overcome these obstacles giving rise to excellent water quality surveillance programmes.

Why Monitor?

The reason we monitor and analyse water quality is primarily to determine

- State - is the water quality at this site good or bad? How does it compare to other sites within the catchment or surrounding catchments? Does the water quality comply with environmental guidelines?

- Trends - has the water quality improved or deteriorated over time? Is this trend environmentally significant?

- Compliance - has the sewage treatment discharge upstream caused the water to deteriorate? Has this discharge had an adverse impact on the health of the river downstream?

- Research - trialing new sampling techniques or gathering data to better understand the resource

Before embarking on any water quality monitoring programme, it is important to have defined the question you are trying to answer.

The primary goal of monitoring is to provide information which can be used to improve our water management. The following diagram illustrates how water quality monitoring and analysis contributes to the overall monitoring cycle.

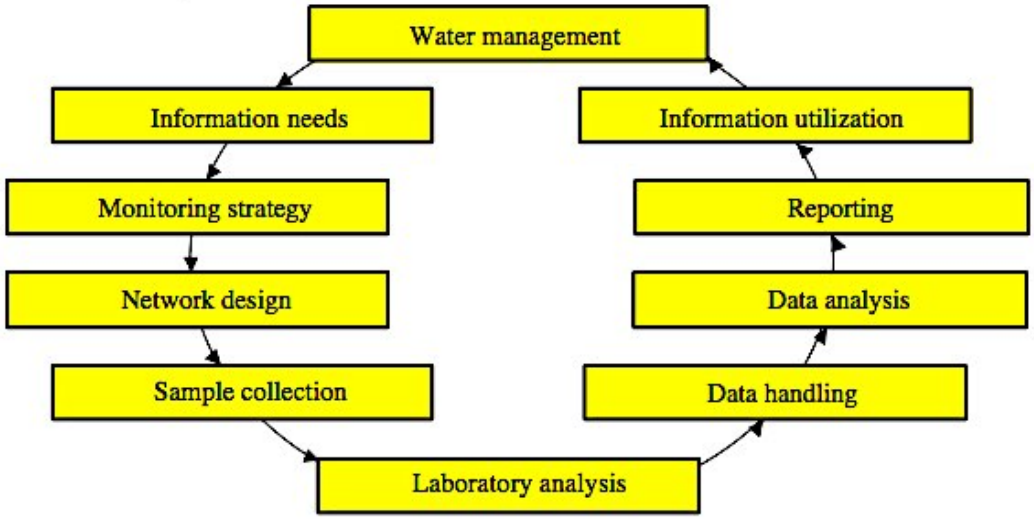


Figure 6: The monitoring cycle

Starting from the water sample collection, it gets sent to a laboratory who then send us the data. The data is then examined for any outliers (extreme high or low values that are perhaps laboratory error) contact the lab, re analyse to confirm, remove any symbols from the data set e.g. < or > values to produce a data set ready for analysis, analyse the data and report on it effectively turning data into information which then goes to the water manager who then says “has this answered my question? if not why not?, what then are my information needs? or have my information needs changed?, if they have then review the overall monitoring strategy, possibly alter the network of sites and then get back to the sampling.

In terms of reviewing your monitoring strategy and network design there are some caviots to be aware of. Firstly if your question is to determine the state of the environment in your region, then you need to be careful not to move or delete sites unless absolutely necessary. This is because long term data sets have far greater information value than short term data sets. With longer term data sets you are in a better position to identify seasonal (e.g. warmer water temperatures in summer) or cyclical trends(e.g. 11 year el nino or la nina) from other patterns in water quality.

For state of the environment monitoring you are wanting to know the ‘long term state and trends in water quality’. NIWA’s rivers national water quality network has been untinkered with since its conception in 1989 furthermore the laboratory analysis methods have not changed since this time to ensure that improved analytical techniques do not bias any trends in the data.

Another caviot to be aware of is that if you change the management question too much, then it may be necessary to consider a completely new monitoring programme rather than tinkering with the older monitoring programme which was not designed for that purpose.

To give you an example, I was once asked to conduct a targeted investigation on how polluted a stream was. After a couple of months monitoring, I was in a position to state that the stream was particularly polluted. My manager then said “ well how polluted is this stream compared to other streams?”, to which I replied “ you have changed the question”. The manager was of the opinion that I could just tack on another site from another catchment and therefore provide further information as to how polluted this stream was compared to a stream in a catchment nearby. Unfortunately we were three quarters through the budget for the project and to suddenly bring in

additional sites would result in a small data set for the second question to the point that it did not have good statistical precision to answer it.

Water Quality Monitoring Program Checklist

Ideally your monitoring survey design should include

•monitoring the appropriate water quality variables (e.g. dissolved oxygen, pH, turbidity etc) that are likely to indicate change as a result of the activity being investigated.

•monitoring at the appropriate replication and frequency to determine trends. Sites should also be selected based on their representativeness and should include at least one control site (a site that is not subject to the effect being studied)

•standard methods should be used for water sampling, laboratory analysis and statistical analysis to ensure consistency in reporting

•a periodic review of the programme to increase its effectiveness (e.g. has it already answered the question?, if not why not?, should newer techniques be deployed to improve the information? has the programme revealed other water quality problems?)

•a quality assurance programme to ensure reliability of data

•a budget that is not subject to major cuts, otherwise the question may not be answered.

Course Structure

This course comprises 7 sections including

•Water quality

•Water quality variables to measure

•Estimating pollution loads, yields and final concentrations

•Water quality indices

•Site selection and survey design

•Preparing data for exploratory analysis

•Graphical techniques to demonstrate water quality

•Statistical analysis of data and tools to determine environmental significance

Water Quality Variables To Measure

There are probably over one thousand water quality variables that could be measured from a water body. There is obviously not time or pages to cover all of these, therefore in this section I will focus on water quality contaminants that are commonly encountered in rivers , lakes and streams in agricultural or urban environments. The variables have been divided according to physical, chemical, biological and microbiological.

Water temperature

Water temperature is an essential variable to be measuring in any aquatic ecosystem. This is because it governs the productivity and metabolism of the ecosystem (how quickly things are processed). Many water quality variables (such as pH, dissolved oxygen and electrical conductivity) are dependent on the water temperature therefore having the temperature reading at hand gives greater meaning for these other measurements. Most hand held water quality meters have temperature compensation in their readings i.e. they adjust the pH, dissolved oxygen or conductivity readings to compensate for the water temperature. The solubility (ability to dissolve) of some toxic compounds (e.g. ammonia) are also temperature dependent, therefore measuring temperature is a must.

The water temperature of a water body is largely governed by the surrounding climate. Water temperatures in rivers and lakes are diurnally variable (vary during the day). For rivers the temperature peak usually comes around 3 p.m. in the afternoon while for lakes it is closer towards the evening. The temperature of the water will determine what species of plants and animals inhabit a water body therefore monitoring water temperature helps biologists interpret whether the absence of a species may be due to pollution or a natural physical feature of the water body.

Variables that affect a waterway's temperature include:

1. The colour of the water. Most heat warming surface waters comes from the sun, so waterways with dark-coloured water, or those with dark muddy bottoms, absorb heat best.
2. The depth of the water. Deep waters usually are colder than shallow waters simply because they require more time and energy to warm up.
3. The amount of shade received from shoreline vegetation. Trees overhanging a lake shore or river bank shade the water from sunlight. Some narrow creeks and streams are almost completely covered with overhanging vegetation during certain times of the year. The shade prevents water temperatures from rising too fast on bright sunny days.
4. The latitude of the waterway. Lakes and rivers in cold climates are naturally colder than those in warm climates.
5. The time of year. The temperature of waterways varies with the seasons.
6. The temperature of the water supplying the waterways. Some lakes and rivers are fed by cold mountain streams or underground springs. Others are supplied by rain and/or surface run-off. The temperature of the water flowing into a lake, river or stream helps determine its temperature.
7. The volume of the water. the more water there is, the longer it takes to heat up or cool down.
8. The temperature of effluents dumped into the water. When people dump heated effluents into waterways, the effluents raise the temperature of the water.

pH

Water consists of hydrogen ions (H+) and hydroxide ions (OH-) A pH-measurement is determined by the number of H+ ions and the number of hydroxide (OH-) ions in solution. When the number of H+ ions equals the number of OH- ions, the water is neutral and will than have a pH of about 7. The pH is a very important factor, because certain chemical processes can only take place when water has a certain pH.

The pH of water can vary between 0 and 14. When the pH of a substance is above 7, it is a basic substance. When the pH of a substance is below 7, it is an acid substance. The pH is a logarithmic scale therefore when a solution becomes ten times more acidic, the pH will fall by one unit. When a solution becomes a hundred times more acidic the pH will fall by two units.

The word for pH is short for “pondus Hydrogenium”. This literally means the weight of hydrogen.

The pH of water determines the solubility (amount that can be dissolved in the water) and biological availability (amount that can be utilized by aquatic life) of chemical constituents such as nutrients (phosphorus, nitrogen, and carbon) and heavy metals (lead, copper, cadmium, etc.). For example, in addition to affecting how much and what form of phosphorus is most abundant in the water, pH also determines whether aquatic life can use it. In the case of heavy metals, the degree to which they are soluble determines their toxicity. Metals tend to be more toxic at lower pH because they become more soluble.

pH is diurnally variable due to algal photosynthesis during the day raising pH (caused by the release of carbonates) and algal respiration at night lowering pH (caused by the release of bicarbonates).

Synergistic Effects of pH

Synergy is the process whereby two or more substances combine and produce effects greater than their sum. For example, 2 + 2 = 4 (mathematically). But synergistically, 2 + 2 = much more than 4! Synergy is a mathematical impossibility, but it is a chemical reality. Here’s how it works:

When acid waters (waters with low pH values) come into contact with certain chemicals and metals, they often make them more toxic than normal. As an example, fish that usually withstand pH values as low as 4.8 will die at pH 5.5 if the water contains 0.9 mg/L of iron. Mix an acid water environment with small amounts of aluminium, lead or mercury, and you have a similar problem—one far exceeding the usual dangers of these substances.

The pH of sea (salt) water is not as vulnerable as fresh water’s pH to acid wastes. This is because the different salts in sea water tend to buffer the water with Alka-Seltzer-like ingredients. Normal pH values in sea water are about 8.1 at the surface and decrease to about 7.7 in deep water. Many shellfish and algae are more sensitive than fish to large changes in pH, so they need the sea’s relatively stable pH environment to survive.

Shallow waters in subtropical regions that hold considerable organic matter often vary from pH 9.5 in the daytime to pH 7.3 at night. Organisms living in these waters are able to tolerate these extremes or swim into more neutral waters when the range exceeds their tolerance.

Total Alkalinity

Alkalinity is not a pollutant. It is a total measure of the substances in water that have "acid-neutralizing" ability. Don't confuse alkalinity with pH. pH measures the strength of an acid or base; alkalinity indicates a solution's power to react with acid and "buffer" its pH — that is, the power to keep its pH from changing. Therefore streams with high alkalinity might stand up better to an acid spill than streams of lower alkalinity.

Absolutely pure water has a pH of exactly 7.0. It contains no acids, no bases, and no (zero) alkalinity. The buffered water, with a pH of 6.0, can have high alkalinity. If you add a small amount of weak acid to both water samples, the pH of the pure water will change instantly (become more acid). But the buffered water's pH won't change easily because the Alka-Seltzer-like buffers absorb the acid and keep it from "expressing itself."

Alkalinity is important for fish and aquatic life because it protects or buffers against pH changes (keeps the pH fairly constant) and makes water less vulnerable to acid rain. The main sources of natural alkalinity are rocks, which contain carbonate, bicarbonate, and hydroxide compounds. Borates, silicates, and phosphates may also contribute to alkalinity.

Limestone is rich in carbonates, so waters flowing through limestone regions generally have high alkalinity — hence its good buffering capacity. Conversely, granite does not have minerals that contribute to alkalinity. Therefore, areas rich in granite have low alkalinity and poor buffering capacity.

Dissolved Oxygen

Fish, invertebrates, plants, and aerobic bacteria all require oxygen for respiration.
Much of the dissolved oxygen in water comes from the atmosphere. After dissolving at the surface, oxygen is distributed by current and turbulence. Algae and rooted aquatic plants also deliver oxygen to water through photosynthesis.

The main factor contributing to changes in dissolved oxygen levels is the build-up of organic wastes. Decay of organic wastes consumes oxygen and is often concentrated in summer, when aquatic animals require more oxygen to support higher metabolisms.

Depletions in dissolved oxygen can cause major shifts in the kinds of aquatic organisms found in waterbodies. Temperature, pressure, and salinity affect the dissolved oxygen capacity of water.

Dissolved oxygen concentrations in water is diurnally variable (varies during the day). This is mostly due to photosynthesis by plants (producing oxygen) during the day and respiration (breathing in oxygen) during the night. The variability of the diurnal swings in dissolved oxygen are usually dampened in faster flowing water owing to the constant re aeration provided by the turbulence of the water.

$6CO_2 + 6H_2O$ +light → $C_6H_{12}O_6 + 6O_2$ (Photosynthesis)

$C_6H_{12}O_6 + 6O_2$ → $6CO_2 + 6H_2O$ +Energy (Respiration)

The ratio of the dissolved oxygen content (ppm) to the potential capacity (ppm) gives the percent saturation, which is also an indicator of water quality.

Dissolved oxygen is normally measured using a hand held meter.

Electrical Conductivity

As the name implies electrical conductivity measures the waters capacity to conduct electricity. Polluted waters generally have higher electrical conductivity readings than pristine waters because of all the dissolved materials (or total dissolved solids) in the water giving rise to a greater potential to conduct electricity. There are however some exceptions to this rule because the waters electrical conductivity can be affected by the catchment geology. For example limestone catchment streams tend to have higher EC because of the dissolution of carbonate minerals in the basin. Spring fed streams tend to have higher EC because the water entering the stream has come from groundwater which has spent a long time passing through rocks and dissolving material along the way.

Provided you keep these external factors in mind, electrical conductivity can be a good surrogate for measuring pollution levels in a water body.

Turbidity

Turbidity essentially measures how cloudy the water is. It is usually measured using a hand held meter. Water is placed in a vial and then placed inside the meter whereupon a small beam of light passes through the vial. The turbidity measurement is made by comparing the light intensity at one side of the vial with the other. Light's ability to pass through water depends on how much suspended material is present. Turbidity may be caused when light is blocked by large amounts of silt, microorganisms, plant fibres, sawdust, wood ashes, chemicals and coal dust. Any substance that makes water cloudy will cause turbidity. The most frequent causes of turbidity in lakes and rivers are plankton and soil erosion from logging, mining, and dredging operations.

You also can measure turbidity by filtering a water sample and comparing the filter's colour (how light or dark it is) to a standard turbidity colour chart.

Total suspended solids

Some of the most severe impacts of urbanisation on stream invertebrates are often caused by suspended solids. These are washed into streams either during initial land development or they enter as a result of stream bank erosion created by newly created impervious areas. Snelder and Trueman (1995) in an Auckland study demonstrated a threefold widening of a stream after 85% of the catchment had been urbanised and demonstrated subsequent increases in suspended solids concentrations. The very high suspended solids loads characteristic of streams in recently developed urban areas can take up to 20-30 years to return to levels more typical of mature urban areas (Williamson 1993).

Continuous supplies of suspended sediments can have dramatic effects on stream ecosystems (Ryan 1991). Graham (1990) demonstrated that silt laden algae were less attractive to invertebrate

grazers than clean algae and Ryder (1989) demonstrated that mayfly (*Deleatidium spp.*) nymphs and caddisfly (*Pycnocentrodes sp.*) larvae grazed preferentially on unsilted rather than silted periphyton. Thus suspended sediment loads in urban streams have the potential to reduce invertebrate communities by altering the benthic food quality.

Suspended sediments can also clog interstices of stream bed substrates (Marshall 1974; Ryan 1991) leading to a reduction in water exchange with surface waters and causing the interstitial layer to become deoxygenated. Silt deposition and associated changes in habitat generally bring about a change in invertebrate communities, with a loss of stonefly and mayfly species and an increase in densities of more pollution tolerant fauna (Marshall 1974; Ryan 1991). Total suspended solids is measured by filtering water through a pre weighed paper water filter and then drying the filtrate (paper plus whatever solids were trapped by it). The difference in weight between the pre weighted filter paper and the filter paper after water filtration is the suspended solids content. Normally a measured volume is filtered and the results are expressed in milligrams per litre (mg/l)

Water Clarity

Water clarity in rivers is measured using a horizontal black disc measuring kit. The black disc is placed in a fairly slow flowing area of the river and a measurement is made at which the outer edge of the black disc can no longer be seen using a horizontal viewer.

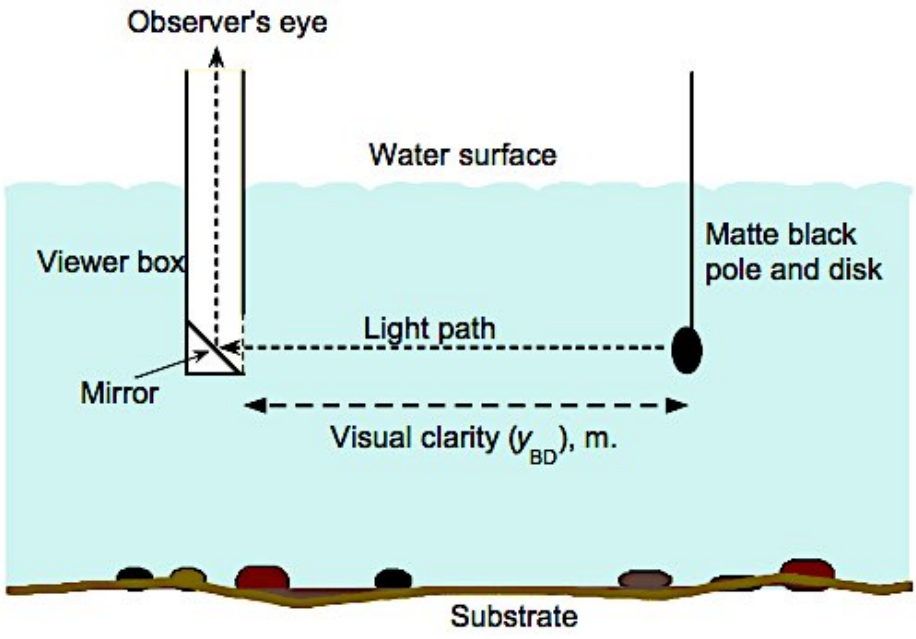


Figure 7: The horizontal black disc apparatus

For lakes a black and white disc is lowered into the water column and viewed using a black box with clear perspex at the bottom. The water clarity in this instance is the distance when the white and black colours can no longer be deciphered.

Figure 8: Secchii disc used for measuring lake water clarity

Measuring water clarity in a lake enables us to determine the photic depth (depth at which plants have enough light for photosynthesis).

There is usually a strong correlation between water clarity, turbidity and suspended solids as they are all measuring similar aspects of water clarity. They are useful variables to use if you wished to monitor the effects of any sediment erosion impacts to a water body. Of the three measurements, water clarity using a disc is the most reliable. This is because when rivers rise towards a flood or fresh (small flood) condition a greater diversity of substrate sizes are suspended in the water column (e.g. fine silts to sand sized particles), this gives rise to a greater variability in measurements of turbidity or suspended solids.

Nutrients

Living organisms need nutrients to live and grow and can be classified according to macronutrients (needed in larger quantities for survival) and micronutrients (needed in smaller quantities for survival). Water quality monitoring programmes usually focus on the macronutrients nitrogen (N) and phosphorus (P) as they are important for driving primary productivity (rates of photosynthesis and plant biomass increase). Because N and P are essential for growth, they are often referred to as limiting nutrients.

If these nutrients reach high concentrations then periphyton (algae that grows on stones) and phytoplankton (planktonic algae) blooms can occur. So while N and P are essential for life, at high concentrations they can be regarded as contaminants (pollutants) owing to the disruption in balance of the ecosystem that is produced. If primary productivity becomes high it can lead to a large depletion in dissolved oxygen due to the respiration required by algal grazers (macroinvertebrates and bacteria) to break down the plant matter. The smothering of the stream bed by periphyton can also lead to the loss of preferred habitat for sensitive species of macroinvertebrates.

Both nitrogen and phosphorus have cycles which I will now discuss as a means to portray how they are used in aquatic ecosystems.

Phosphorus Cycle

The phosphorus cycle is a biogeochemical cycle (one which cycles through living systems and the earths surface) that describes the movement of phosphorus through the lithosphere (surface of the earth), hydrosphere (wherever water is found on the planet), and biosphere (all ecosystems). Unlike many other biogeochemical cycles, the atmosphere does not play a significant role in the movement of phosphorus, because phosphorus and phosphorus-based compounds are usually solids at the typical ranges of temperature and pressure found on earth.

Phosphorus normally occurs in nature as part of a phosphate ion, consisting of a phosphorus atom and some number of oxygen atoms, the most abundant of which is orthophosphate (PO_4^{-3}). Most phosphates are found as salts in ocean sediments or in rocks. Over time, geologic processes can bring ocean sediments to land, and weathering will carry these phosphates to aquatic and terrestrial habitats. Aquatic plants absorb phosphates from the stream or lake substrate, while periphyton and phytoplankton absorb phosphates directly from the water column. The plants absorb the phosphates binding it into organic compounds. The plants and algae may then be consumed by herbivores which in turn can be consumed by carnivores. After death the animal or plant decays and the phosphates then bind to the stream or lake bed sediments.

A great global presentation on the P cycle can be found at http://www.wadsworthmedia.com/biology/starr_udl11_tour/phos_anim.html

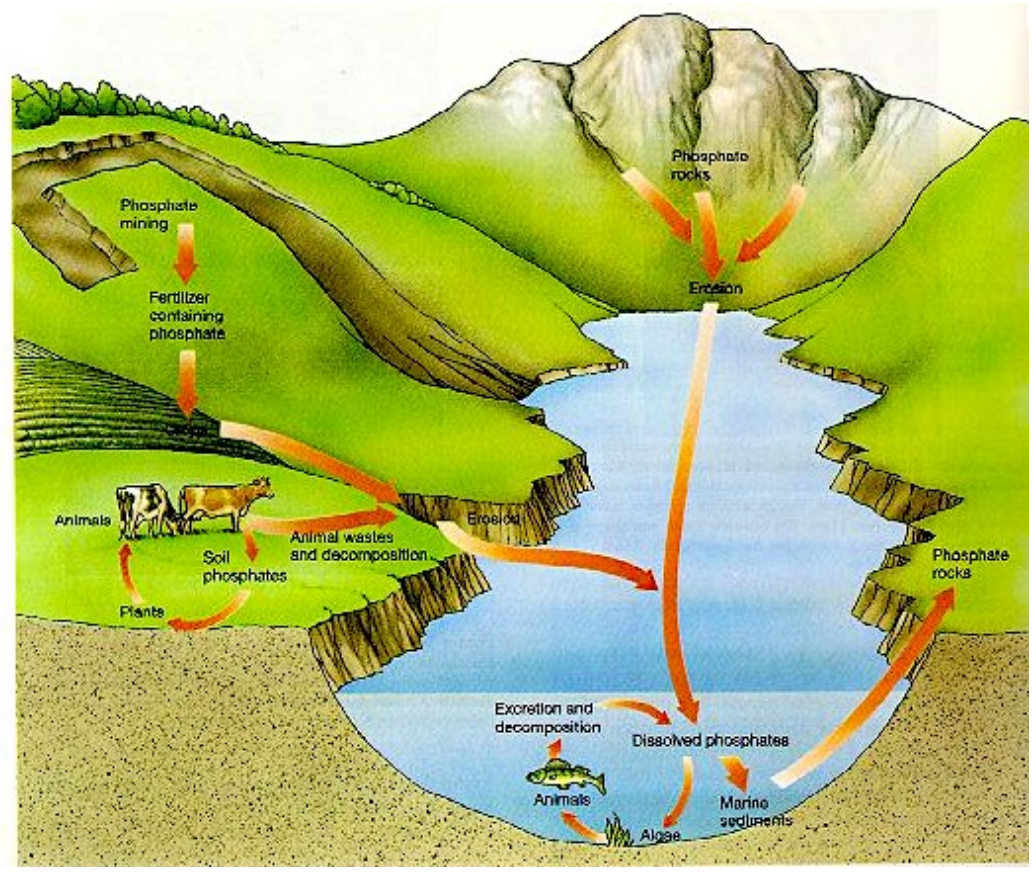


Figure 9: Elements of the phosphorus cycle

Phosphorus is probably the most studied plant nutrient in freshwater aquatic sciences. It is often found to be (and more often inferred as) the nutrient that limits the growth and biomass of algae in lakes and reservoirs. Whether this nutrient is as universally limiting as once believed is debatable, but certainly there is substantial evidence of its importance in many lakes. Numerous correlations and regressions have been constructed linking phosphorus, especially total phosphorus, with variables such as algal chlorophyll, algal weight, and productivity. Because of its possible importance in limiting the growth and biomass of algae and because of the numerous empirical models available for phosphorus, it is an important addition to the list of variables to be measured in water quality monitoring programs. There are two primary types of phosphorus: sediment attached (termed particulate phosphorus) and bio-available (termed soluble reactive phosphorus).

Most water quality monitoring programmes will include monitoring of total phosphorus (TP) to gauge the overall nutrient status of the water body and soluble reactive phosphorus (SRP) assess the amount of phosphorus that is readily available for biological uptake. Total phosphorus minus soluble reactive phosphorus = total particulate phosphorus. Soluble and particulate phosphorus are differentiated by whether or not they pass through a 0.45 micron membrane filter. A discussion of these contaminants follows.

Total Phosphorus

Total phosphorus measures the total amount (particulate and dissolved fractions) of phosphorus in the water column. Both point and non point sources of phosphorus are important contributors of

total phosphorus. Point-source phosphorus comes mainly from municipal and industrial discharges to surface waters. Non point-source phosphorus comes from runoff from urban areas, construction sites, agricultural lands, manure transported in runoff from feedlots and agricultural fields, and human waste from noncompliant septic systems. TP is an important variable for estimating the trophic level of a lake using the trophic level index (see page 39).

Soluble Reactive Phosphorus

Soluble reactive phosphorus (SRP) measures the soluble component of phosphorus in the water column that is readily available for uptake by plants and algae. High concentrations of SRP can result in algal blooms. This phosphorus fraction should consist largely of the inorganic orthophosphate (PO_4) form of phosphorus.

The Nitrogen Cycle

Nitrogen is essential to all life and is an important component of proteins, genetic material, chlorophyll and many other organic molecules. In fact nitrogen is the fourth most common element in living tissues. Plants take up nitrogen as simple inorganic compounds (ammonia and nitrate), while animals secure their nitrogen compounds from plants they have fed on.

The major reservoir for nitrogen on earth is the atmosphere making up 78% of the air we breathe as nitrogen gas (N_2). From the atmosphere nitrogen enters biological systems where it is recycled through numerous pathways. The N_2 molecule is held together by strong chemical bonds and only a few specialised bacteria and cyanobacteria (blue green algae) are capable of fixing N to usable inorganic nitrogen (ammonia). A number of higher plants e.g. legumes house these nitrogen fixing bacteria within their roots so they can use the nitrogen. The only other natural mechanism that convert N to inorganic nitrogen is lightning. In addition to this the industrial fertiliser manufacturing process can convert the N_2 molecule (under intense heat and pressure) into ammonia.

Nitrogen is transported to waterways from sources such as diffuse catchment runoff, urban and industrial drainage, groundwater recharge and nitrogen fixation by cyanobacteria (blue green algae).

Diffuse sources of nitrogen in catchments include animal wastes, decomposing plant and animal matter and agricultural fertilisers. Urban drainage including treated sewage and stormwater runoff which contains animal wastes and fertilisers leached from lawns, gardens and public open spaces. Dissolved nitrogen can also leach into groundwater which slowly seeps into waterways over years (a process called recharge).

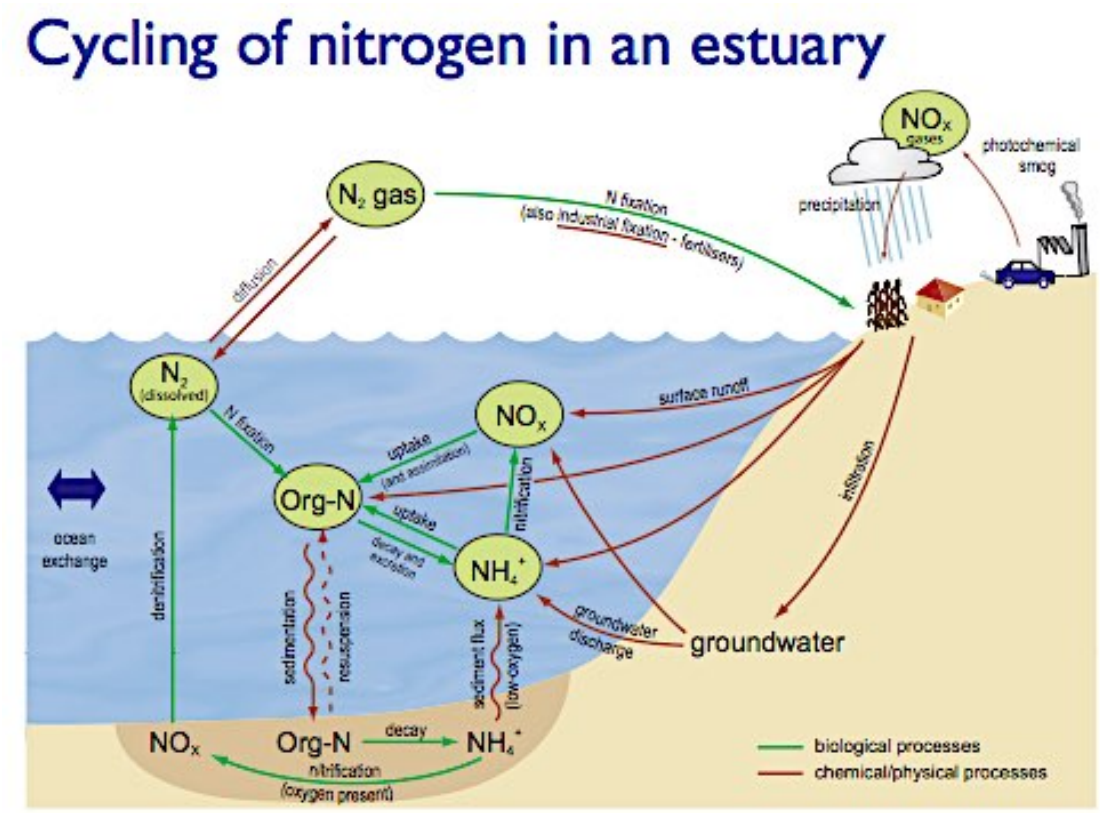


Figure 10: The nitrogen cycle

Once nitrogen has entered the water column either as nitrate or ammonium it can readily be taken up by plants which then convert it to organic nitrogen. As the plants die their decomposing bodies may settle on the bed of the water body whereupon the organic nitrogen is converted to ammonium which can then either be directly re absorbed by plants or alternatively converted to nitrate by a process called nitrification (using nitrifying bacteria). Denitrifying bacteria will convert nitrogen oxides (nitrites, nitrous oxide, or nitrates) back into nitrogen gas which may then be released back into the atmosphere.

Most water quality monitoring programmes will include monitoring of total nitrogen (TN) to gauge the overall nutrient status of the water body and dissolved fractions of nitrogen (total ammonia + nitrate nitrogen = dissolved inorganic nitrogen) to assess the amount of nitrogen that is readily available for biological uptake. A discussion of the components of total nitrogen follows.

Ammonia

Ammonia (NH3+) is a colorless gas with a strong pungent odor. Ammonia will react with water to form a weak base. The term ammonia refers to two chemical species which are in equilibrium in water (NH_3, unionized and NH_4^+, ionized). Tests for ammonia usually measure total ammonia (NH_3 plus NH_4^+). The toxicity of ammonia is primarily attributable to the unionized form (NH_3), as opposed to the ionized form (NH_4^+). In general, more NH_3 and greater toxicity exists at higher pH and lower water temperature.

When dissolved in water, normal ammonia (NH_3) reacts to form an ionized species called ammonium (NH_4^+)

$$NH_3 + H_2O \leftrightarrow NH_4+ + OH^-$$

This is a shorthand way of saying that one molecule of ammonia reacts with one molecule of water to form one ammonium ion and a hydroxyl ion. From the doubled headed arrow we can tell that the reaction can go either way and hydroxyl ions and ammonium ions could combine to form ammonia and water. This is precisely what happens as the pH of water increases; that is the water becomes more alkaline. You may recall that alkalinity is caused by an increase in hydroxyl ions. An increase in hydroxyl ions (or alkalinity) pushes the equilibrium to the left and more unionized ammonia is formed.

Unioinised ammonia is highly toxic to aquatic life if present in great enough concentrations. The Australia and New Zealand Environment Conservation Council (ANZECC, 2000) water quality guidelines provide total ammonia thresholds above which deaths of aquatic animals are likely to occur.

Nitrate and nitrite nitrogen

The term "nitrate nitrogen" refers to the nitrogen present which is associated with the nitrate ion ($NO3^-$). This nomenclature is used to differentiate nitrate nitrogen from other forms (e.g. ammonia nitrogen or nitrite nitrogen), etc. The concentrations are usually expressed in parts per million (mg/l or g/m^3) of nitrogen.

Nitrate reactions to nitrite in fresh water can cause oxygen depletion, this is because nitrites in the water column are oxygen scavenging. Thus, aquatic organisms depending on the supply of oxygen in the stream may die. If ingested in high enough concentrations nitrites react directly with methemoglobin which destroys the ability of red blood cells to carry oxygen. This can result in the death of the animal. In fish the condition is called “brown blood disease” while in infants the condition is called “blue baby syndrome”. Water with nitrite concentrations exceeding 1.0 mg/l should not be used for feeding infants. Recent research suggests that nitrate nitrogen concentrations above 1.0 mg/l can also have adverse effects on the success of trout egg survival (Hickey et al 2009).

Dissolved Inorganic Nitrogen

Inorganic nitrogen may exist in the free state as a gas N2, or as nitrate NO3-, nitrite NO2-, or ammonia NH3+. Dissolved inorganic nitrogen represents the combined inorganic nitrogen sources that are readily available for plant uptake to stimulate algae or plant growth. If concentrations of dissolved inorganic nitrogen become very high, algal growths may proliferate resulting in adverse effects on the aquatic environment.

Total Kjeldahl Nitrogen

Total Kjeldahl Nitrogen (TKN) represents the readily oxidizable organic fraction of nitrogen in the water column and is an important variable to measure particularly at waste water outfalls.

Measures of oxygen demand or organic loading

Waste water characteristically has a low oxygen concentration owing to the microbiological respiration occurring and / or chemical reactions occurring within it. When waste water enters

natural water it causes a decline in the oxygen concentration of the receiving water. Oxygen demand can be biological (termed biochemical oxygen demand) or chemical (termed chemical oxygen demand). The biochemical oxygen demand (BOD_5) test involves measuring the dissolved oxygen content of the water sample at day one and then incubating the sample for five days and measuring the dissolved oxygen content again. The difference in dissolved oxygen between the beginning and end of the incubation period is the biochemical oxygen demand.

For chemical oxygen demand, a strong oxidising agent is added to the water sample (potassium dichromate) which breaks down the organic matter in the sample. During this process the potassium dichromate gets reduced into $Cr3^{+}$. The amount of Cr^{3+} is determined after oxidization is complete, and is used as an indirect measure of the organic contents of the water sample.

In many environmental waters the concentration of biochemical oxygen or chemical oxygen demand is very low and is often below detection of the laboratory (called the detection limit). Therefore these variables are best measured in waste water only. Fortunately there is another test called total organic carbon (TOC) which can be analysed from environmental waters and it determines the amount of organic matter that is available for biological breakdown.

Total organic carbon is an important variable to measure as it also indicates the carbon that is available to bind to heavy metals in the water of which the dissolved fraction (termed dissolved organic carbon or DOC) determines the bioavailability. The alkalinity and pH of the water will also determine how much of the metal leaches from the sediment of the stream to become bioavailable.

Contaminants In The Urban Environment

Stormwater runoff in an urban or industrial environment characteristically contains four types of pollutants; suspended solids, nutrients, toxicants and general rubbish (Williamson 1993; Argue ad Pezzaniti 1997).

Stormwater runoff may also contain high concentrations of heavy metals, hydrocarbons and toxic chemicals such as organochlorines (Williamson 1993; Mackepeace et al; Snelder and Trueman 1995; Sharpin and Breen 1997). Of these contaminants, the heavy metals of lead, copper and zinc are commonly monitored.

Lead

Lead is highly toxic to aquatic life and it is the dissolved fraction (acid soluble lead) which can be readily bioavailable. Common anthropogenic (human derived)sources of lead include wheel weights from vehicles, vehicle wear and emissions, lead paint and lead heads on nails while catchment soils and rainfall also provide background sources.

Copper

Copper is also highly toxic to aquatic life, again with the dissolved fraction (acid soluble copper) which is readily bioavailable. Common anthropogenic sources of copper include brake pad wear, roof runoff, garden products (like fungicides), and copper plumbing.

Zinc

Zinc is also highly toxic to aquatic life with the dissolved fraction (acid soluble zinc) being readily bioavailable. Common anthropogenic sources of zinc include tyre wear, vehicle emissions, roof surfaces, paints and natural background sources in catchment soils.

All heavy metals are measured in parts per million (mg/l or g/m^3). Concentrations of lead have declined over the years owing to the change to lead free additives in our fuels however regular exceedances of environmental guidelines for zinc and copper are known to occur in urban and industrial streams, usually following small sized rainfall events which have been preceded by longer dry periods. This phenomenon is known as the 'first flush' effect as the wash off from roads is transported to the stormwater reticulation system the rainfall initially carries a very concentrated loading of metals owing to the accumulated metals having been deposited during the dry period. Real time monitoring (undertaken at short 5 minute intervals) of electrical conductivity and flow demonstrates the first flush phenomenon in the diagram below.

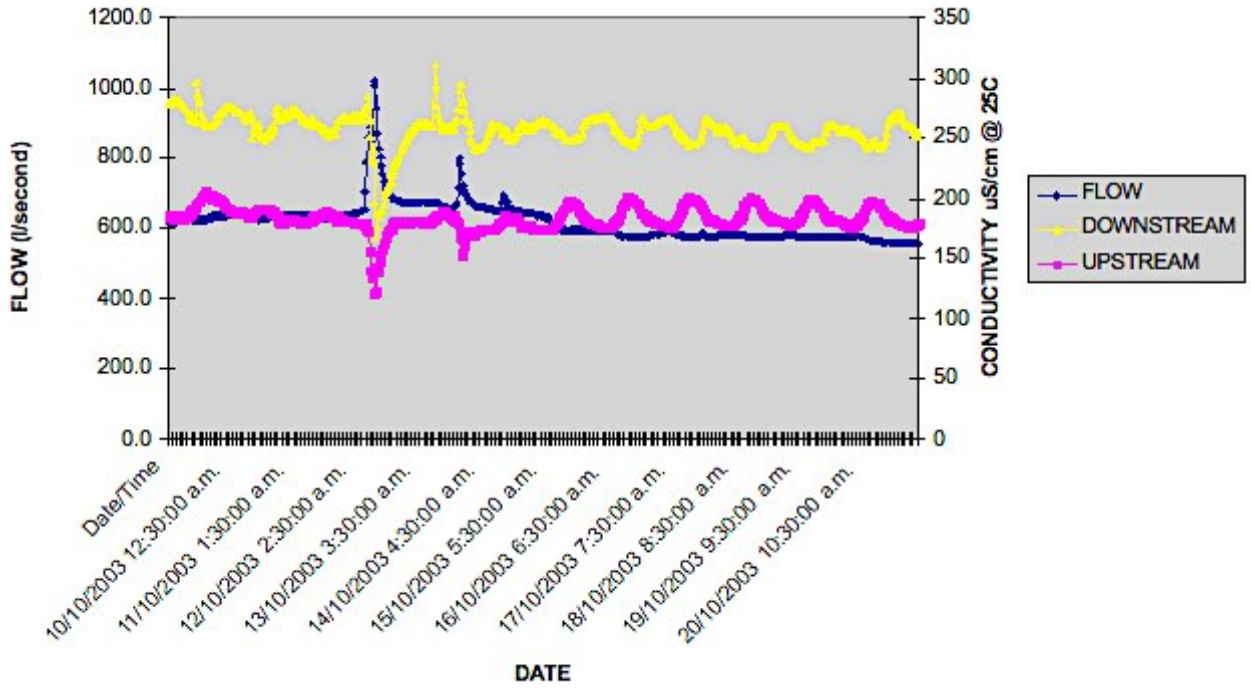


Figure 11: Results of real time monitoring

Often the stream water may be less toxic than the underlying sediments due to the sorption of metals from the water column to the sediments. It is therefore important to also monitor metal concentrations (mg/kg) in stream sediments. Heavy metals tend to bind to the smaller fractions of stream sediments (usually < 20 um) so if the stream is a cobble /gravel system you may have trouble getting enough sample in this size class. Normally the laboratory would like about half a kilo of sediment for analysis. Stream bed sediments provide an important sink for stormwater derived contaminants. You can sample the sediments using an open ended plastic tube and scooping the sediment up, decanting any water slowly out then placing the sediment in a plastic bag, alternatively you can use an ekman grab sampler which looks like a bunch of spring loaded jaws.

Be sure to focus on the upper layers of the stream bed as this is where the most recent contamination is likely to have settled. The laboratory will report back not only the metal concentrations but also the fraction sizes of particles within the sample, this enables you to get a grasp of how much of the stream bed sediment is most suitable for metal sorption. Note it may also pay to have total organic carbon to assist with comparing organic contaminant concentrations against sediment quality guidelines. The ANZECC sediment quality guidelines provide two thresholds, the ISQG- Low at which the onset of biological effects could occur and ISQG-high at which biological effects are expected.

Polycylic Aromatic Hydtrocarbons

Polycyclic aromatic hydrocarbons (or PAHs) are common atmospheric pollutants of the urban environment and are present in fossil fuels or are released into the environment as a result of incomplete combustion of wood, oil products, coal, fat, or tobacco.

Polycyclic aromatic hydrocarbons are lipophilic meaning they mix more easily with oil than water. The larger compounds are less water-soluble and less volatile (i.e., less prone to evaporate). Because of these properties, PAHs in the environment are found primarily in soil, sediment and oily substances, as opposed to in water or air. Stream sediments can contain PAHs that may have been deposited by the erosion of contaminated soil or sediments. They will have a petrochemical smell to them which is where the aromatic word comes in. They are certainly worth monitoring in heavily urbanised or industrial stream catchments.

Monitoring Water Microbiology

Microorganisms are present everywhere in our environment, in soil, air, food and water. Also called microbes, microorganisms are living organisms, generally observable only through a microscope. Our exposure to them causes harmless microbial flora to establish in our bodies, although some microbes are pathogens and can cause diseases. These diseases are considered waterborne if the pathogens are transmitted by water, to infect humans or animals that ingest the contaminated water. Diseases transmitted by water are primarily those found in the intestinal discharges of humans or animals The presence of microbial contaminants in drinking water has plagued humans throughout history.

Water borne disease ingestion can be caused by

- direct ingestion e.g. drinking untreated water

- accidental ingestion - caused by swimming or perhaps water skiing

- inhalation of aerosols - yes we can get sick by breathing in tiny droplets of contaminated water

- working in high risk areas e.g. sampling effluent from a municipal sewage discharge - occupational health and safety requirements of sampling help minimise this risk

For bathing beach water quality water samples are collected at popular bathing sites at mid calf depth. The reason we take samples at this depth is because babies and toddlers who are likely to be most vulnerable to contaminated water are usually paddling at these depths. Regional Councils advertise the results of microbiological water quality at their website and the local newspaper to alert the public if any bathing areas are unsafe for bathing.

Stormwater runoff usually has a very high bacteria concentration owing to the faeces that may get washed down the drain from dog droppings and many cities still have combined sewer/stormwater overflow systems. What this means is that during dry or small rainfall events, both the stormwater and sewage go to a municipal treatment system for treatment, however during high rainfall events, the water received by the pipe network exceeds its capacity so the overflow (a mixture of stormwater and sewage) will flow directly to a river or the sea (depending on where the urban settlement is located) untreated.

Rural runoff is also very high in bacteria concentrations owing to the faeces of domestic stock (cows, sheep, goats, deer) being washed down the catchment by rainfall which then enters rivers which then discharge to sea.

Water borne disease organisms are mainly bacterial (Campylobacteria, Salmonella,Shigella, Yersinia and toxigenic E.coli), protozoal (Cryptosporidia and Giardia) or viral (enterovirus or norovirus). If someone is exposed to these microbes and accidentally ingests them they can expect a severe case of gastroenteritis. It should be noted here that most gastroenteritis is caused by ingesting poor quality food or from contact with animals, a much smaller fraction is caused by playing or swimming in water.

Measuring water microbes directly is very expensive and time consuming, to get around this problem regional councils and public health authorities measure indicator bacteria in the water i.e. bacteria that are commonly associated with other water microbes of public health concern.

Once a water sample has been taken, it is immediately stored in a chilli bin (or eski) with ice and transported to a laboratory. The laboratory then passes the water through a filter. The filter paper is then transferred to a nutrient agar medium (kind of like jelly) which feeds the bacteria and they are placed in an incubator for 24 hours. During this time the bacteria form colonies (the correct term is colony forming units) which show up as dots on the filter paper. The filter paper is marked with grid lines. The laboratory analysis then counts the number of dots on the filter paper. So supposing 500 dots are counted on the filter paper and 200 ml of water had passed through the filter then the resulting concentrations of indicator bacteria is 250 cfu / 100 ml.

For fresh waters the preferred indicator bacteria for monitoring bathing safety is Escherichia coli (*E.coli*). E.coli is commonly found in the lower intestine of warm blooded animals and has shown to have a good correlation with other microbes in fresh water. Faecal coliform bacteria are also common in the intestines of mammals and often measured for shellfish harvesting or stock water.

For estuarine or salt water environments, Enterococci bacteria also commonly found in the gut of mammals are monitored. Both E.coli and enterococci are expressed in concentrations of cfu/ 100ml.

Detecting the source of contamination

Looking for the source of bacteriological contamination in a catchment could be likened to looking for a needle in a hay stack. The reason I say this is that microbes can only be seen with a microscope, therefore it is difficult to know exactly where the source of contamination could be

coming from. Fortunately there are some tools we can use to help us gain some understanding of source areas and types of contamination.

The first question people may ask is “is the source of contamination human or animal derived?”. If we know the answer to this question then we are in a position to provide appropriate catchment management improvements, e.g. if the source of contamination is human derived then perhaps there are leaking pipes in the sewage reticulation network or perhaps there are failing on site waste water treatment systems operating in the area, alternatively if the source of contamination is predominantly animal derived then the appropriate action might involve fencing stock from water ways, providing overhead bridges for stock crossings or providing wetland buffer areas strategically located within the catchment.

Human derived waste water usually contains complex organic compounds called fluorescent whitening agents (FWA) contained in washing powders. The FWAs adhere to woven fabrics and make clothing appear brighter and cleaner, however a lot of the FWAs also do not adhere to the clothing and end up in grey water (water discharged from the washing machine) which then gets discharged to the waste water treatment facility. The presence of FWAs in a natural water body (stream or beach area) would indicate that human derived waste water is present.

Fortunately there is a method that can finger print whether waste water is predominantly derived from human or animals. The chemical method is called faecal sterol analysis. Because humans and other animals have different concentrations of steroids in their faeces, the faecal sterol test is able to show us whether the contamination of the water is likely to be human or animal derived.

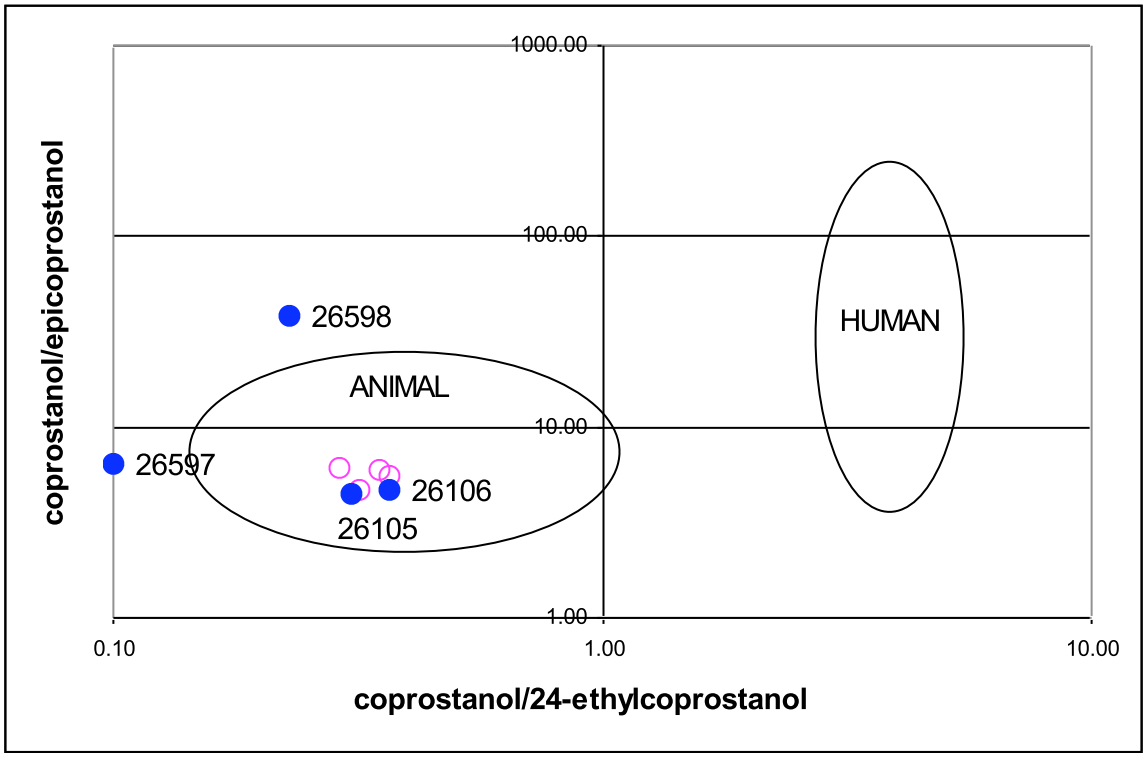


Figure 12: results of a faecal sterol analysis undertaken in the Puhokio Stream Catchment of Hawke’s Bay

Because enterococci and E.coli have different survival rates (they are generally killed off by sunlight) the ratio of Enterococci / E.coli in a water course can help us understand how close we might be to the source of contamination. E.coli bacteria generally survive in water for shorter periods compared to enterococci (Donnison 1998), therefore if we took samples down a catchment and found one of the sites had a higher Enterococci/ E.coli ratio, this site is likely to be closer to the source of contamination.

Tracer dye tests are sometimes on wastewater treatment systems if the a water course nearby is suspected to be contaminated by failing treatment systems. Common dyes include fluroscene (a yellow colour that becomes brighter with sunlight) and potassium permanganate (a red colour). Mixed success has resulted from these tracer dye tests because it is often uncertain how long a pollution plume may take to reach the natural receiving water.

Monitoring Stream or Lake Ecology

Freshwater ecosystems contain many species of invertebrates (animals lacking a back bone e.g. snails, crustacea, insect larvae, worms etc) and vertebrates (fish), many of which have known pollution tolerances. By sampling the biology of the water course we are often able to gauge the overall stream health of that system. At this point it is pertinent to point out the levels of biological organisation at which stress may be experienced.

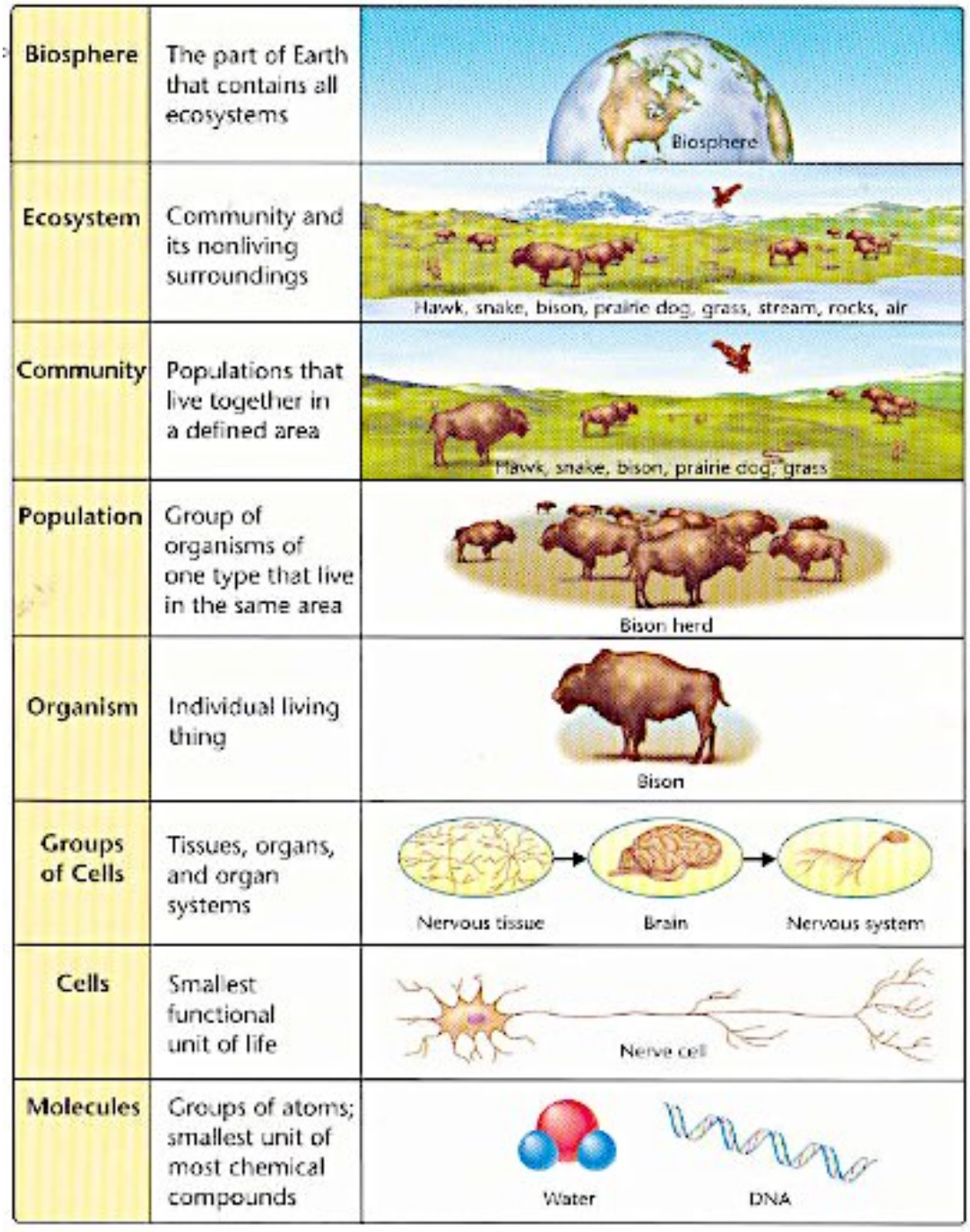


Figure 13: Levels of biological organisation. Source: http://mwsu-bio101.ning.com/profiles/blogs/2263214:BlogPost:2543

From the above diagram we can see that there are many levels of biological organisation at which stress can be felt. When we are monitoring aquatic ecosystems it is important to keep this in mind. Often we do not have the technology to understand what stress may be felt by a plant or animal at the cellular or cell organelle level (single cell organisms are probably an exception to this).

For fish we are able to determine any stress of the organs by autopsy, however this is probably not possible for smaller species of aquatic animals like macroinvertebrates as they can be very small.

For the remainder of this section I will examine the effects of eutrophication (increasing nutrient loading) and sedimentation on a water course as the effects of other toxicants (e.g. heavy metals, pesticides, herbicides, poly aromatic hydrocarbons, estrogens, etc) and their synergistic effects (combined effects) are less well understood.

We now know that if a water course becomes polluted with excess nutrients, that algal blooms can occur. In streams this results in the stream bed becoming covered in green slime (called periphyton) or in lakes phytoplankton blooms may occur. In both instances productivity is being stimulated at the base of the food chain (plants) which will give rise to effects further up the food chain (herbivores and carnivores).

Chlorophyll a

Chlorophyll is the green molecule in plant cells that carries out the bulk of energy fixation in the process of photosynthesis. Besides its importance in photosynthesis, chlorophyll is probably the most-often used estimator of algal biomass in lakes and streams. Chlorophyll is also relatively easy to measure.

Chlorophyll itself is actually not a single molecule but a family of related molecules, designated chlorophyll a, b, c, and d. Chlorophyll a is the molecule found in all plant cells and therefore its concentration is what is reported during chlorophyll analysis. The amount of plant matter growing in a water body (a measure of the productivity of the ecosystem) can be measured by sampling for chlorophyll a.

For rivers and streams, we often sample a fixed area of the stream bed (a bottle cap area on a stone) and send this sample away for chlorophyll a analysis. In lakes we filter a measured volume of water then send the filtrate (filter paper plus trapped phytoplankton) to the lab.

The laboratory procedure involves boiling the sample in acetone to extract the chlorophyll content, allowing it to cool and then reading the absorbance (light absorbance value) using a spectrophotometer. For the stream or river samples the units are in mg chlorophyll per metre squared of stream bed while in lakes the units are mg chlorophyll a per milliliter of water.

Macroinvertebrates

Macroinvertebrates comprise insect larvae, molluscs (snails and bivalves), crustaceans and soft bodied animals such as oligochaetes and planarians (flat worms). They are an important component of the ecosystem to monitor for as they

•are an important food source for fish and wading birds

•they are easy to collect due to them being prevalent in most water and being highly abundant

•identification is relatively easy provided you have the appropriate keys

•they have a high hysteresis value due to their sedentary habits and relatively long life cycles. They therefore represent how the water quality has been over the past 6 months - 2 years

•they have a graded response to a broad spectrum of pollutants (this is due to the diverse phyla within the group).

They are termed macro because the animals examined are larger than 0.5 mm body length which is the mesh size recommended for sampling them. The macroinvertebrates can be sampled using a kick net or a surber sampler. The kick net procedure involves placing a D shape net on the stream bed and kicking the stones immediately upstream of the net. The aquatic macroinvertebrates become dislodged from the stones and wash into the net. The surber sampler uses a similar procedure only the area of stream bed to be disturbed is defined by an aluminium frame which folds out in front of the net. Because the surber sampler is sampling in a defined area of stream bed it provides a more accurate estimate of macroinvertebrate densities.

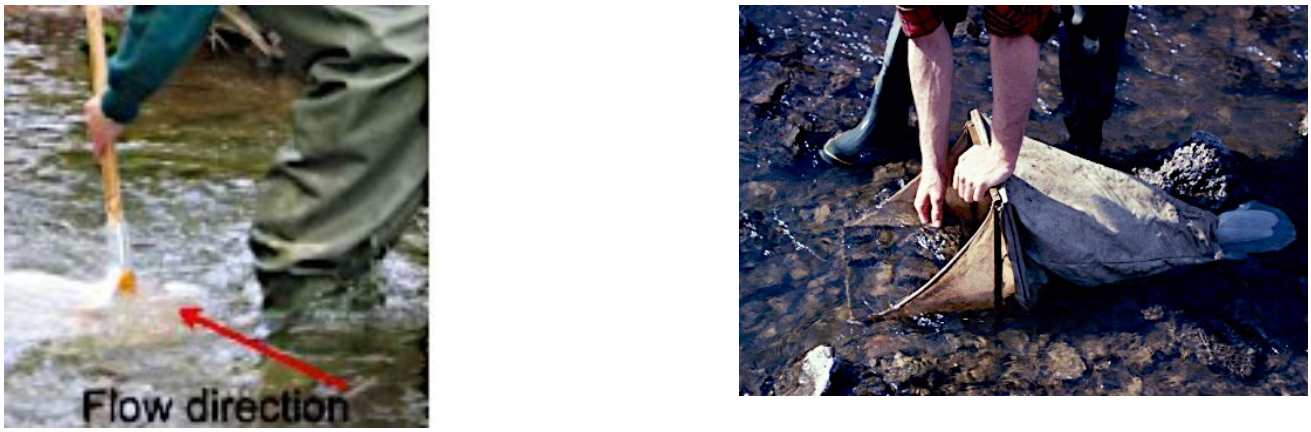


Figure 14: Kick net sampling (left) and surber sampling (right) aquatic macroinvertebrates

Once the macroinvertebrates have been sampled, they are transferred to a plastic pottle and preserved in either ethanol or isopropyl. This kills them and preserves them for microscopic analysis. The taxonomist (person who goes through the sample in the lab) will identify each animal to a minimum standard of taxonomy (mostly to genus) and this data can then be used in the formulation of a biotic index of water quality called the macroinvertebrate community index (see water quality index section).

Fish

Fish are also an important component of the aquatic ecosystem to measure. They are generally less sensitive to pollution than macroinvertebrates as they have a greater capacity to swim away from pollution, however their presence or absence in a stream can be an important indicator of habitat or fishery value.

Fish are usually monitored using an electro fishing machine. As the name implies, the machine has a wand which passes an electrical current into the water column. This direct current (DC) electrical pulse induces what is called forced swimming of the longtitudinal (lengthways) muscles of the fish towards the wand. The fish become stunned as a result of the electric current in the water and they are then quickly swept up in a sweep net. The fish will be identified, counted and their length recorded before they are returned to the water. The stunning of the fish usually occurs for a matter of seconds after which they recover back to normal health. There are sometimes a few casualties for those fish that may have touched the wand however this normally equates to a very low percentage of individuals.

The native fishery value of the water course can be evaluated based on the number and diversity of fish present. The most common index used in New Zealand is termed the index of biotic integrity (see water quality indices section).

Undertaking a Stream Gauging

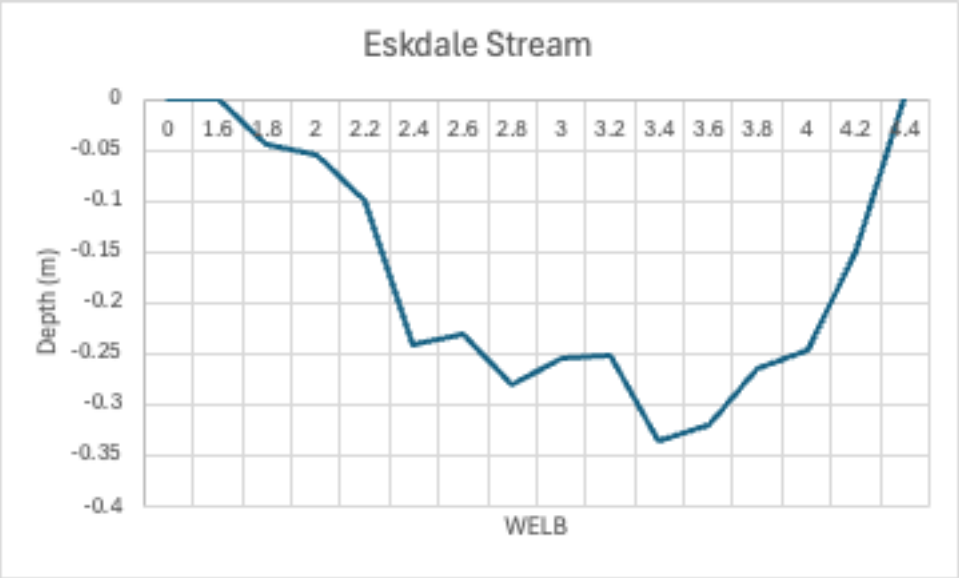
A stream gauging adds enormous value to any stream monitoring you are undertaking. It enables you to gain an understanding of :

- Available habitat for fish and macroinvertebrates particularly if depth, velocity and substrate are recorded
- Water velocities within the stream reach - this could explain why slow swimming fish may be absent above a particular point
- The comparative size of the river or stream in L/s or m^3/s
- If combined with water quality sampling it enables you to determine the contaminant loading and yield of the catchment (see next section).
- It enables you to flow adjust your data for time series analysis - contaminant concentrations are usually flow dependent

The procedure is reasonably straight forward. First of all you measure the wetted width of the stream. The protocol is from the true left bank. You then measure the depth of the water at fixed interval distances from the left bank. Ideally you should have at least 12 depth measurements. Like all things more is better , less is not so good. Once you have measured the depths its then time to get out your velocity meter and measure the velocity at 0.6 of the depth at each transect point. The reason 0.6 depth is advocated is because it is where average velocities are experienced in the water column. The surface water is very fast compared to the bottom due to friction occurring, so overall the best measurement is taken at 0.6. depth.

Once your velocity measurements have been made for a quick field discharge estimate you simply calculate the average and multiply this by the cross sectional area of the water. This is calculated by multiplying the wetted width by the mean depth.

Alternatively you could calculate the area of each individual trapezoidal cell so the first is from the waters edge left bank to the first point of the transect distance multiplied by the depth and divided by 2. The reason we divide by 2 is because the stream edge is like a triangle. The remaining cell areas are calculated as trapezoids until you get to the waters edge right bank where the area is again divided by 2. In general the trapezoidal area technique is the better method as it is considered more accurate for irregular shaped stream beds. A schematic and data is provided overleaf for the Eskdale Stream.



WELB	Depth	Area m2	Velocity (m/s)	Flow m3/s	11:00 AM
0	0				18/5/2022
1.6	0				
1.8	-0.045	-0.0045	0.1	-0.00045	
2	-0.055	-0.011	0.1	-0.0011	
2.2	-0.1	-0.02	0.1	-0.002	
2.4	-0.24	-0.048	0.05	-0.0024	
2.6	-0.23	-0.046	0.1	-0.0046	
2.8	-0.28	-0.056	0.05	-0.0028	
3	-0.255	-0.051	0.05	-0.00255	
3.2	-0.25	-0.05	0.05	-0.0025	
3.4	-0.335	-0.067	0.05	-0.00335	
3.6	-0.32	-0.064	0.05	-0.0032	
3.8	-0.265	-0.053	0.05	-0.00265	
4	-0.245	-0.049	0.05	-0.00245	
4.2	-0.15	-0.03	0.05	-0.0015	
4.4	0	0			
			Sum	-0.03155	
Wetted width	Average Depth (m)	Area (m2)	Average Velocity (m/s)		
2.8	-0.1846667	-0.5170667	0.06538462	-0.0338082	

Sometimes you may have a perched culvert immediately upstream of your gauging site. While this is bad for fish passage migrations it can be a godsend for undertaking a bucket gauging.

So the method is very straight forward. One person has the stop watch the other has the bucket. The bucket is timed for the period that flow exits the culvert and fills the bucket. The volume of water is then measured using volumetric flasks or containers and the flow = volume/ time = L/s.

An alternative method to measure the volume is to record the depth in the bucket and use

π r^2 * bucket depth

Be cautious with this because the bucket must have the same diameter at the base as the mouth of the bucket. If you're sure the bucket has straight sides then no problem. You could also try simply weighing the bucket empty then weighing the bucket with the water in it. 1 litre = 1 kg.

Estimating Pollution Loading, Yield and Final Concentrations

In the previous chapter we discussed various water quality variables that can be measured from a water body. If we have sampled a measured concentration of a contaminant from a river or stream and we measured flow at the time of sampling, we are in a position to estimate the pollution load which is essentially the concentration of the contaminant multiplied by the flow.

Pollution load (mg/second) = concentration (mg/litre) * flow (litres/second)

The pollution load can give us an estimation of the amount of pollution leaving a catchment over a given period of time. Some caution must be exercised in making this calculation because most of the contaminant is likely to leave the catchment during high flows when a lot of catchment runoff is occurring. If we only sampled water during low flow periods, our estimation of pollution leaving the catchment would be an under estimation.

To give you an example of an instantaneous loading, supposing we sampled water at the Hutt River and it was found that the soluble reactive phosphorus concentration was 0.010 mg/l and the flow at the time was 1500 l/ second. Then the pollution loading at that time was:

Pollution load = 0.010*1500 = 15 mg/second

If we were to base this one survey to calculate the annual loading we could multiply up the number of seconds to make a year which is

15 *60*60*24*365 = 473040000 mg pollution load in one year. Because a mg (1/1000 of a gram) is a small unit of measurement, we could divide this result by 1 million to give 473.04 kg/year. As I indicated earlier, calculating a loading based on one survey is likely to be a bad estimate as the flows vary throughout the year.

Ideally the pollution load should be estimated at a site in which real time data (telemetry) for flow is available. It is unlikely that you will have measured the contaminant concentration every second of the day therefore there are three methods you can use to determine the annual loading for a catchment they are averaging, interpolation and correlation.

Averaging is where you simply multiply the average flow for a month by the average concentration of the contaminant in a month and add these monthly loadings to get the annual loading estimate. Alternatively interpolation involves drawing a line between each monthly loading estimate and calculating the daily loading between each point. Finally the correlation method involves firstly plotting the contaminant vs flow relationship e.g. nitrate nitrogen on the y axis, flow on the x axis. From this, a regression equation can be calculated which describes the relationship of flow with the contaminant in the form of y (concentration of contaminant) = e.g. 0.00012* flow. Then it is just a matter of multiplying the flow at each instant by this scaling factor to give the instantaneous concentration. The resulting instantaneous concentrations are then multiplied by the real time flow values to give instantaneous loadings which are then added up to give the annual loading estimate.

I have tried all three methods of estimating loadings and the results are reasonably similar.

While calculating the pollution loads coming from various catchments is important information, what's also important is understanding the catchment yield. This is where we divide the loading estimate by the catchment area e.g. supposing that at the point that we sampled the Hutt River, it had a catchment area of 500 km^2 (50,000 hectares) then the instantaneous yield for the catchment would be :

473.04/50000 = 0.00946 kg/ha or to make it more readable, 9.46 g/ha

Catchment yield estimations are particularly important for identifying catchments that are most polluting. To give you an example, supposing we compared the loadings of two rivers that had similar nitrate nitrogen concentrations (Hutt River 0.250 mg/l, Waipoua River 0.228 mg/l). First we'd examine the loadings for which the Hutt Rivers flow was 1500 l/s, where as the Waipoua River flow was 200 l/s.

Hutt River Loading = 0.25*1500 = 375 mg/second*60*60*24*365/1000000 = 11826 kg/yr

Waipoua River Loading = 0.228*200 = 56 mg/second*60*60*24*365/1000000=1766.02 kg/yr

From these instantaneous loadings we see that the Hutt River is delivering more pollution but now let's examine the catchment yields for which at the point of sampling the Hutt River catchment area was 50000 Ha whereas the Waipoua River catchment area was 7000 Ha

Hutt River Yield = 11826/50000 = 0.236 kg/ha/yr

Waipoua River Yield= 1766.02/7000 = 0.252 kg/ha/yr

So in the above example although the pollution load coming from the Hutt River was over 10 times higher than the Waipoua River, the yield of the Hutt Catchment was 6.3% lower than the yield of the Waipoua Catchment. It could be that the Waipoua River is being more intensively farmed which is why it has been estimated as having a higher yield.

Loadings and yields provide important information to the scientist particularly if needing to decide where best to place catchment restoration initiatives to improve water quality.

Often we need to estimate the final concentration in a river after it has received a discharge of wastewater. If the concentration of the contaminant and the flow is known for the discharge pipe and the receiving water, then the final concentration of the contaminant can be estimated using the following formula:

Final concentration = $C_1V_1 + C_2V_2 / (V_1+V_2)$

For example supposing a meat works factory was wanting a consent to discharge treated offal waste into a river for which good flow statistics and water quality was known. Let's call the river the Wairoa River near the township of Wairoa. The flow statistics for this river might be mean annual low flow (the average low flow event in a given year) = 2000 l/second and the mean nitrate nitrogen concentration for this river during low flow events was 0.7 mg/l and the consent holder wanted to discharge nitrate nitrogen with a maximum permitted discharge of 100 mg/l nitrate nitrogen at 20 l/second, then the final concentration is

0.7*2000 + 100*20 / 2020 = 1.68 mg/l nitrate nitrogen

Clearly this would be unacceptable and the consent applicant would need to treat the effluent to a better standard. Supposing some treatment facilities were brought in which got the quality of the effluent down to 10 mg/l, then the final concentration of nitrate nitrogen in the river would be

0.7*2000 + 10*20 / 2020 = 0.79 mg/l nitrate nitrogen

When making these estimations, I always estimate the worst case scenario, i.e. I would probably use a 1 in 5 year mean annual low flow combined with the maximum permitted discharge at the highest concentration that the consent holder is seeking. I have seen consulting reports previously where they use mean annual flows (the average flow of the receiving water) which does not represent worst case scenario. Something to be mindful of.

Water Quality Indices

From the previous chapter you would agree that there are many water quality variables to measure in our freshwater environments (streams rivers, lakes and wetlands). In many instances it is necessary to summarise the water quality information using indices. This would make the reporting less complex and easier to understand by water managers who do not have a strong technical understanding of water quality science. In this chapter I will outline some indices that are commonly used in rivers, lakes and wetlands.

Macroinvertebrate Water Quality Indices

When macroinvertebrates are sampled from a river or stream, the data generated can be used to formulate indices that assess ecosystem health using the number and abundance of taxa (identified groupings, e.g. genus, family, order) identified within each sample.

The macroinvertebrate community index is probably the most commonly used biotic index used in New Zealand. It was initially labelled as an index of organic enrichment, however there is nothing wrong with using this index for assessing impacts of sedimentation on a stream also. Each taxonomic group (taxon) identified in the sample is given a biotic score based on what is known about the sensitivity of those animals to pollution (e.g. a biotic score of 10 = highly sensitive, a biotic score of 1 = insensitive). The biotic scores of all the taxa are then added up and divided by the number of taxa and multiplied by a factor of 20 to give the macroinvertebrate community index (MCI).

As an example supposing I found 5 taxa at a stream site for which the biotic scores were:

Taxa Biotic Score

Deleatidium may fly = 8

Coloburiscus mayfly = 9

Potamopyrgus snail = 4

Oligochaeta worm = 1

Chironomus midge =1

Total 23/5 taxa * 20 = MCI of 92

The author of the MCI Dr. John Stark categorised MCI values for the Taranaki Ring plain streams as follows (Stark 1985):

Quality Class	Stark (1988) Descriptors	Threshold
Excellent	Clean Water	MCI >120
Good	Possible Mild Pollution	MCI 100-120
Fair	Probable moderate pollution	MCI 80-100
Poor	Probable severe pollution	MCI <80

Table 1: MCI threshold values and descriptors

So in the example above we could classify the water as moderately polluted. The MCI has proven to be a robust measure of ecosystem health and has been recommended as the standard state of the environment reporting tool for regional councils.

The quantitative analogue of the MCI is called the quantitative macroinvertebrate community index. This is where the quantity of individuals in each taxon is incorporated into the formulation of the index. So taking the previous example, supposing the abundances of each taxon were as follows:

Taxa	Biotic Score	Abundance	Abundance loaded score
Deleatidium may fly = 8	*	20	160
Coloburiscus mayfly = 9	*	9	81
Potamopyrgus snail = 4	*	10	40
Oligochaeta worm = 1	*	2	2
Chironomus midge =1	*	11	11
Total number of individuals		52	
Total Loaded score			294
QMCI			294/52 =5.65

Because the QMCI incorporates the number of individuals in each taxa, it is considered a more sensitive index than the MCI. Surber sampling is recommended for this index as the number of individuals could be biased by the area of stream bed being sampled. The QMCI is recommended for specific consent or compliance monitoring in which subtle changes in community composition need to be assessed. Stark (1988) provides the following descriptors for the QMCI

Quality Class	Stark (1988) Descriptors	Threshold
Excellent	Clean Water	QMCI >6
Good	Possible Mild Pollution	QMCI 5-6
Fair	Probable moderate pollution	QMCI 4-5
Poor	Probable severe pollution	QMCI < 4

Table 2: QMCI thresholds and descriptors

So in the above example the water is classified as mildly polluted. Note the water grading has changed and this is because we are using more information. From the data you can see that there are more pollution sensitive animals in the sample (29 may flies vs 23 lower scoring taxa) which has resulted in the index classing the water as less polluted.

Another biological index commonly reported on is the EPT index (Lenat 1988). This index calculates the number of taxa belonging to the orders ephemeroptera (may flies), plecoptera (stone flies) and trichoptera (caddis flies) so in the above index the EPT value = 2 because there are 2 taxa of mayflies. The index is also sometimes reported on in terms of the percentage of individuals that belong to those orders. So in the above example, there are 29 mayflies and there

are 52 individuals in total, therefore the % EPT = 29/52* 100 = 55.8. The information derived by the EPT indices are not recommended for state of the environment reporting rather they are recommended to be used for providing information to the stream ecologist reviewing the report or paper. These numbers give the ecologist a 'feel' for the composition (make up) of the aquatic macroinvertebrate community.

Normally the taxa richness of a sample is reported, in the above example there are 5 taxa. Taxa richness is not strictly a measure of stream health because slightly enriched streams tend to have a higher taxa richness than pristine streams. Furthermore low taxa richness can be associated with quite sterile environments with extremely pure water. However taxa richness goes some way to providing the reader with an indication of biodiversity value of the stream although this is probably better expressed in terms of how the biodiversity compares to a reference condition for that stream.

The community loss index does just that! It estimates the loss of taxa between comparison samples and reference samples. As an example supposing in the current example that we re sampled the stream after a known disturbance occurred.

Taxa Before	Taxa After the Disturbance
Deleatidium may fly	Deleatidium may fly
Coloburiscus mayfly	
Potamopyrgus snail	Potamopyrgus snail
Oligochaeta worm	Oligochaete worm
Chironomus midge	Chironomus midge
	Culex mosquito
	Physella snail

The formula for the CLI is

CLI = number of taxa in reference condition - number of taxa in common / number of taxa at impacted state

therefore in the above example the CLI= 5-4/6 = 0.16

Generally CLI values above 0.4 represent a major change in biodiversity value so in the above example we would indicate that the biodiversity of the system has not been impacted. The CLI has not been widely used in New Zealand and I think that is because it can be easily biased by the presence of rare taxa that are simply rare because they're rare! More replicate samples may be required to give the CLI greater precision. Replicates and precision are described in chapter 7.

The Index of Biotic Integrity

The index of biotic integrity is a tool used to assess river condition for fish based on fish monitoring and it was developed by Dr. Mike Joy of Massey University. The calculation of the index is quite complicated as it needs to address not only the number of native and exotic fish species present but also takes into consideration the distance inland (from the sea) and the altitude of the site. The

reason these latter details are considered is because approximately half of New Zealand's native fish species are diadromous meaning at some stage in their life cycle they are either in the rivers or at sea. These represent both anadromous and catadromous fish. Anadromous fishes spend most of their adult lives at sea, but return to fresh water to spawn. Catadromous is a term used for a special category of marine fishes who spend most of their adult lives in fresh water, but must return to the sea to spawn.

For fish to undertake these migrations to sea and from sea to freshwater, they must have adequate passage up and down the rivers that they migrate. Often these migratory pathways have been altered by man made structures such as hydro electricity dams, farm dams or elevated culverts that prevent upstream and or downstream migration. This impedes the fish's expected distribution which means that the river condition is likely to score lowly on the IBI scale. Fish assemblages recorded from the surveys are compared to a reference condition (a stream in which no barriers to fish exist, or exotic fish are present and the stream is unmodified).

The IBI scores: excellent (58 - 60), good (48 - 52), fair (40 - 44), poor (28 - 34), and very poor (0 - 22). Calculation of the IBI involves using an excel macro.

Other more generic tools to use for river condition include the New Zealand Freshwater Fish Database which can be used to see where fish monitoring surveys have been undertaken previously and a point fish clic tool that was developed for regional councils by Dr. Mike Joy. The point fish clic tool is a GIS (geographic information systems) layer that enables the viewer to see where fish surveys have been undertaken before and what species of fish have been found there. There is also an additional layer that gives the probability of finding a particular fish species on any reach of river in New Zealand. The following maps provide some outputs from point fish clic.

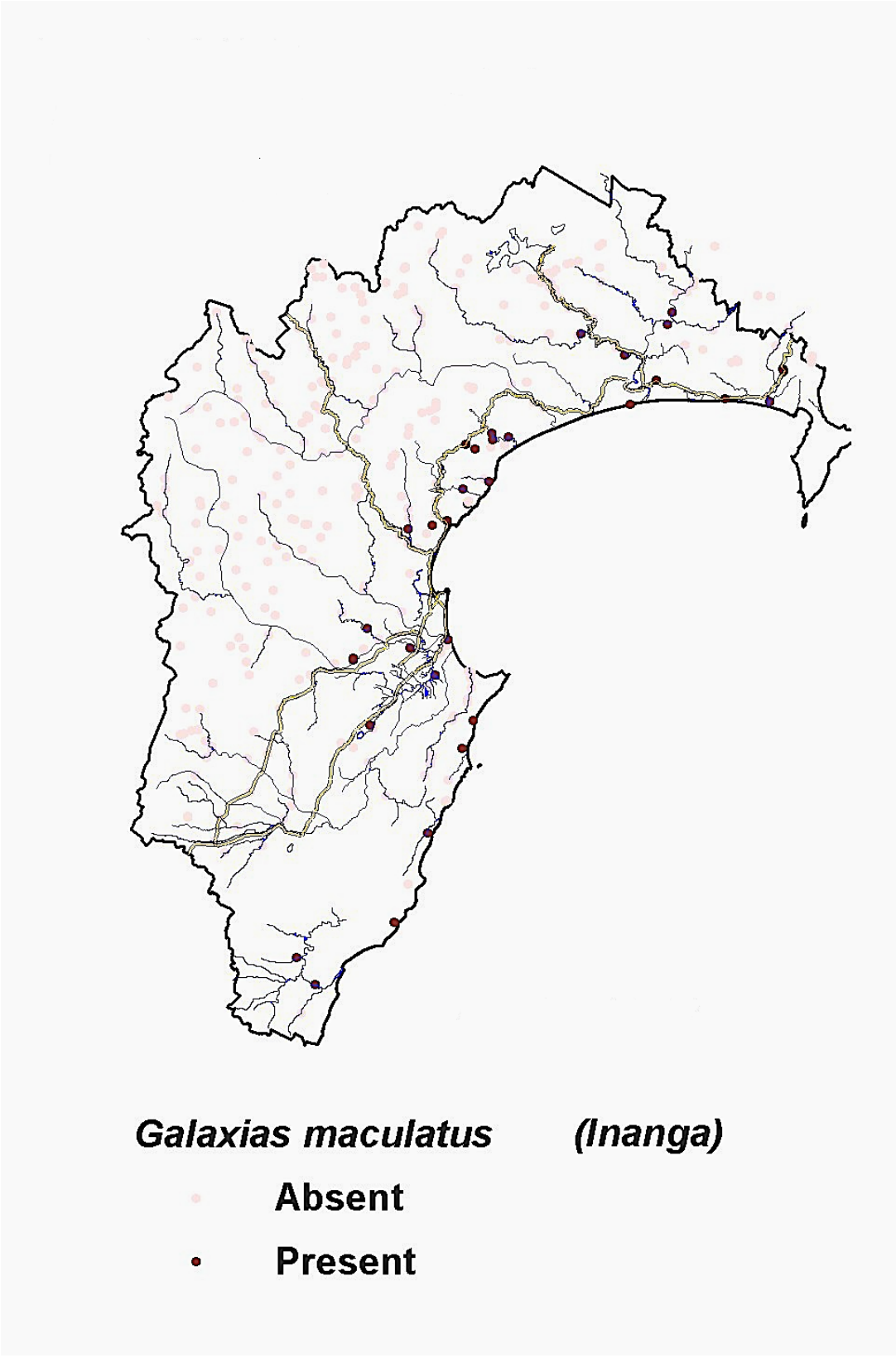


Figure 15: Locations where fish surveys have been undertaken in the Hawke's Bay Region before.

Figure 15 shows the locations at which fish monitoring surveys have been undertaken before. The darker dots indicate the sites at which Inanga have been recorded. Using this information and knowledge of this fish species habitat preferences, a probability map can also be drawn up.

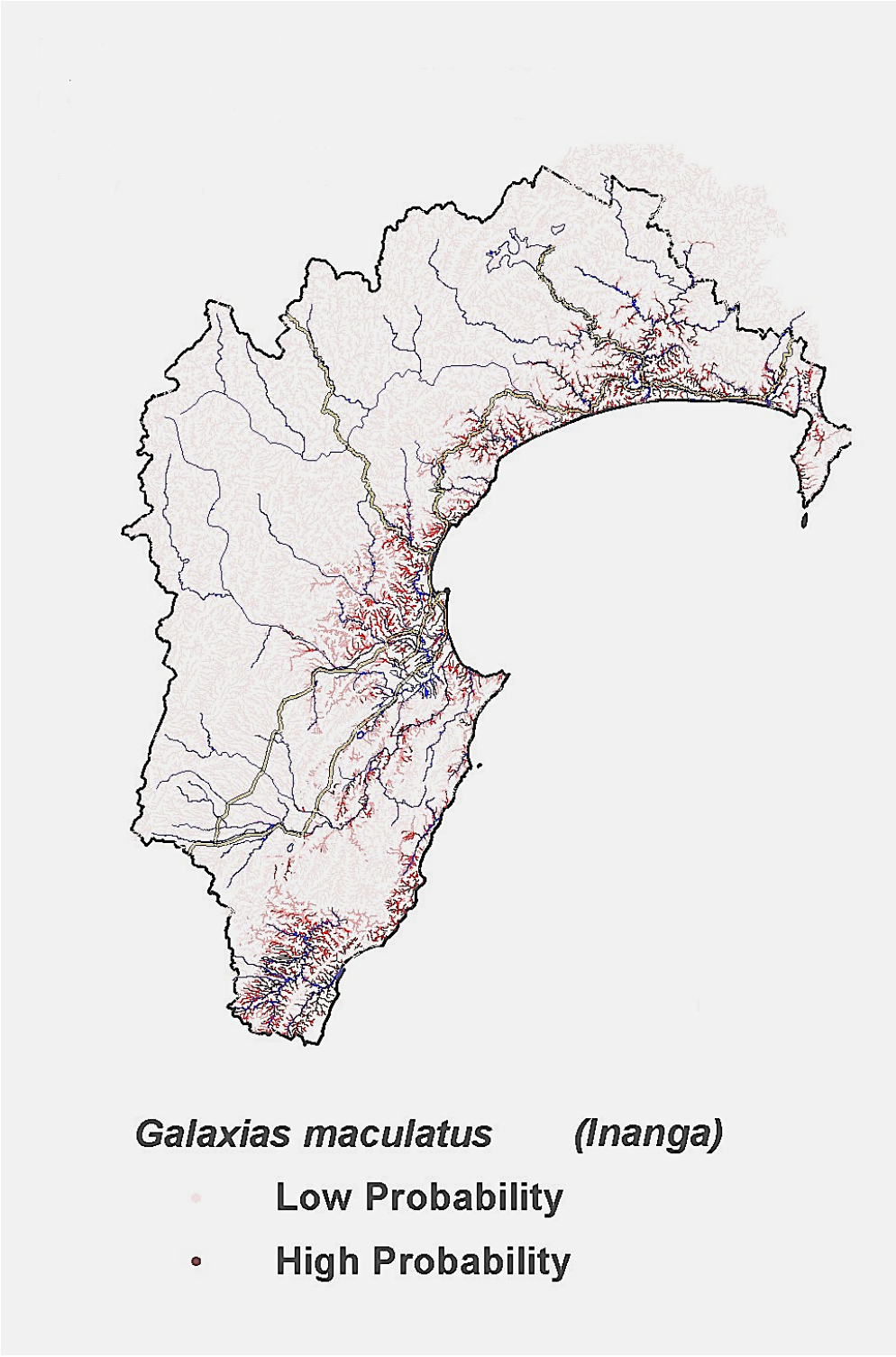


Figure 16: Fish probabilities for inanga.

Figure 16 shows that inanga are more common towards the sea, generally in low elevation areas. This agrees with what we know about inanga in that they are poor climbers and often are in close proximity to river mouths.

Lake Indices

LakeSPI (pronounced "Lake Spy") is an information and management tool that uses submerged plant indicators (SPI) for assessing the ecological condition of New Zealand lakes. A good quality lake would be characterised by having few invasive plant species with a rich diversity of native plant species that can grow to a good depth owing to the clear water.

The underlying principle of LakeSPI is that any New Zealand lake can be characterised by the composition of native and invasive plants growing in the lake and the depth to which they grow.

Once key submerged plant indicators have been identified and recorded using the LakeSPI survey method, LakeSPI applies a simple scoring system to generate 3 LakeSPI indices (or scores):

- Native Condition Index — This captures the native character of vegetation in a lake based on the diversity and quality of native plant communities.
- Invasive Impact Index — This captures the invasive character of vegetation in a lake based on the degree of impact by invasive weed species.
- LakeSPI Index — This is a synthesis of components from both the native condition and invasive condition of a lake and provides an overall measure of the lake's ecological condition.

By using submerged plants as indicators, the effects of water quality, and the impacts of catchment management and aquatic weed invasion can be assessed. This is just one of the many uses of LakeSPI. Lake SPI results are freely available from the following web page.

http://lakespi.niwa.co.nz/lakeSearchResults.do

The following chart shows some results of lake rehabilitation work being undertaken in Lake Tutira. Endothall spray (an aquatic chemical herbicide) was applied at the shore line areas where vehicles off load boats and sterile grass carp were introduced to control the invasive weed Hydrilla. The results shows that the overall ecological condition of the lake is improving.

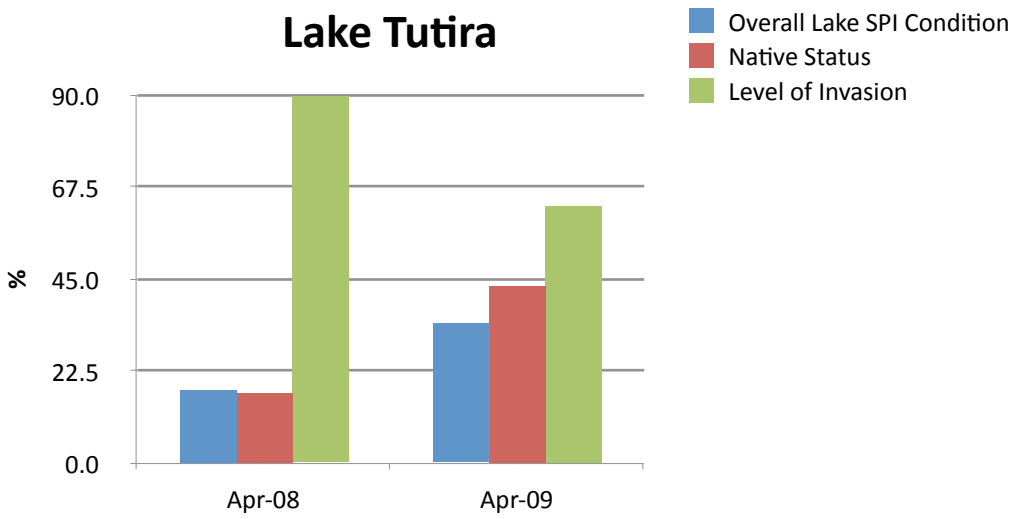


Figure 17: Lake Spi results for Lake Tutira

Trophic Level Index

In New Zealand, the Trophic Level Index (TLI) is widely used to measure changes in the nutrient (trophic) status of lakes in New Zealand. This index considers phosphorus and nitrogen levels, as well as visual clarity and algal biomass. Because water quality is seasonally variable, a TLI value for a lake is not calculated until a years worth of data has been collected. For each of the TLI variables (total nitrogen, total phosphorus, secchi disc clarity and chlorophyll a) an annual average is calculated for which a trophic (pollution) designation is assigned.

To determine the tropic level condition from each individual variable, Burns (2000)recommends the following equations.
TLc = 2.22 +2.54log(Chl a)
TLs = 5.1 + 2.27(log1/SD-1/40)
TLp = 0.218 + 2.92 log(TP)
TLn = -3.61 + 3.01 log(TN)

And finally the TLI = ¼(TLc+TLs+TLp+TLn)

Lake Type (Trophic State	Water Quality Descriptor	Trophic Level	Chla (mg/m^3)	Secchi Depth (m)	TP (mg/m^3)	TN (mg/m^3)
Ultra-microtrophic	Very Good	0-1	0.13-0.33	31-24	0.84-1.8	16-34
Microtrophic		1-2	0.33-0.82	24-15	1.8-4.1	34-73
Oligotrophic	Good	2-3	0.82-2.0	15-7.8	4.1-9.0	73-157
Mesotrophic		3-4	2.0-5.0	7.8-3.6	9.0-20.0	157-337
Eutrophic	Poor	4-5	5.0-12.0	3.6-1.6	20.0-43.0	337-725
Supertrophic	Very Poor	5-6	12.0-31.0	1.6-0.7	43.0-96.0	725-1558
Hypertrophic		6-7	>31.0	<0.7	>96.0	>1558

Table 3: Trophic state designations and descriptors

There is generally good agreement of the trophic state indicated by the four TLI variables because when you have high nutrient concentrations, you will have high chlorophyll a concentrations and poor secchi disc water clarity (owing to the absorbance of light by the phytoplankton). Alternatively when you have low nutrient concentrations you are likely to have low chlorophyll a concentrations (owing to less phytoplankton growth) and good water clarity (secchii depth).

Percentage Average Change

Percent Annual Change (PAC) allows an estimate to be made of the probability of change of trophic level of a lake with time. To achieve this, the data on the trophic state variables (Chla, SD, TP and TN) are deseasonalised before plotting them in a time-trend plot to increase the sensitivity of the trend detection procedure. These values and the values from the deseasonalised procedure are plotted as a function of time. Calculation of the PAC value for a lake over the analysis period is then based upon determining only significant slopes for each variable from the p value of the time trend plot. If a p value is <0.05 then the null hypothesis of no trend is rejected and the trend is considered statistically significant (Burns et al 2000). The PAC is determined by calculating the slope of the regression line of the residuals and dividing this value by the average value of the variable (secchi depth, chlorophyll a, total nitrogen, total phosphorus) during the period of observation. If there is no significant change observed with a variable, it is given a PAC value of zero. The average PAC value for a lake is then determined from the averaging of PAC values of

the four variables together and determining the p value of their average. For determination of the probability of change of trophic level, Burns et al. (2000) gives the following guidelines:

P<0.1: definite change

0.1-0.2 = probable change

0.2-0.3 = possible change

p>0.3: no change

PAC values are calculated only from significant trends (P<0.05) determined by parametric analysis of deseasonalised data.

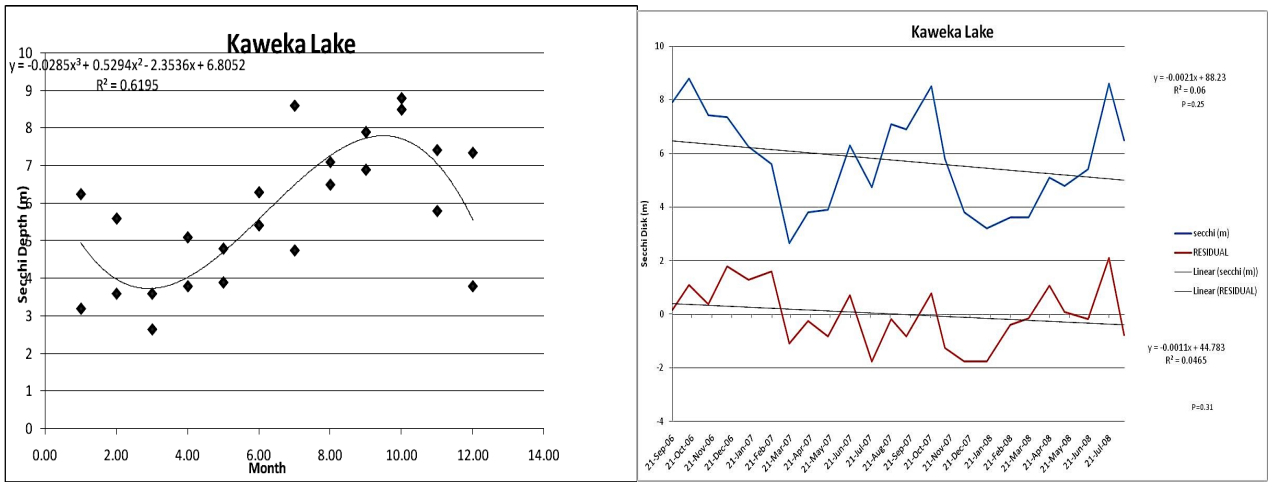


Figure 18: Some water clarity readings taken from 2 years worth of data at Kaweka Lake

Figure 18 shows some some water clarity readings (measured by secchi disc) taken from Kaweka Lake. the graph on the left shows that there is significant seasonality in that the water is much clearer in the months of October compared to March. The plot on the right shows the raw data plot (blue line) and the residual plot (red line) for this variable. Both plots show a downward trend over time however neither of the trends are very significant (p<0.1).

Hypolimnetic Volume Oxygen Depletion Rate

The fifth key variable in assessing lake water quality is dissolved oxygen and the hypolimnetic volume oxygen depletion rate (HVOD) which gives a measure of how quickly oxygen is consumed within the hypolimnion during the course of the year. This value is usually unique to a lake but changes in the HVOD of a lake with time are indicative of change of trophic level.

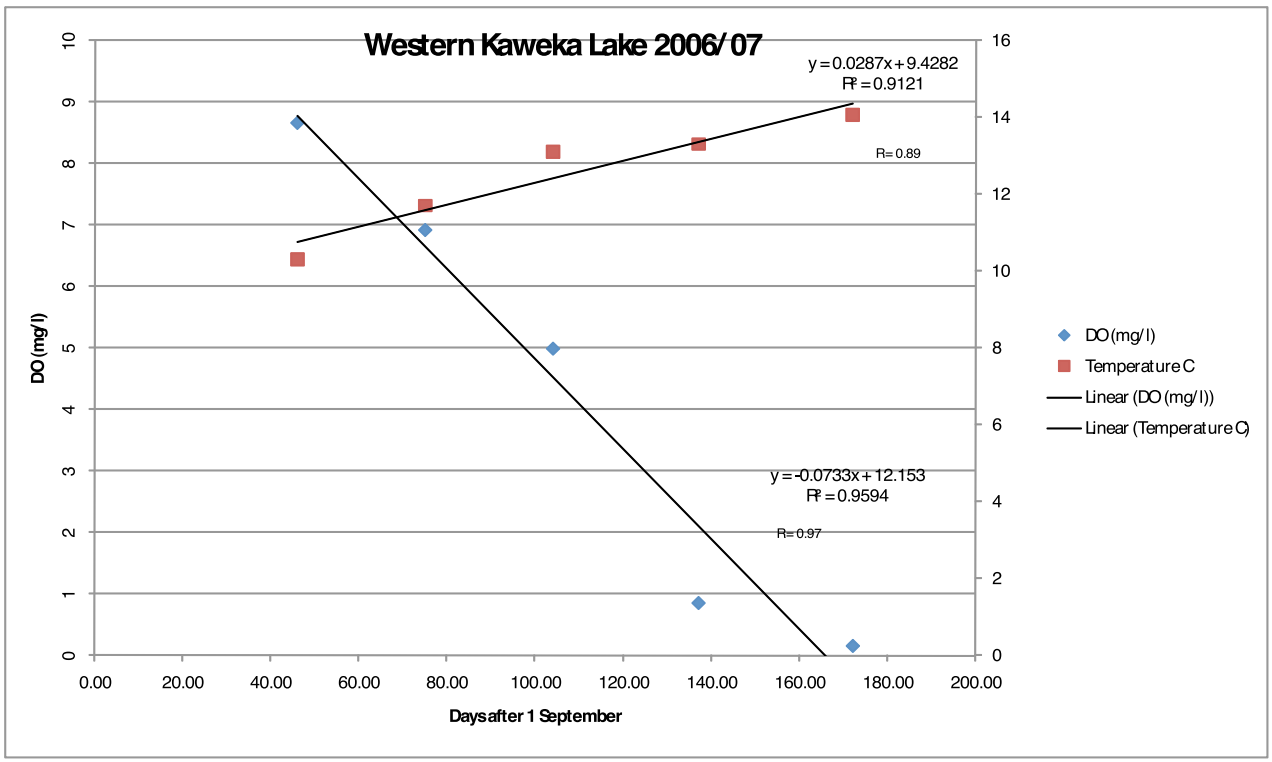


Figure 19: Hypolimnion oxygen depletion rate in Kaweka Lake

Figure 19 shows that it takes about 110 days for all of the oxygen in the hypolimnion to have been consumed. Lakes with quicker rates of oxygen depletion are likely to be more eutrophic as it indicates that the organic matter is getting processed much faster.

The Water Quality Index

One of the difficult tasks facing environmental managers is how to convey their interpretation of complex environmental data into information that is understandable and useful to non-technical people, including the general public (Ott, 1978). This is particularly important in state-of-environment reporting where the intended audience, including policy makers, can not be expected to understand “raw” environmental data such as that for water quality. Internationally, there have been a number of attempts to produce a single value measure or index, that meaningfully integrates the data pool. The philosophy of such an index is broadly analogous to other indexes widely used in economics, such as stock indexes (e.g. Dow) and consumer price indices.

The water quality index provides a summary of water quality using 6 variables to give an overview of the suitability of the water for swimming.

E.coli or faecal coliforms are included in the index to assess the safety for contact recreation from faecal contamination while pH is included in the index as the pH of the water can affect irritation of the bathers eyes. Colour and visual clarity (or turbidity) are also included in the index as these variables are often considered “before diving in”. Finally dissolved nutrients (soluble inorganic nitrogen and soluble reactive phosphorus) are also included in the calculation of the index as high concentrations of nutrients are known to cause considerable algal blooms which can deter any bather from diving in.

For each of the variables, a sub index value is read off a curve (which has been generated by expert opinion and some maths) which is then recorded . An example follows:

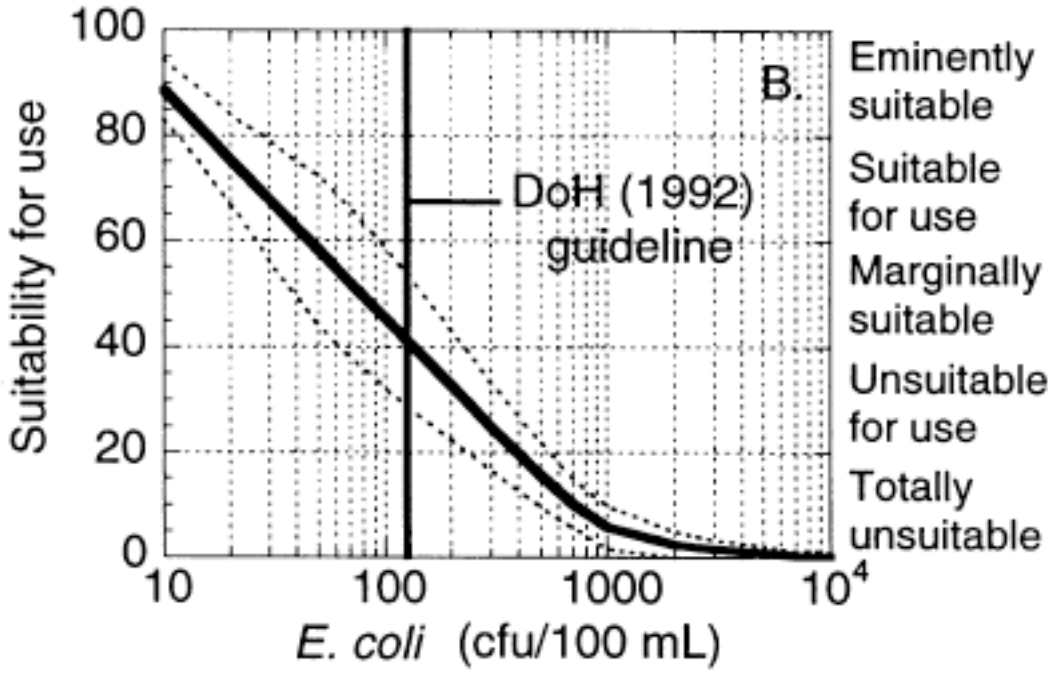


Figure 20: Sub index curve generated for E.coli

From figure 20 you can see that if E.coli concentrations reached 300 cfu / 100 ml the corresponding sub index value is just under 20 falling in the Totally Unsuitable category. Alternatively if E.coli concentrations were at 100 cfu / 100 ml, the sub index value would be 40 indicating marginally suitable water for swimming.

This process is undertaken for each water quality variable of the WQI and the final WQI value is that sub index value that scores the lowest. For example supposing the sub index values generated for the WQI were:

E.coli	40
pH	50
Colour	55
Water Clarity	30
Soluble Inorganic Nitrogen	45
Soluble Reactive Phosphorus	35

Then the overall WQI value would be 30 as the lowest sub index value was 30 (that assessed for water clarity).

The water quality index was first developed in the 1980s and was more recently revised in 2001 by the National Institute of Water and Atmospheric Research (NIWA). It hasn't had much uptake from regional councils however it has formed part of national state of the environment reporting by NIWA.

Site Selection and Survey Design

For any water quality monitoring program, your selection of sites and water quality variables will depend on the objectives of the study and all studies have a limited budget. Therefore the water quality scientist has to provide some rationale for the site selection and survey design of a monitoring programme.

For state of the environment monitoring programs, the site selection will be driven by achieving geographical representativeness, diversity of stream types, inclusion of reference sites (for comparisons) and vehicular access.

For consent monitoring programs, the objective is to determine whether an impact has or is occurring as a result of the activity. In an ideal world the Before After Control Impact (BACI) design is recommended. This is where monitoring is undertaken before the consent is granted, after the consent is granted and monitoring sites are located at control (sites that are not subject to the consent activity) and impact sites (sites that are subject to the effects of the consent activity). However time and budget constraints may not afford the luxury of before and after monitoring in which case we have to deal with comparing control sites from impacted sites only.

It is important to note here the difference between control sites (not subject to the effect being measured) from reference sites (near pristine and not subject to any anthropogenic impacts). Reference sites are important to have for state of the environment monitoring because changes in water quality may be due to natural occurrences (e.g. due to a high rainfall year, or a drought year) but may be less critical for consent monitoring programmes.

Water quality is inherently variable both over time (temporal variability) and space (spatial variability). Therefore to determine whether an impact has occurred, it is often necessary to take replicate (multiple) water samples from replicate sites. It has often been said that little can be said from one water quality sample and I tend to agree as it often merely represents a snap shot in time and without further sampling you are not in a position to determine the true variability of the water quality at that monitoring site.

When making comparisons of water quality at various sites, it is important that you are comparing ‘apples with apples’, by this I mean it is important that the sites you are comparing have very similar features of climate, topography, geology, and land cover as these ‘natural features’ of a catchment can drive the water quality of the receiving water a lot more than what land use impacts can. Fortunately our rivers have been classified by NIWA using a tool called the river environment classification (REC, Snelder et al 2003). This is a geographic information system (GIS) layer which enables you to determine the classification for any stream in New Zealand. The following table provides the factors used to determine the river environment classification.

Factor		Descriptors
Climate	Catchment climate	Cool extremely wet, Cool wet, Cool Dry, Warm extremely wet, Warm Wet, Warm Dry
Topography	Catchment topography	Glacial mountain, Mountain, Hill, Low elevation, Lake, Spring, Wetland, Regulated
Geology	Catchment geology	Hard sedimentary, Soft sedimentary, Alluvial, Volcanic Basic, Volcanic Acidic, Plutonics, Miscellaneous
Land Cover	Catchment land cover	Bare ground, Indigenous Forest, Exotic Forest, Scrub/Tussock, Pastoral, Urban
Network Position	Position of the section of river within the catchment (stream order)	Low order, medium order, high order
Valley- Landform	Landform of the valley the river section is located within	HIgh, Medium or Low Gradient

Table 4: River environment classification descriptors

As a general rule, high altitude rivers tend to have excellent water quality often because they have cool water temperatures, ample rainfall which regularly flushes the stream clean, a hard sedimentary geology which is surrounded by indigenous vegetation and a high gradient which means that the water velocity is fast and well aerated. At the other end of the spectrum ward dry low elevation streams tend to have poor water quality because they have warm water temperatures, less rainfall to flush the stream clean, a soft highly erodible geology which is usually surrounded by pastoral or urban land cover and a low gradient, meaning water velocity is slower. Most of the descriptors in table 4 are self explanatory however the network position probably needs a bit more explanation.

Stream order is based on the number of branches a stream network has, the following diagram shows this.

1st order stream 2nd order stream

When studying stream order, it is important to recognize the pattern associated with the movement of streams up the hierarchy of strength. Because the smallest tributaries are classified as first order, they are often given a value of one by scientists . It then takes a joining of two first order streams to form a second order stream. When two second order streams combine, they form a third order stream, and when two third order streams join, they form a fourth and so on.

If however, two streams of different order join, neither increases in order. For example, if a second order stream joins a third order stream, the second order stream simply ends by flowing its contents into the third order stream, which then maintains its place in the hierarchy. First and second order streams are also called headwater streams and constitute any waterways in the upper reaches of the watershed. In Hawke's Bay 95% of its stream length is in 1st or 2nd order streams. It is estimated that over 80% of the world's waterways are first through third order, or headwater streams.

Unlike the smaller order streams, the medium and large rivers (3rd to 8th order) are usually less steep and flow slower. They do however tend to have larger volumes of runoff and debris as it collects in them from the smaller waterways flowing into them.

The following map shows some streams and rivers that have a warm wet low elevation soft sedimentary pastoral stream river environment classification.

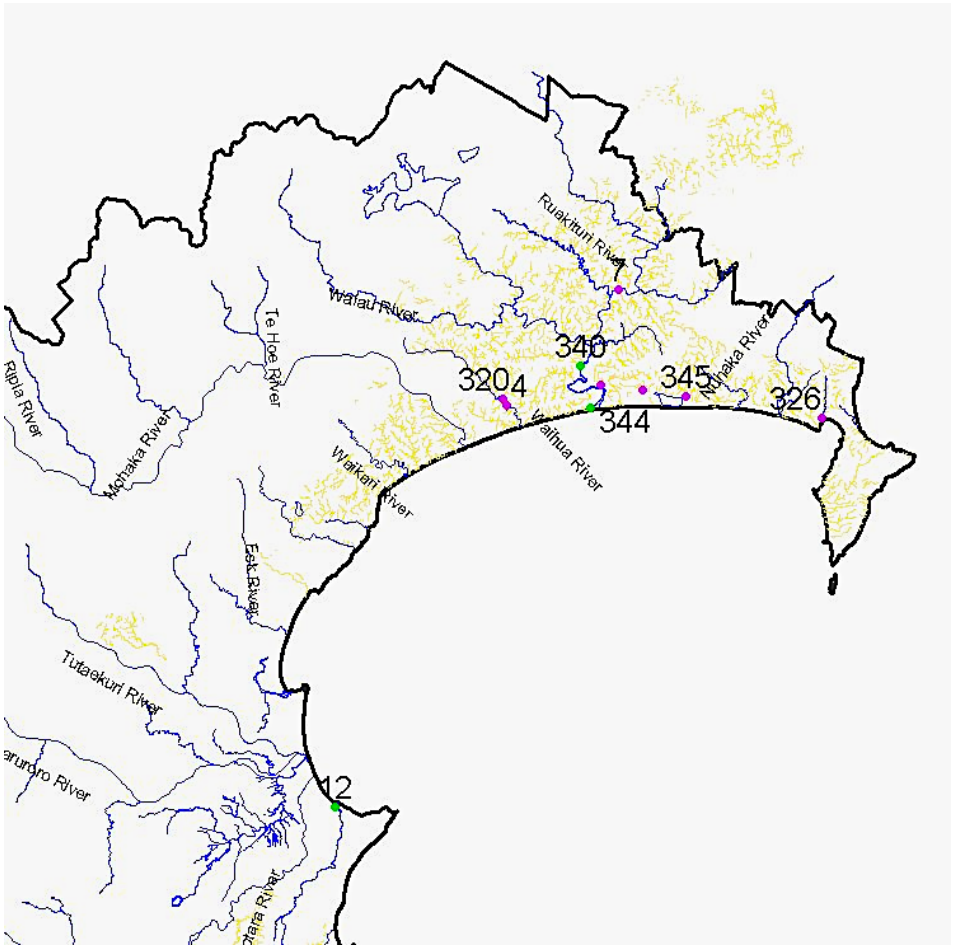


Figure 21: Warm wet low elevation soft sedimentary pastoral streams and rivers of the northern Hawke Bay.

Figure 21 shows where I can find warm wet low elevation soft sedimentary pastoral streams. (the shaded areas). This map makes it useful for me if I was wanting to make a meaningful comparison of water quality or ecology in rivers such as the Ruakituri, Nuhaka, lower Waiau or Waikari Rivers.

Survey Design.

Unlike a census in which every individual of the population is contacted counted and evaluated, water quality monitoring programmes rely on sampling a proportion of the population i.e you cannot analyse every molecule of water in the Hutt River for nutrient concentrations, as you would need to completely drain the river to do that.

When sampling from a population you need to be aware of measurement error and sampling error. Measurement error is that which may occur e.g. as a result of a laboratory fault, or perhaps your dissolved oxygen meter was low on batteries. As the name implies they relate to an error made during measurement. Alternatively sampling error is that which arises in your estimation of a population statistic (e.g. the mean or median value). Supposing the soluble reactive phosphorus 3 hourly concentrations in a river during a day were, 0.03 mg/l, 0.04 mg/l, 0.02 mg/l, 0.01 mg/l, 0.05 mg/l, 0.02 mg/l, 0.03 mg/l and 0.04 mg/l then the average concentration would be

0.03+0.04+0.02+0.01+0.05+0.02+0.03+0.04 /8 = 0.03 mg/l

However if you had sampled every 12 hours, your estimation of the mean concentration would be 0.01+0.04/2 = 0.025 mg/l

Alternatively you may have sampled once in the middle of the day for which you would report a daily concentration of 0.1 mg/l.

Clearly the degree of error in your estimation of a population statistic is dependent on your sample size and frequency. There will always be error in sampling from a population and it is for this very reason that statistics must be conducted on sampling data as it can provide you with a level of confidence (called the confidence interval) in statements you make about the overall population you are studying (e.g. dissolved oxygen concentrations, or nutrient concentrations or ecological health of a river).

Sampling Error

You will often hear of two types of error in statistics

Type I error is “thinking you have found a statistically significant difference when in fact there isn’t one”. Or as statisticians would say falsely rejecting a true hypothesis. In compliance terms think of it as convicting an innocent person!

This type I error risk is symbolised by the greek letter alpha (α) For example if my alpha value is set at 0.05 it means I run a 5% risk of making this type of error.

Type II error occurs when you think there is no statistically significant difference when in fact there is. Or as statisticians would say a Type II error occurs when we fail to reject a false hypothesis. In compliance terms think of it as letting the guilty go free! The risk of a type II error is symbolised by the greek letter beta (β). Thus if my beta value is set at 0.1 it means I run a 10% probability of making this type of error.

The lower the type I error risk is set (e.g. $\alpha = 0.05$), the greater the likelihood that you will make a type II error.

In statistics, a sample is the collection of water samples you have gathered to make an inference about the population you are sampling. With environmental sampling your main goal should be to achieve representativeness with the sample you have collected from the water body. Furthermore your sampling should be as objective (unbiased) as possible. The sampling design can be a number of types depending on the objectives of the study.

Randomised Sampling

As the name implies, the sampling is completely random through space and time. This can achieve good representativeness of the population provided an adequate sample size has been taken however there are also downsides to this type of sampling such as the difficulty in randomly organising survey teams e.g. " can you work this Sunday but not next Tuesday?" "Can you not take any annual leave because I might randomly need you to come in to do some water quality sampling?".

To be honest the random design is pretty inflexible. Also if the population is spread out geographically, then considerable cost may be involved in achieving representativeness unless the sample is grouped in some way. A common technique to solve this issue is to use stratified random sampling. This is where the population is divided into 'strata' (e.g. runs, riffles or pools of a river) which then enables the survey team to divide resources equally amongst the strata to ensure a similar degree of sampling effort and representativeness is achieved for each strata (or habitat type). Alternatively you also have the option to place more effort on the larger strata, for example 1st and 2nd order streams can represent up to 95% of overall stream length in a region, therefore the scientist could decide to stratify the rivers by size and place more sampling effort on the smaller streams to achieve overall representativeness. This latter example is called proportional stratified sampling in which the sample size in each stratum is proportional to the size of the stratum. Proportional sampling is appropriate when all the strata are equally variable. When this is not the case the more varied strata should be sampled more heavily.

Systematic Sampling

Systematic sampling is where the sampling is undertaken at a regular fixed interval e.g. river water quality monitoring for state of the environment monitoring is normally undertaken on the same day of each month at a site. This enables the time series to be evenly distributed through time. This type of sampling design has the advantage of being able to organise survey teams more easily and to allocate more resources when staff may be away. Some people have argued that systematic sampling will not achieve good representativeness as the regular interval sampling may miss peaks and troughs between each sampling event. However as the data sets become increasingly larger in size, this concern is likely to become less so.

A disadvantage of systematic sampling is that it makes the sampling programme inflexible if other emergencies arise (e.g. the Rena disaster called for assistance with assessment and clean up however not everyone could be available to help owing to state of the environment monitoring commitments).

For aquatic macroinvertebrate sampling, systematic sampling often does not work. This is because of the effects of freshes or floods which flush invertebrate larvae out of the stream bed and deposit

them downstream or flush them out to sea. The recommended time to sample macroinvertebrates is 2 weeks after a fresh or flood as this 2 week window enables most invertebrates to recolonise the substrate of the river. Therefore the sampling that results from macroinvertebrate monitoring is not entirely systematic or random. Sometimes the regular systematic February Summer (worst case scenario) sampling may be postponed to the next month. While this may introduce some seasonal bias (cooler water temperatures in March, perhaps less periphyton growth) there are methods that can be used to remove seasonal bias (see next section on data analysis).

Preparing Data For Exploratory Data Analysis

Prior to undertaking any analysis it is necessary to prepare the data so that it is in the correct format for analysis. As a general rule if the data for each site is temporally ordered by rows and the measured variables are in adjacent columns, you're half way there. Most statistical software packages will not accept character variables (letters, words) unless that column specifies 'character', furthermore any numerical columns (columns with numbers) should not have any letters or symbols (e.g. < 0.005, >1000) otherwise the data cannot be analysed.

Laboratory data may often contain < or > symbols to indicate the limits of detection of the laboratory e.g. a soluble reactive phosphorus reading of < 0.003 mg/l means that the concentration was less than 0.003 mg/l and that the laboratory cannot confidently indicate the true value below this limit. Alternatively a faecal coliform bacteria concentration of >10,000 cfu/ 100 ml indicates that the laboratory cannot confidently tell you how much above 10,000 cfu / 100 ml was in the water sample as this represents the maximum value the laboratory can confidently specify the bacteria count. As a general rule if you need to remove detection limit symbols from your data (< and > symbols), you should change < limits to half their value and > limits to their value. So in the previous discussion all <0.003 entries would be changed to 0.0015 and all > 10,000 cfu / 100 ml values would change to 10,000 cfu / 100 ml.

For data that has been collected in the field (e.g. habitat quality assessments, flow estimations or substrate assessments) the data has not been electronically received form a laboratory therefore the data entry should be rechecked by a second person. This ensures that an independent second audit has picked up on any errors in data entry.

Exploratory Data Analysis

For all data whether it has been electronically received or manually entered by field staff, the analyst should look for any extreme outliers. For example, supposing you were looking at some nitrate nitrogen data received from a laboratory and the values for the site were:

Date	Site ID	Site Name	Nitrate Nitrogen (g/m3)
26/1/2010	16	Hunters Ck	0.157
28/2/2010	16	Hunters Ck	0.327
26/3/2010	16	Hunters Ck	0.267
24/4/2010	16	Hunters Ck	0.359
25/5/2010	16	Hunters Ck	3.97
27/6/2010	16	Hunters Ck	0.259
26/7/2010	16	Hunters Ck	0.395

Clearly the sampling of 25/5/2010 is extraordinarily high compared to the rest of the data. I would be asking myself "Why is the nitrate nitrogen value on this date so high?", "What were meteorological conditions like on this date?", "Could this be laboratory error? If so phone the laboratory to ensure a mistake wasn't made at their end", "Could some contamination of the water

sample have occurred?" "Did the sample get to the laboratory within 24 hrs?", "Was the sample adequately chilled on receipt?"

All of these questions can be answered by a few phone calls to the laboratory and the field staff. If all indications that no error was made then you have the option of either accepting the value in your data set or deleting the data point on the grounds that it represents an outlier that is not representative of the population of interest (in this case nitrate nitrogen concentrations of Hunters Creek). There are a plethora of statistical tools available as a decision support if you are uneasy about using that data value (called outlier analysis). Should you choose to remove the data point from your data set, you need to clearly state in the technical report write up why you chose to do this.

The influence the outlying data point will have in your statistical analysis will depend on the testing you are doing. For example, for parametric statistical testing (testing for which the data must conform to a bell shape curve), the outlier is likely to affect the result. Alternatively if the data is subject to non parametric testing (testing for which the data does not need to conform to any distributional pattern), the outlier is unlikely to affect the end result.

Another useful thing to do is to plot a time series of the data. This enables me to not only look out for extreme values in the data set but also for other anomalies such as stepped changes which may represent a change in laboratory detection limits. Exploring the time series plot enables us to look for any gaps which represent when sampling was not undertaken. It may have been that flows were too high to safely monitor the water, perhaps a tree fell across the road preventing access. We have to accept that gaps in data will occur due to unforeseen circumstances. In most circumstances the missing data point is unlikely to badly affect the quality of your data analysis. However if the need arises there are methods available to 'fill the gap' such as taking the mean value of the two points adjacent to the gap or alternatively using a correlated value from another variable, an example follows.

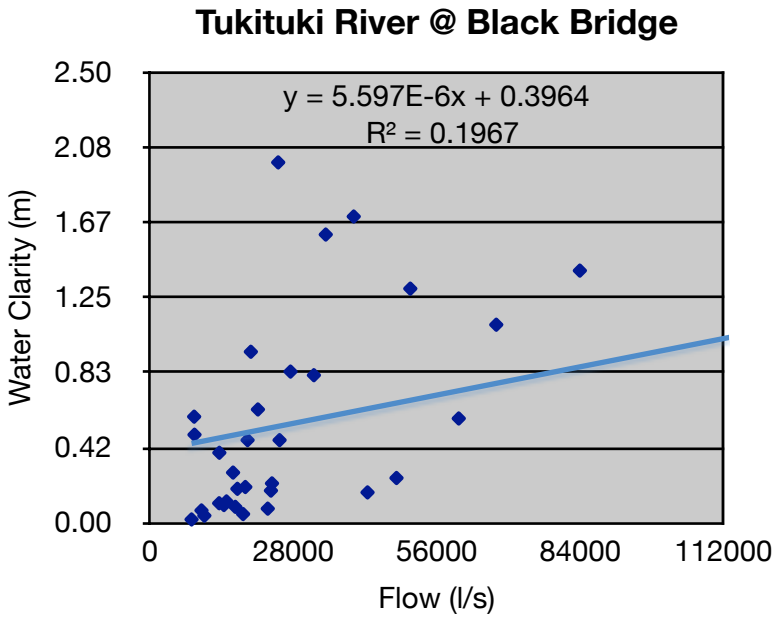


Figure 21: Water clarity vs flow for the Tukituki River

Supposing you were unable to survey the water on a particular day, however the flow recorder located at the site was relaying data back to the office and it read 2800 l/second in the Tukituki River at Black Bridge. The correlation of water clarity with flow from historical records was found to be that in the figure 21. Then from this we could estimate the water clarity on the day to be = 5.597^{-6} *2800 +0.3964 = 0.49m water clarity.

Keep an eye out for seasonality with your exploratory data analysis. This can be easily done by simply viewing the time series or better still viewing seasonal box and whisker plots some examples of seasonal data are provided in figures 51 and 52 on page 107.

If your data demonstrates seasonality it may be necessary to deseasonalise the data (see Seasonal Kendall Trend Test p114) particularly if you are looking to do time series analysis.

Before I proceed with any mention of statistics it's probably best to refresh your memory on some commonly used statistical notation.

Common Notation In Statistics

μ the population mean
X_i the ith numerical datum in a sample
$\bar{X}$ the arithmetic mean
Mode – the most frequently occurring value
Median – the middle value when the data is ordered from smallest to largest
Geometric mean – is the nth root of n samples e.g. if n = 6 then the geometric mean would be equal to the $\sqrt[6]{X_1X_2X_3X_4X_5X_6}$ NB ensure your data set does not have zero values in it for this calculation otherwise your geometric mean will = zero. Often used in bacteriological standards.
Range = largest value to the smallest value
Interquartile range = median of the top half of the data – the median of the bottom half of the data, denoted by IQR
Variance = the sum of all datum deviations from the mean,
i.e $S^2 = 1/n\text{-}1 \sum_{i=1}^{n} (X_i-\bar{X})^2$ using n-1 degrees of freedom.
Standard deviation = the positive square root of the variance.
Coefficient of variation = standard deviation / mean

Transforming Data To Achieve Normality

In many natural processes, random variation conforms to a particular probability distribution known as the normal distribution which is the most commonly observed distribution. The shape of the normal distribution represents that of a bell, so it is sometimes referred as the bell curve.

The bell curve can be described using two statistics, the mean and the standard deviation and has the following characteristics

•68% of the data will fall within one standard deviation of the mean

•95% of the data will fall within two standard deviations of the mean

•99% of the data will fall within three standard deviations of the mean

•the bell shape curve is symmetrical about the mean (or average)

•the curve is unimodal (single peak)

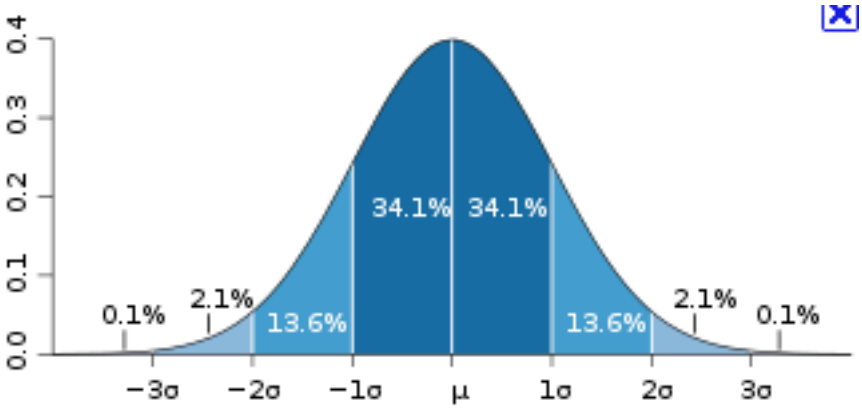
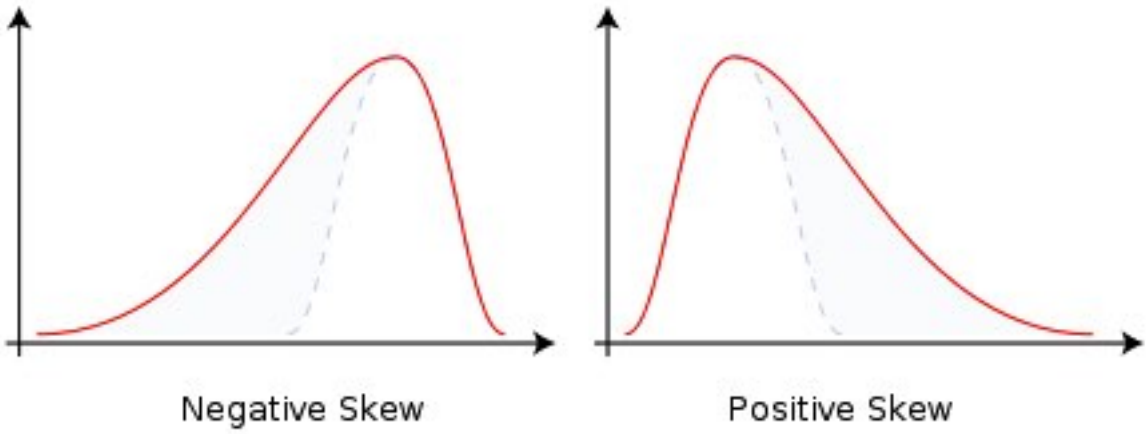


Figure 22: The normal distribution

Parametric statistical testing requires that the data's probability distribution conforms to the bell shape curve, furthermore if populations are to be compared (e.g. the turbidity levels at site A compared to site B) parametric statistical testing requires that the populations being compared have the same variance (degree of variability).

These data characteristics often do not exist with raw water quality data (which instead are usually positively skewed) so a mathematical transformation may need to be applied to the data. The type of transformation will depend on how skewed the raw data is. By skewed I mean how wonky the probability distribution curve is.



For normally distributed data, the mean and median are equal whereas with negative skewness the mean of the data is less than the median (called left skewed) while positive skewness results when the mean of the data is greater than the median (called right skewed).

Depending on how skewed the data is, the following table shows some mathematical transformations that can be applied to the data set to achieve normality.

Power	Formula	Name	Result
3	x3	cube	stretches large values, use for left skew data
2	x2	square	
1	x	raw	
1/2	$\sqrt{x}$	square root	squashes large values, use for right skewed data
0	log(x)	logarithm	
-1/2	$-1/\sqrt{x}$	reciprocal root	
-1	-1/x	reciprocal	

Table 5: Mathematical tools for transforming data.

Checking for Skewness

There are a plethora of tools available for the researcher to analyse the skewness of their raw data. Perhaps the simplest of which is to look at a histogram. A histogram is a frequency bar chart of the data that is divided into intervals and each bar of the chart displays the frequency (or count) of data points that fall within those intervals.

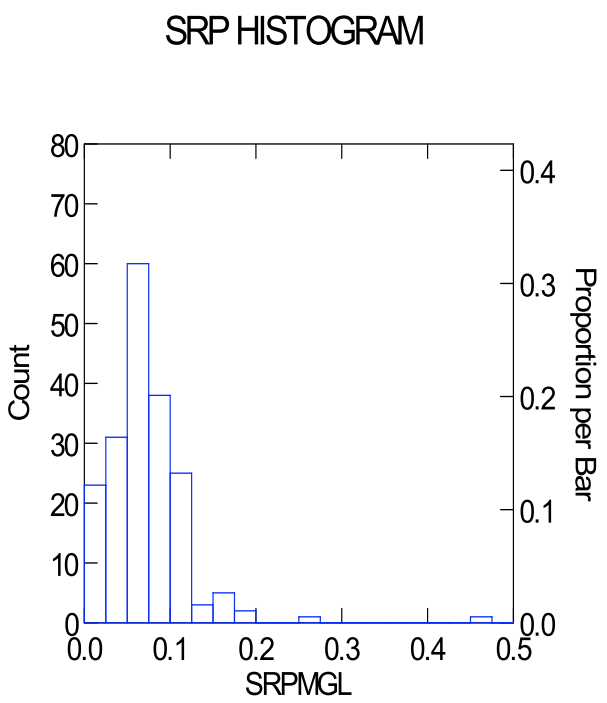


Figure 23: Histogram of soluble reactive phosphorus concentrations (mg/l) in the Clive River.

Figure 23 shows the histogram of soluble phosphorus concentrations in the Clive River of Hawke Bay. You can see that the data is heavily right skewed with a lot of counts between 0 to 0.2 mg/l and then there are less counts above 0.2 mg/l. The data clearly is not symmetrical and would warrant a transformation if parametric statistics was to be performed on it.

Another useful tool to examine skewness is the box and whisker plot. This is probably the most commonly used statistical graphic used in water quality science.

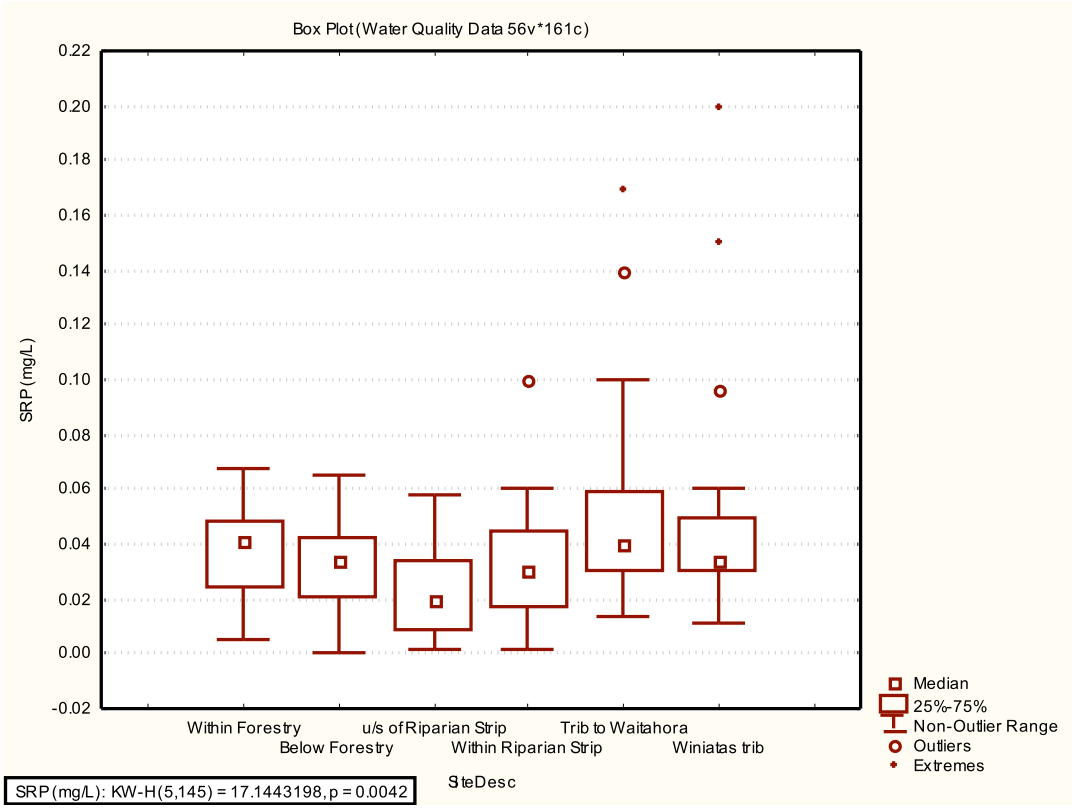


Figure 24: Box and whisker plots of soluble reactive phosphorus concentrations at selected sites of the Waitahora Stream Catchment of Northern Hawke's Bay.

Figure 24 shows the soluble reactive phosphorus concentrations for a group of monitoring sites in the Waitahora Stream catchment. For each site a box and whisker is displayed. The central box represents the data that lies between the 25th and 75th percentiles which means it shows the range of values in which the middle 50% of the data lies. Within each box is a smaller box representing the median which is the middle value of the data if the data were ranked from highest to lowest. The whiskers of each box represent data below the 25 percentile and above the 75th percentile that are not considered outliers while outliers are signified by a circle and extreme outliers are signified by a dot.

You can see from the plot that Winiatas Tributary (the right hand side box and whisker) is quite skewed in that the median is very close to the lowest 25 % of the data. Furthermore there is a wide spread of outlying points at the higher values. This is right skewed data however the statistical testing (lower left box) conducted on the data is a non parametric test (which makes no assumptions about normality), therefore transforming the data is not necessary.

It should be noted here that there is no standard format for box and whisker plots, they are generally custom made according to the software package you are using, so read up on what the graphic symbolises in the help section of the programme.

Another tool for examining skewness is a statistical test called the Anderson Darling Test in which the data is tested for how far from the expected normal distribution the data is. The difference between the observed value (from the raw data) and the expected value (from a hypothetical normal distribution) is called a residual. The residuals are plotted on a normal probability plot and if

the residuals conform closely to the expected observation (in this instance a straight line) then the data is considered normally distributed. Some examples follow:

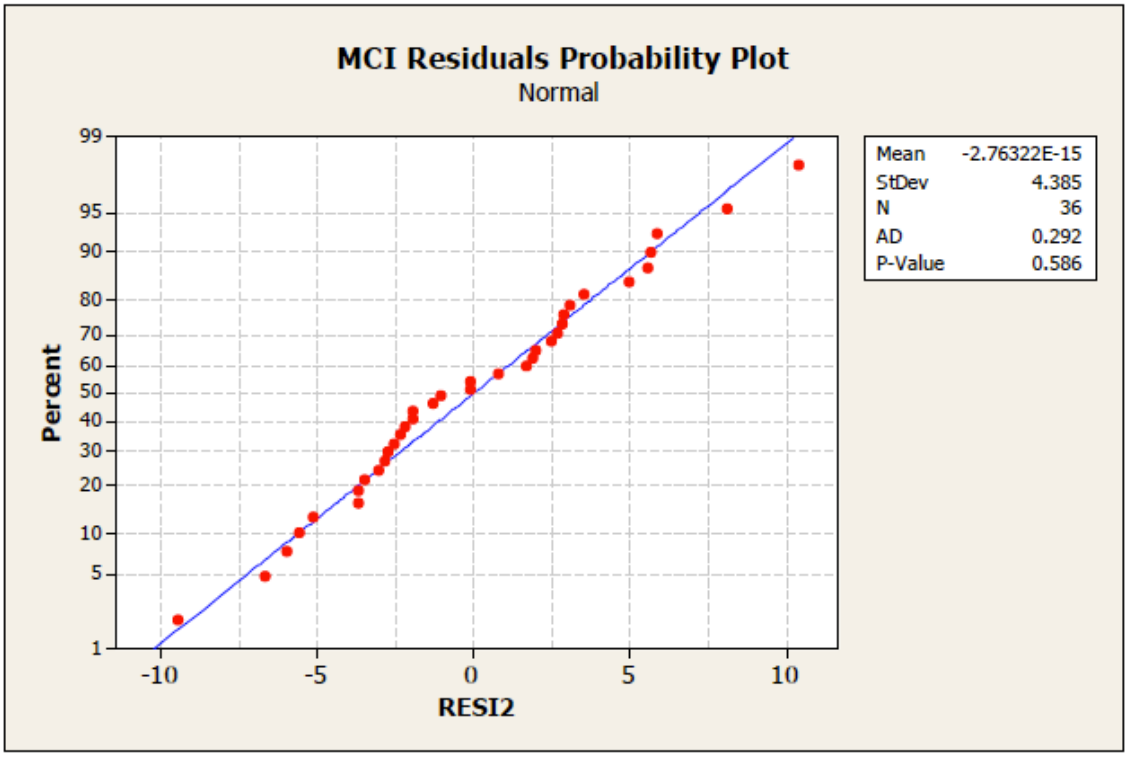


Figure 25: Normal probability plot of MCI residuals for the Mohaka River in Summer

Figure 25 shows that the residuals (red dots) conform closely to a straight line and the Anderson-Darling test (AD) indicates that normality has been achieved.

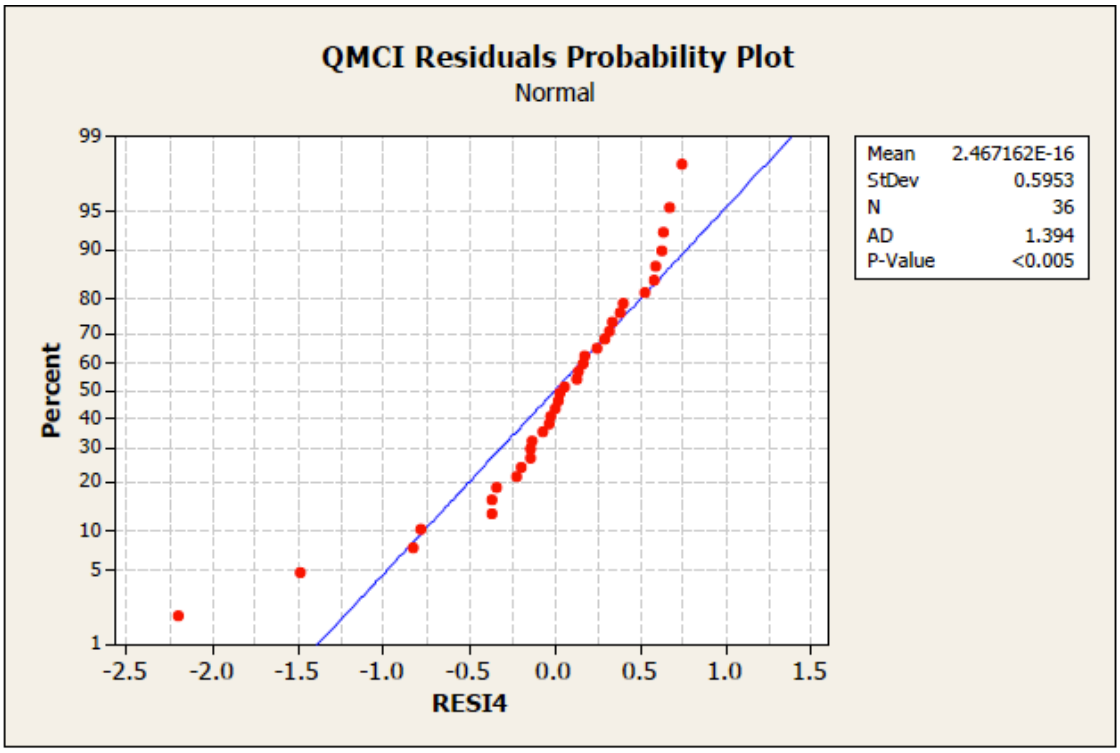


Figure 26: Normal probability plot for QMCI residuals of the Moahaka River

Figure 26 shows that the residuals (red dots) do not conform to the expected normal distribution (blue line) at all. The Anderson Darling test (AD) confirms that the data is not normally distributed. Keep in mind that if you subject a data set to a transformation to achieve normality that you can over transform, and example is given during the workshop.

Testing Hypotheses

A hypothesis is a proposed explanation for a phenomenon. In the case of the Anderson Darling Test the null hypothesis (default assumption) is that the data is normally distributed unless demonstrated otherwise. If we chose an alpha value of 0.05 ((probability of a type I error (see page 43)being less than 5%)) then we reject the null hypothesis if a p value is less than 0.05. So in figure 26, the null hypothesis that the data is normally distributed is rejected in favour of the alternative hypothesis that the data is not normally distributed. However in figure 25, we fail to reject the null hypothesis as the p value is less than the critical alpha value of 0.05.

It is important to understand that the point null hypothesis can never be proven. A set of data can only reject a null hypothesis or fail to reject it. For example, if comparison of two groups (e.g.: treatment, no treatment) reveals no statistically significant difference between the two, it does not mean that there is no difference in reality. It only means that there is not enough evidence to reject the null hypothesis (in other words, one fails to reject the null hypothesis).

There are a plethora of tools out there for examining skewness of data sets, each with their own strengths and weaknesses however the intricacies of these tests is beyond the scope of this course. If you've understood the three methods I've described to date, it's likely to be all you need for examining water quality data.

Graphical Techniques To Demonstrate Water Quality

There are many graphical tools that can be used to demonstrate water quality. In this chapter I will outline some graphical techniques that can be used to demonstrate:

• Site comparisons and compliance with environmental guidelines

•Correlations of water quality variables using the scatter plot matrix (SPLOM)

•Similarity of biological communities

•Trends in time series data

Site comparisons

Sometimes all that is needed to compare sites is a simple bar chart. This is where each bar of the chart represents a site's value.

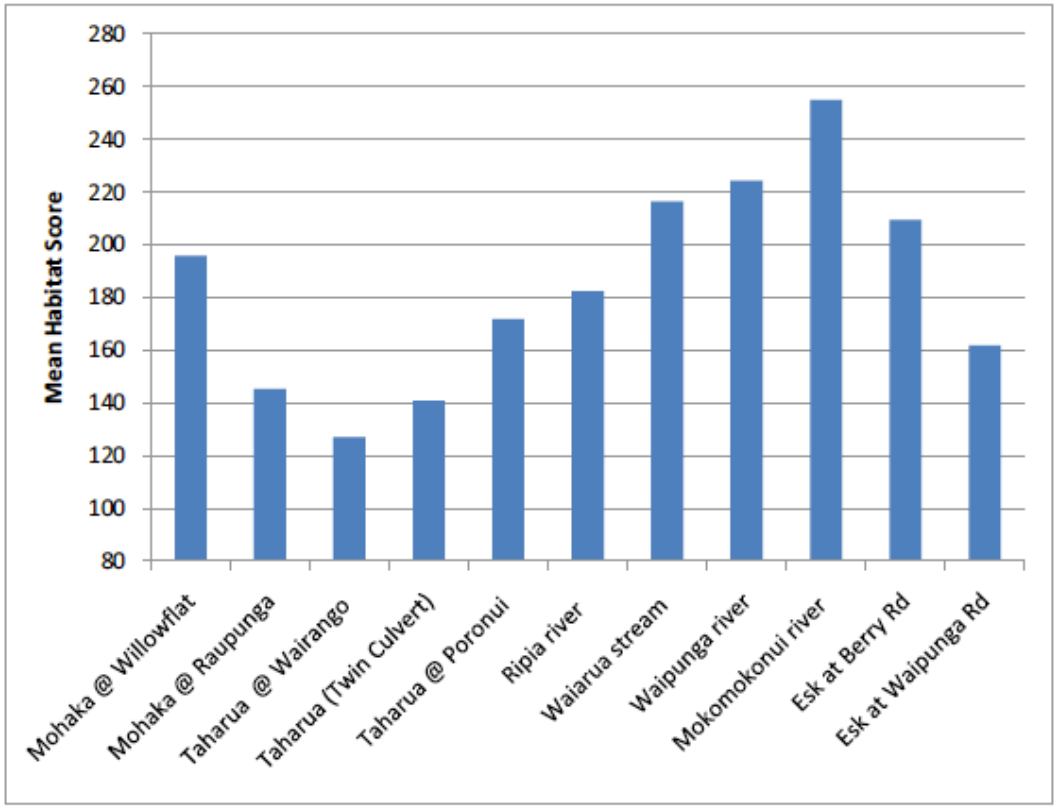


Figure 27 : Mean habitat scores for river sites within the Mohaka and Esk River catchments

Figure 27 shows that on average habitat quality is very high in the Mokomokonui River and comparatively low at the Taharua River at Wairango Station. What the diagram doesn't show us is the variability around each mean value for each site. The best tool to show such differences is the box and whisker plot.

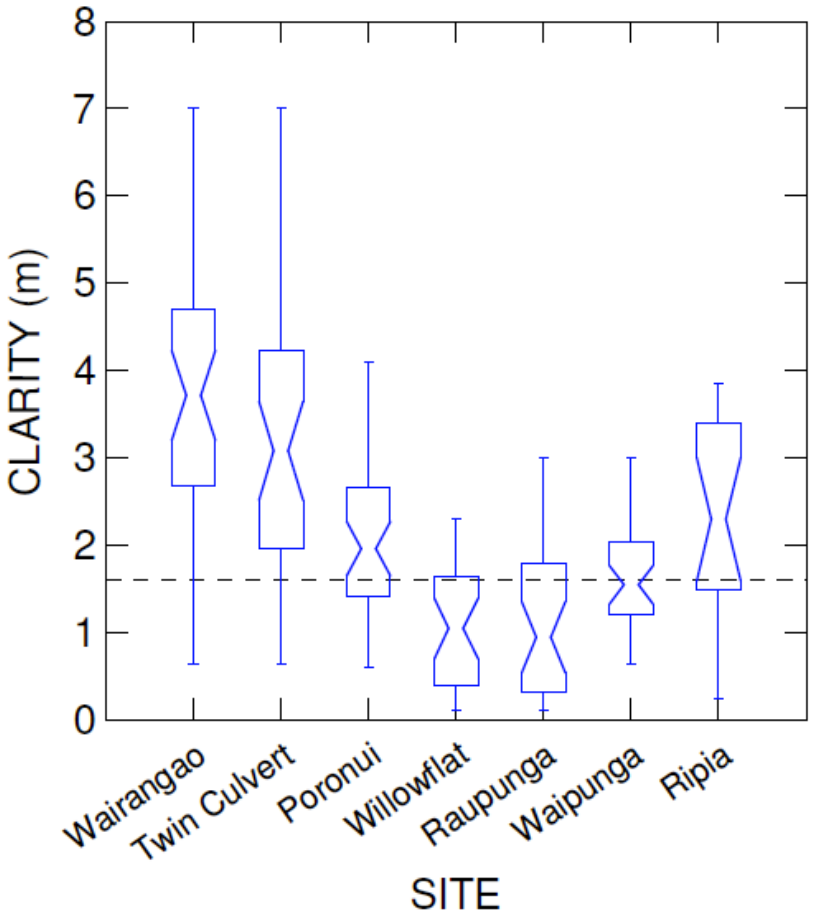


Figure 28: Box and whisker plot of water clarity measurements taken from selected sites.

Figure 28 is more informative than figure 25 because it is showing the variability of readings taken for each site. In this plot the box represents the interquartile range or range of values from the 25th to the 75th percentile of the data while the upper whisker represents values above the 75th percentile and the lower represents the range of values from the minimum to the 25th percentile. Each box in the diagram has a narrow waist which represents the median value of the data set while the sloped lines extending from the waist represent the 95% confidence interval. The confidence interval shows the range of values you estimate the true population median value to lie within (based on the sampling you have done). For example the Wairango site has a 95% confidence interval between about 3.2 to 4.2m. If the confidence intervals of any sites do not overlap then we are confident ($p < 0.05$) that the sites have statistically significant median values. For example the Poronui site has statistically significantly higher median water clarity than the Willowflat or Raupunga sites.

In figure 28 a dotted line is also shown which represents a water quality guideline for bathing (1.6m clarity). Clearly the Wairango, Twin Culverts, Poronui and Ripia sites demonstrate better compliance with this water clarity guideline than other sites. Compliance with guidelines might also be demonstrated at a site with a simple time series plot. This has the added advantage of telling the researcher not only how often compliance was breached but also when it was breached.

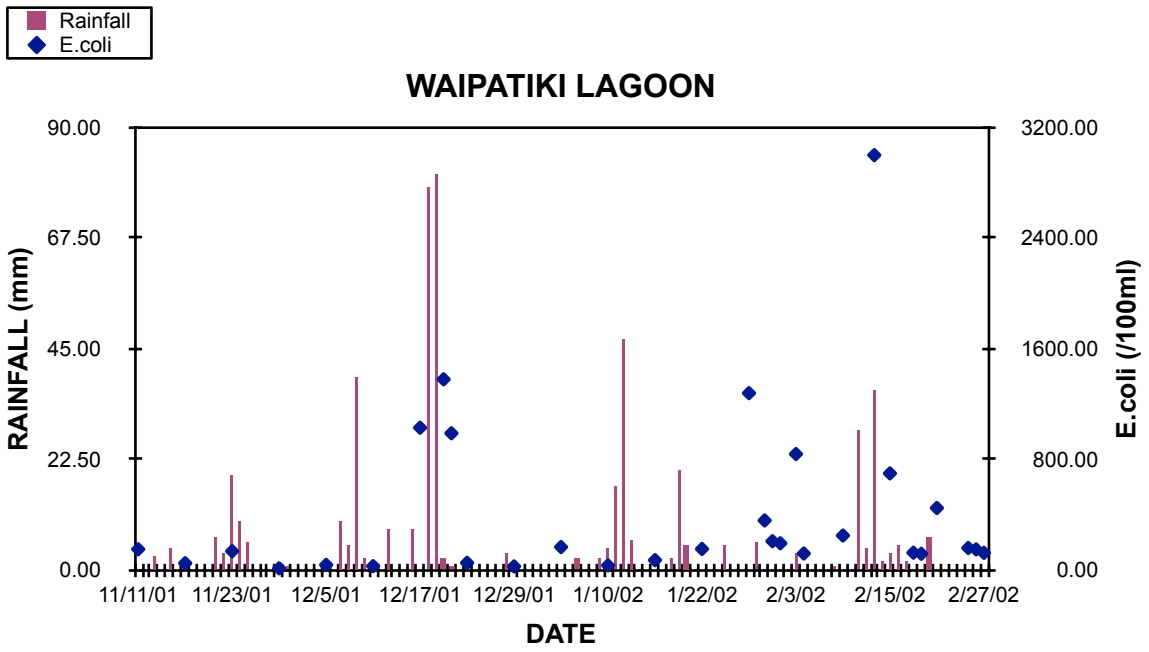


Figure 29: Rainfall and E.coli readings taken from Waipatiki Lagoon

Figure 29 gives the scientist some insight as to when E.coli concentrations (blue dots) at Wapatiki Lagoon were in compliance with the recreational water quality guideline . Also adding a second y axis of rainfall enables the researcher to see if the exceedances in E.coli were associated with high rainfall events (red bars).

Sometimes compliance with guidelines can be shown as a bar chart, however if no error bars are attached then it needs to be stated that the % compliance is based on the number of samples taken. That way you are informing the reader that the % compliance does not relate to sampling undertaken every single second of the day, rather it is based on the sampling you have undertaken.

Correlations with other variables or sites

Sometimes it may be useful to investigate the correlation of water quality variables with each other. This may give justification for using correlated values for one variable if data is missing (see page 46) or may be used to examine the influence one site may have on another (e.g. how much influence does the incoming tributary have on water quality downstream?). A quick method for examining the correlation of water quality variables can be done using a scatter plot matrix, an example follows.

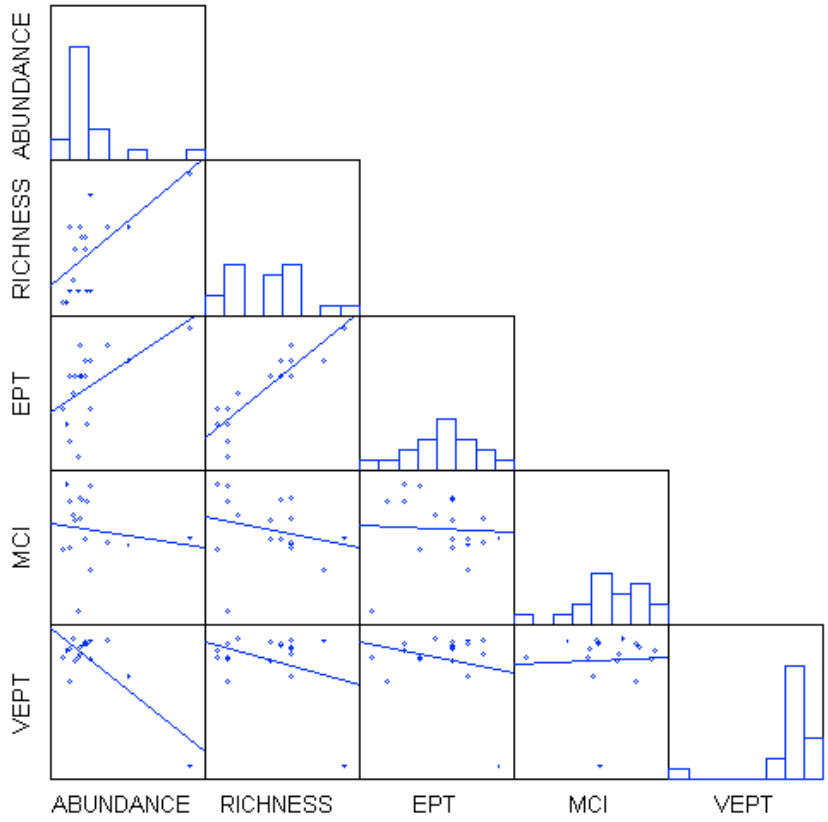


Figure 30: Scatter plot matrix of a host of water quality variables.

Figure 30 shows a scatter plot matrix of 5 biological water quality variables with each other. Data that conforms to a straight line indicates a good linear relationship while those that do not conform to a straight line show a poor linear correlation. This does not mean that the relationship is necessarily poor as other lines of fit may explain the relationship, (e.g. a quadratic relationship). The histogram of each variable is also demonstrated in the SPLOM to give the viewer an indication of any skewness. The scatter plot matrix tool is a quick and easy way to examine relationships of water quality variables or sites.

In this example we can see that the percentage of sensitive macroinvertebrate taxa (VEPT) declines as the overall abundance of macroinvertebrates increases (bottom left). Alternatively the MCI (macroinvertebrate community index) tends to increase as the percentage of sensitive macroinvertebrate taxa increases (bottom right). This is expected as the MCI is a biological index of water quality and we would expect a higher proportion of sensitive taxa corresponding to a higher MCI value. In general I wouldn't compare more than 10 variables in a scatter plot matrix, otherwise it can become too cluttered and you can lose track of what you are comparing easily.

Examining the similarity of biological communities

The most simple method of displaying the similarity of biological communities is to display the biological community composition at each monitoring site using a relative abundance stacked bar chart. This is where the percentage of the fauna pertaining to a specific taxonomic group is displayed relative to each other. An example is shown overleaf.

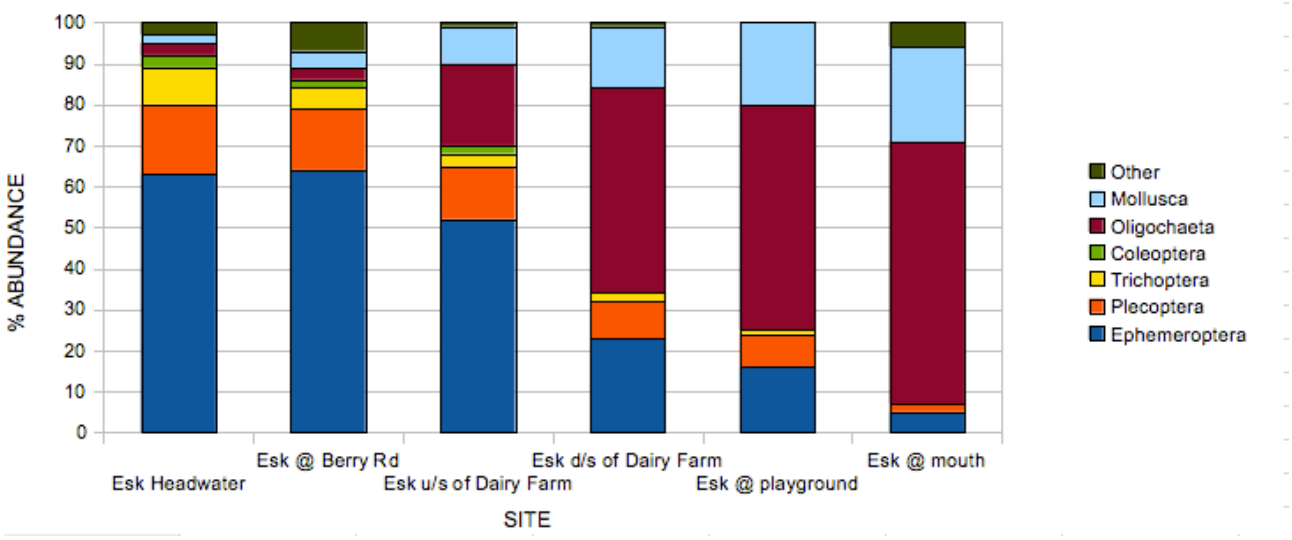
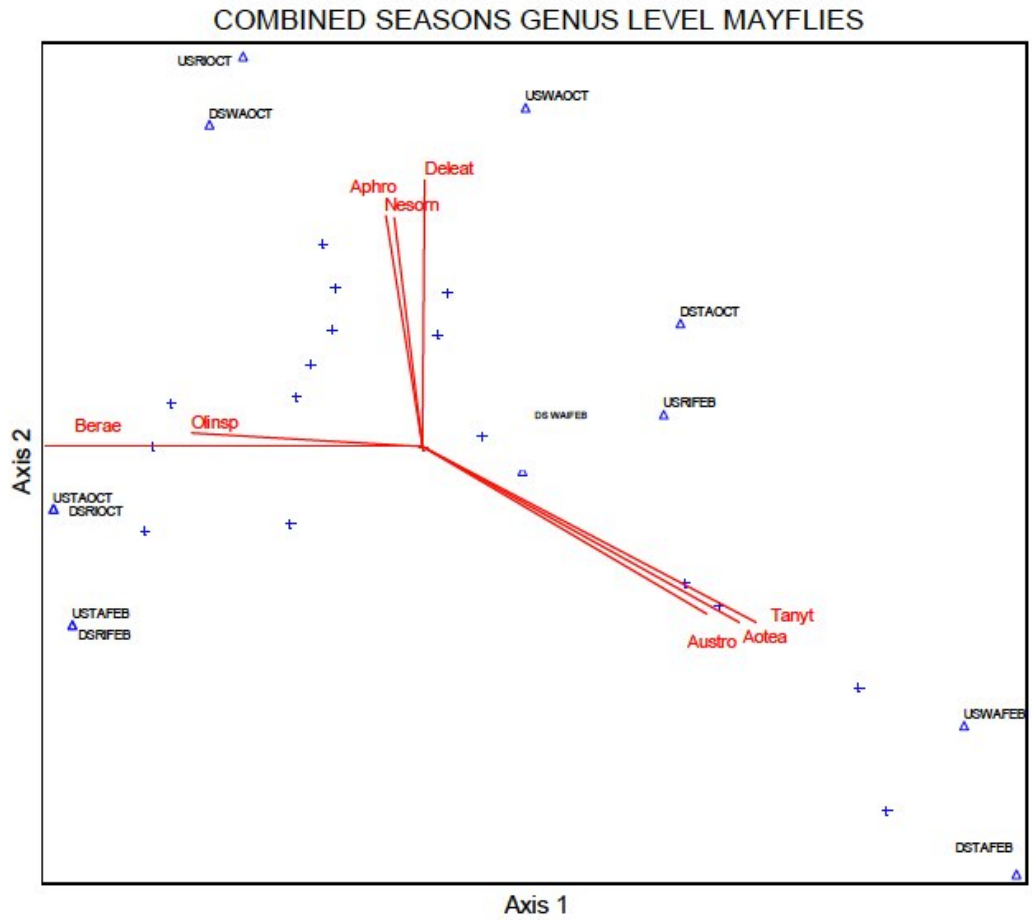


Figure 31: Relative abundance bar chart of macroinvertebrate taxonomic groups

Figure 31 shows the change in macroinvertebrate community composition as you head from the headwaters (left hand bar) to the sea (right hand bar). There is a clear pattern emerging that shows a decline in sensitive groupings (Ephemeroptera, Plecoptera and Trichoptera) and an increase in pollution tolerant groupings (Mollusca and Oligochaeta) as you move from headwaters to the sea.

Another graphical tool commonly used in biology is called a multivariate ordination. It's a bit like a map in that those sites that are in closer proximity to each other on the bi plot are more similar. Conversely those sites further apart on the ordination are more dissimilar. The ordination will also show vectors (lines going in various directions) which indicate the influence certain species may have had in distinguishing sites. To avoid clutter on the plot, site names and taxa names are normally abbreviated. An example follows over leaf.



Coefficients of determination for the correlations between ordination

```
distances and distances in the original n-dimensional space:
R Squared
Axis Increment Cumulative
1 .658 .658
2 .213 .872
```

Figure 32: A multivariate ordination of macroinvertebrate taxa sampled from the Mohaka Catchment.

Figure 32 shows the multivariate ordination (in this instance non metric multidimensional scaling) of the macroinvertebrate taxa sampled from the Mohaka River catchment. Axis 1 explains 65.8% of the variation in the data while Axis 2 explains a further 21.3% of the data. In total this means that 87.2% of the variability of the data is explained by this biplot. You will see that there are 8 taxa representing 8 vectors. Tanyt, Aotea and Austro are stretching the sites towards the low right while Berae and Olinsp are stretching the sites towards the left. In terms of axis 2, Aphro, Nesam and Deleat are stretching the sites upwards. The correlations of the taxa with the axis is also displayed as part of the output. This enables the ecologist to look for key indicator taxa that are likely to have large differences amongst the sites which gives rise to their uniqueness.

Pearson and Kendall Correlations with Ordination Axes N= 12

Axis:	1			2			3		
	r	r-sq	tau	r	r-sq	tau	r	r-sq	tau
Austro	.811	.658	.667	-.622	.387	-.286			
Colhum	-.448	.200	-.375	.458	.210	.125			
Deleat	.074	.006	.156	.786	.618	.719			
Nesorn	-.260	.067	-.188	.728	.529	.688			
Zeland	-.531	.282	-.531	-.223	.050	.094			
Aotea	.856	.732	.531	-.638	.407	-.406			
Berae	-.934	.872	-.969	.006	.000	.094			
Olinsp	-.731	.534	-.688	.178	.032	.063			
Pycno	.064	.004	.156	.579	.335	.406			
Archi	-.310	.096	-.425	.172	.030	.110			
Hydora	-.293	.086	-.281	.523	.274	.406			
Aphro	-.292	.085	-.125	.731	.534	.625			
Austro	.138	.019	.156	.035	.001	-.094			
Maori	.426	.181	.406	-.043	.002	-.094			
Orthoc	-.559	.313	-.219	-.450	.202	-.406			
Tanyt	.878	.771	.563	-.637	.406	-.438			
Oligo	.525	.276	-.016	-.571	.326	-.740			

Table 6: Correlations of taxa with axes of ordination

Table 6 shows the correlations of the macroinvertebrate taxa with the axes of the ordination. I have highlighted those tax that show a good correlation ($r > 0.6$) which are also displayed as vectors on the biplot. Note a strong negative correlation on axis 1 means a stretching to the left while a strong positive correlation on axis 1 means a stretching to the right. On axis 2 a strong positive correlation means stretching upwards versus a negative correlation means a stretching downwards.

Another interesting pattern to look at in the plot is the closeness of the sites.

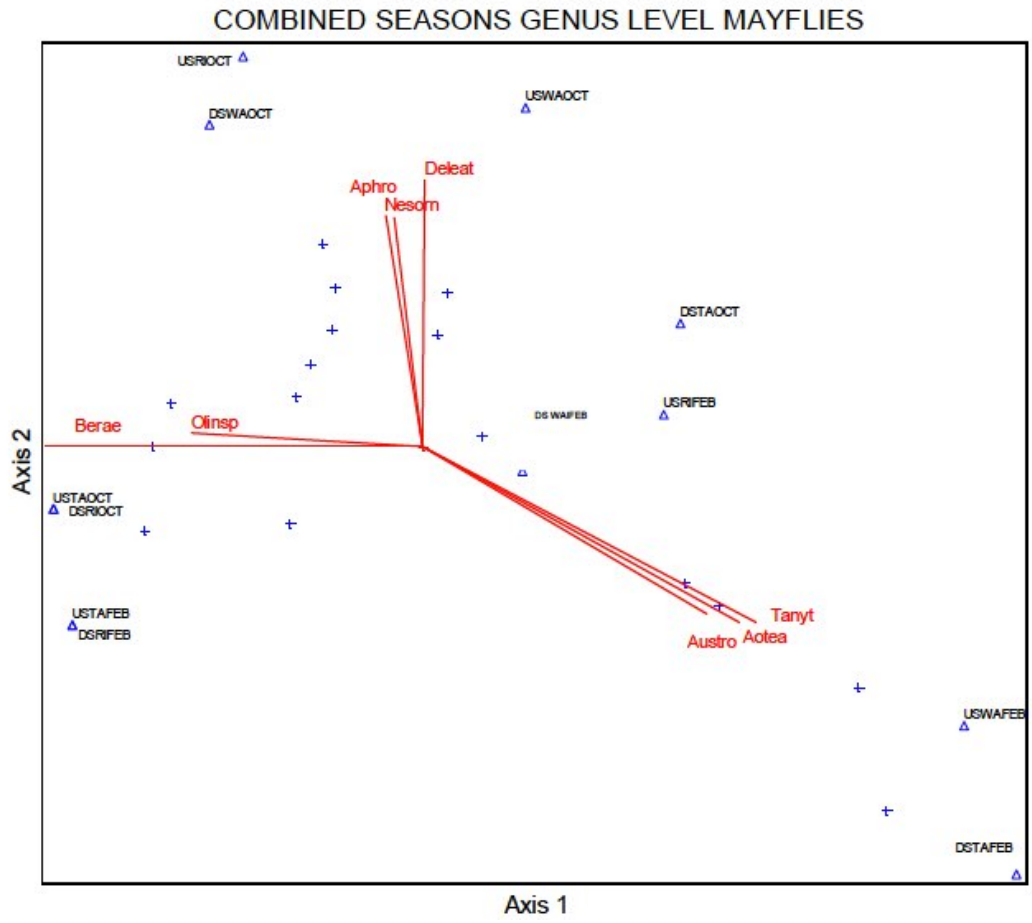


Figure 33: Repeat viewing of the ordination.

From the ordination plot you can see that some sites are very distant to each other e.g. USTAFEB (Mohaka River upstream of the Taharua confluence sampled in February) is to the left on Axis 1 while DSTAFEB (Mohaka River downstream of the Taharua confluence sampled in February) is to the lower right on Axis 1. It is likely that changes in abundance of Berae and Olinsp (Beraeoptera and Olinga species of caddisfly) were responsible for this change in macroinvertebrate community structure.

Along axis 2 you can see a large separation of USRIPOCT (Mohaka River upstream Ripia River confluence sampled in October) and DSRIPOCT (Mohaka downstream of Ripia River confluence sampled in October). It is likely that there was a significant change in the abundance of Aphro, Nesam and Deleat (Aprophila crane fly, Nesameletus may fly and Deleatidium may fly) which were responsible for this change in change in community structure.

There are many patterns within the biplot that the ecologist will examine to gain a sense of relativity in terms of changes to macroinvertebrate community structure for reporting. In the above example the ecologist can discuss changes to community structure at the sites as well as the effects the season of sampling (OCT = October, FEB = February). The changes in community structure can be statistically tested as part of the output of the ordination which can be used as a decision support tool for the ecologist when discussing changes.

Statistical Analysis And Tools To Determine Environmental Significance

So far we have examined some methods to prepare data for analysis and how to explore patterns in the data from which hypotheses may be generated. As mentioned earlier a hypothesis is a statement used to explain a phenomena which can be tested for its statistical significance. We need to keep in mind that statistics is only a decision support tool so in some instances it may not be used owing to what the scientist may know about confounding factors that may have biased a particular result. However it is important to provide an overview of some statistical methods that are used in water quality science which I shall now demonstrate.

Regression Analysis

Regression analysis is about fitting a model to data! Congratulations, you have made it this far, you are hereby a statistical modeller!

WHEN I MIGHT USE IT

- To generate synthetic E. coli values e.g. you may have a long record of Faecal coliforms dating back to the 1970s but you only have E.coli values since 2000. You are also likely to have a period of overlap as you brought E.coli monitoring into your programmes (at HBRC they allowed about a 18 month overlap). By modeling the relationship of Faecal coliforms : E. coli, we're able to generate synthetic E.coli values for years gone by. This has some pitfalls in that you can get false positive Faecal coliforms but it might be worth a crack.
- Often you might have a great data set of monthly water chemistry values but there may be a few months missing in the data series (perhaps the site wasn't monitored on these days). By modeling the correlation of flow to the water quality variable concerned you may be able to estimate (by a logistic regression) what the water quality was like on these sampling occasions. An example of this was provided on page 50.
- To describe a relationship
- To forecast trends into the future

Generally the function of a regression can be described as:

fit= a + bx

where a = the y intercept and b is the slope of the line

Furthermore an observation y can be described as:
y= fit + residual. So a residual is the difference between the fitted line and the datum point.

The following graph provides you with some understanding of the line of best fit, the slope of the line, the y intercept and calculation of a residual.

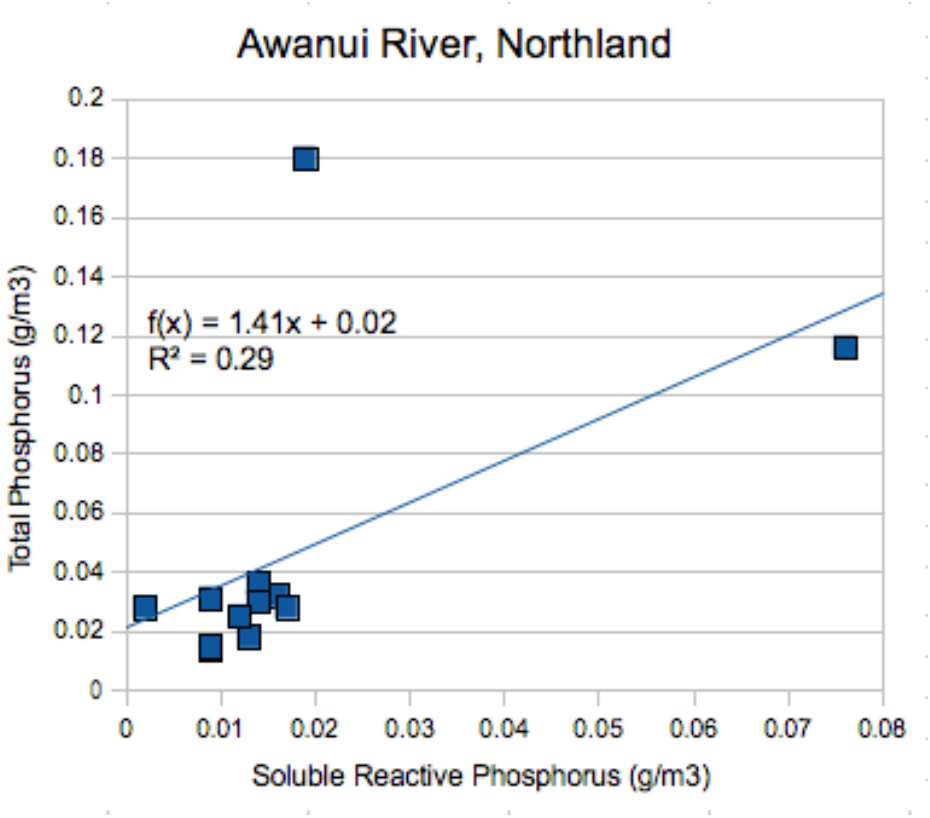


Figure 34: Regression of total phosphorus vs soluble reactive phosphorus for the Awanui river

Figure 34 shows an example of a regression of total phosphorus vs soluble reactive phosphorus. In this example the y intercept is 0.02 while the slope of the line is 1.41. The residual value for the outlying data point of 0.18 mg/m3 TP is about 0.125 as this is the difference between the value (0.18) and the modeled line of best fit (the blue line). The f(x) equation describes the relationship of TP to SRP while the R^2 value says that this model explains only 29% of the variability of the data set.

Obviously, we would like the model to fit the data as closely as possible which means that the fit will explain most of the variation in the observations. The residuals, then should be free of patterns and represent random variation about the fit. We are likely to also place further restrictions on the residuals such as requiring them to be symmetrically distributed and have constant variance over the full range of fitted values. These assumptions will be required for performing statistical tests on the regression model. For confirmatory analysis we add the requirement that the residuals should at least approximate a normal distribution.

Residual Analysis

When the residuals are plotted against fitted values, the residuals should fall into a horizontal band. If the spread tends to increase with the location, a shrinking transformation such as the square root or logarithm is needed.

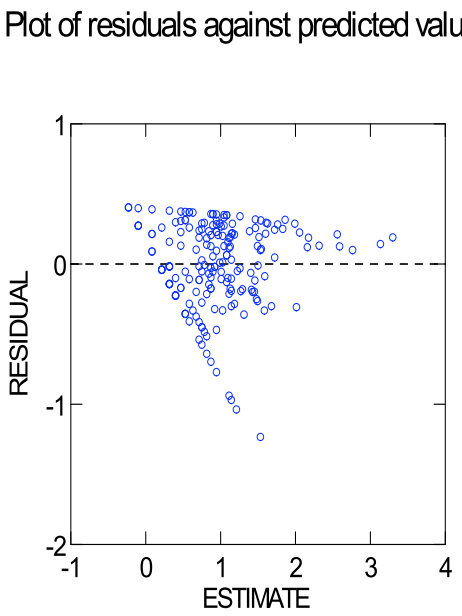


Figure 35: Residual plot

Often people forget to undertake residual analysis which could be a major mistake if they are to use the model further. Consideration should also be made of the F test and t test which are also provided in a regression analysis output.

F – the F Test is a statistical test of the goodness of fit of the line for with the null hypothesis is H_o:the model explains an insignificant proportion of the variation of the dependent variable Y.

t- the t test is a statistical test that tests for the significance of the coefficients (i.e is the slope of the line significantly different from zero) for which the null hypothesis is H_o:the slope is zero

A simple scatter plot will give you a good idea on what to expect in terms of the correlation coefficient of the regression.

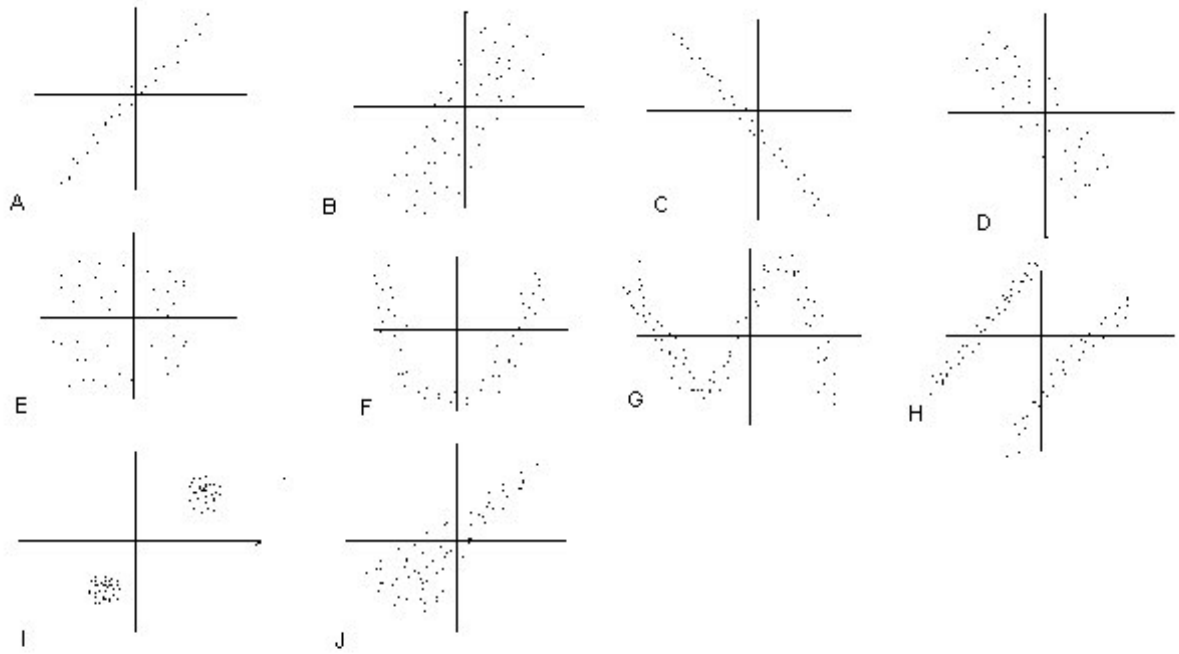


Figure 36: Possible scatter plot outcomes

Figure a provides strong evidence that a positive slope would fit the data well
Figure b indicates moderate evidence that a line of positive slope would fit the data
Figure c indicates provides strong evidence that a negative slope would fit the data well
Figure d indicates moderate evidence that a negative slope would fit the data well
Figure e f and g indicate a correlation coefficient close to zero, with e there seems no pattern but f indicates a quadratic relationship $Y = a + bx + cx^2$ and g a cubic relationship $Y = a + bx + cx^2 + dx^3$
In h there seems to be a linear relationship within each subgroup even though the correlation coefficient is close to zero – perhaps this data is better described as two individual series. On the other hand the subgroups in i do not indicate a linear relationship even though the correlation coefficient over both groups is positive.

The correlation coefficient in j is positive and a line would be a reasonable fit although the scatter diagram indicates that both Y and X have skewed distributions suggesting that a transformation is necessary.

In the following example I provide some evidence of where a mathematical transformation can drastically improve a regression model. Don't walk away from your regression just because it shows a poor relationship, it could be improved significantly simply by applying a different scale (e.g. log base 10) to the data.

Interpreting A Regression Analysis

In this example I provide a regression of E.coli against Faecal coliform bacteria.

Dep Var: ECOLICFU1 N: 12 Multiple R: 0.295 Squared multiple R: 0.087

Adjusted squared multiple R: 0.087 Standard error of estimate: 279.101

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
FCCFU100M	0.032	0.032	0.295	1	1.023	0.328

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	81474.192	1	81474.192	1.046	0.328
Residual	856868.808	11	77897.164		

*** WARNING ***
Case 187 has large leverage (Leverage = 0.958)
Case 187 is an outlier (Studentized Residual = -8.720)
Case 187 has large influence (Cook distance = 219.250)
Case 190 is an outlier (Studentized Residual = 5.171)

Durbin-Watson D Statistic 1.555
First Order Autocorrelation 0.221

Table 7: A regression analysis output

The first part of the output describes the model,

In this instance I have set the Y intercept to zero because if there are 0 Faecal coliforms then I expect 0 E.coli. So in this instance an intercept is not displayed.

The slope is 0.032, i.e E.coli = 0.032* Faecals wow what an incredibly bad model, look at the R sq value 0f 0.087, i.e this model explains 8.7 % of the total variance. Furthermore the p value of the t test (slope significance) indicates that the slope of the regression is not significantly different to zero. The F test also confirms that the amount of variation explained by the model is not statistically significant. The output delivers some warning messages about outliers having large leverage or influence on the model. Therefore I could try a shrinking transformation which squashes larger values to reduce their influence (see page 52). On this basis I would suggest trying a log:log model. In the following example I have simply applied a log 10 transformation to the data.

Dep Var: LOG10ECOLI N: 12 Multiple R: 0.971 Squared multiple R: 0.944

Adjusted squared multiple R: 0.938 Standard error of estimate: 0.513

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
LOG10FAECAL	0.818	0.06	0.971	1	13.63	0.000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	48.395	1	48.395	184.01	0.000
Residual	2.895	11	0.263		

*** WARNING ***
Case 187 is an outlier (Studentized Residual = -4.331)

Durbin-Watson D Statistic 1.646
First Order Autocorrelation 0.138

Table 8: The regression output following a log 10 transformation

Again I have set the intercept to 0 as indicated by the statement at the top 'Model contains no constant". The slope is 0.818 i.e log 10 E. coli = 0.818* log 10 Faecal coliforms. The performance of this model has improved dramatically compared to the previous one, look at the R sq value of 0.944 i.e 94.4% of the total variance is explained by this model. Also note the significance of the slope is highly significant 0.000 just means I needed to have set digital settings on the output lower but suffice to say the p value is <0.001. Therefore the null hypothesis that the slope of the regression is not significantly different to zero is rejected.

You will recall I mentioned the F test is a statistical test of the goodness of fit of the line. In this instance the p value of the F test is 0.000 which suggests we can reject the null hypothesis that the model explains an insignificant proportion of the variation of the dependent variable Y.

Note the output now only shows a warning message for case 187 which is recognised as an outlier and it does not appear to be influencing the regression model as in the previous example. I cannot stress the need for residual analysis as it gives you a good understanding whether there is any pattern as well as how much of an outlier is one residual compared to the rest. Some common residual patterns are provided overleaf.

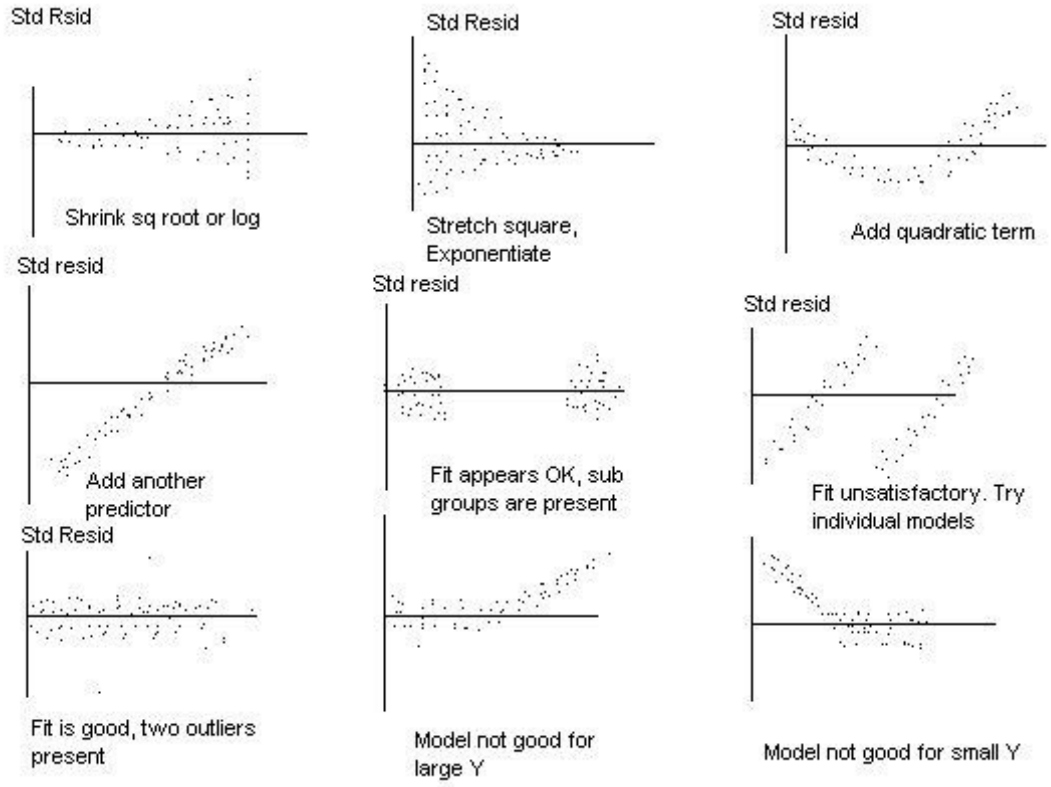


Figure 37: Some possible patterns in residuals

Remember a good model should demonstrate the following

- •an even spread of residuals above and below the fitted line
- •no evidence of autocorrelation i.e independence
- •no pattern
- •normality

What can be done if the regression is poor

- •Try using a different predictor variable
- •Transform the Y variable
- •Add an additional predictor variable e.g. Synthetic E. coli = 0.7 Faecals + 0.09 Flow +0.1 Suspended Sediments
- •Split the data - could it be that the relationship is dependent on the x value range
- •Try a non parametric alternative (Spearman rho) which is based on ranks

Before I venture into non parametric statistics, lets first look at the option of adding more predictor variables (Multiple Regression Methods)

As with simple linear regression models a number of assumptions must be made

- •The predictor variables X are known without error
- •The dependent variable Y is related to X according to a straight line (for one variable), a plane (for two variables) or a hyper plane (for more than 2 variables)
- •There is a random variability of Y about the fitted model. i.e y = fit + residual
- •For t and F statistics , The variability in Y about the line (or plane) is constant and independent of the X variables.

- The variability of Y about the line (or plane) follows a Normal distribution i.e the distribution of Y (given a certain value of x) is Normal
- The distribution of Y given X = x_i is independent of Y given X=x_j i.e there is no autocorrelation amongst the predictor variables.

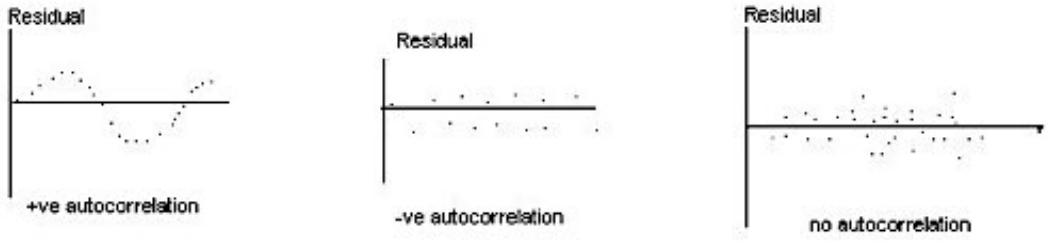


Figure 38: Patterns of autocorrelation

This last condition is often why multiple linear regression models are not really very good for water quality modelling i.e because so many water quality variables are autocorrelated to some degree (e.g. Flow, suspended sediments, indicator bacteria, clarity, chla, MCI etc). There are two types of autocorrelation illustrated above.

Multiple Linear Regression

Before undertaking a multiple regression analysis it is important to first look at a correlation matrix of the predictor variables. Correlation is usually defined as a measure of linear dependence between random variables. The correlation coefficient must lie within the interval [-1,+1] with a value of zero implying no correlation. Pearsons r is a widely accepted measure of the strength of a linear relationship, especially in linear regression.

The autocorrelation matrix enables us to explore the correlations between the predictor variables (X1, X2, X3, X4...., in this case % Dairying, Median Flow, SU/HA, Stream Order, Porosity) and the dependent variable Y (median nitrate N). The matrix also enables us to investigate the correlations that may exist amongst the predictor variables, if the matrix shows a high degree of intercorrelation amongst the predictor variables it suggests that not all of these variables are needed to explain Y(median nitrate N).

	MEDIANNITRA	%DAIRYING	MEDIANFLOW	SUHA	STREAMORDER
MEDIANNITRA	1.000				
%DAIRYING	**0.935**	1.000			
MEDIANFLOW	-0.732	-0.595	1.000		
SUHA	**0.929**	0.991	-0.615	1.000	
STREANORDER	-0.562	-0.349	0.891	-0.375	1.000
POROSITY	0.555	0.702	0.005	0.696	0.240

Table 9: The autocorrelation matrix

This correlation matrix shows that % Dairying and SU/HA are more strongly correlated with median nitrate N, while the other variables are less correlated to median nitrate N. The output also shows that there is autocorrelation amongst the predictor variables especially SU/Ha and % Dairying. This

indicates that perhaps only one of these predictors is needed to adequately describe the variability in the Y variable (nitrate N)

Multiple linear regression is generally undertaken in a stepwise fashion, here I describe 3 types.

Forward Stepwise

Forward stepwise regression – here we place the predictors sequentially into the model in the decreasing order of their additional SSR. Thus the first variable to enter the model is the one that is most highly correlated with the response variable Y. The second best predictor to be placed into the model is the one that has the next largest additional SSR and so on.

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
Out		Part. Corr.						
2	%DAIRYING	0.935	.	.	1.00000	1	48.587	0.000
3	MEDIANFLOW	-0.732	.	.	1.00000	1	8.061	0.025
4	SUHA	0.929	.	.	1.00000	1	44.253	0.000
5	STREAMORDER	-0.562	.	.	1.00000	1	3.238	0.115
6	POROSITY	0.555	.	.	1.00000	1	3.121	0.121

Table 10: The initial multiple linear regression output

The initial regression shows that % Dairying has the highest correlation so this is the most strongly correlated variable to the Y variable Nitrate N. Therefore this is the first variable to enter the model.

```
Forward stepwise with Alpha-to-Enter=0.900 and Alpha-to-Remove=0.010
Step # 1 R =  0.935 R-Square =  0.874
Term entered: VDAIRYING
```

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	%DAIRYING	0.026	0.004	0.935	1.00000	1	48.587	0.000
Out		Part. Corr.						
3	MEDIANFLOW	-0.615	.	.	0.64591	1	3.641	0.105
4	SUHA	0.051	.	.	0.01725	1	0.016	0.904
5	STREAMORDER	-0.711	.	.	0.87846	1	6.131	0.048
6	POROSITY	-0.399	.	.	0.50741	1	1.137	0.327

```
Step # 2 R =  0.968 R-Square =  0.938
Term entered: STREAMORDER
```

Table 11:Step 2 of the forward stepwise regression output

Once % Dairying has entered the model, the next best predictor variable to enter the model is stream order

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	%DAIRYING	0.024	0.003	0.841	0.87846	1	59.866	0.000
5	STREAMORDER	-0.154	0.062	-0.269	0.87846	1	6.131	0.048
Out		Part. Corr.						
3	MEDIANFLOW	0.102	.	.	0.11453	1	0.053	0.827
4	SUHA	-0.174	.	.	0.01627	1	0.156	0.710
6	POROSITY	0.242	.	.	0.23980	1	0.312	0.600

```
Step # 3 R =  0.935 R-Square =  0.874
```

Table 12: Step 3 of the forward stepwise regression output

And finally the model is recommended as no other predictor variables show significant correlation to Y.

```
Analysis of Variance
```

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	7.085	1	7.085	32.650	0.002
Residual	1.085	5	0.217		

```
*** WARNING ***
Case            2 is an outlier        (Studentized Residual =        2.592)

Durbin-Watson D Statistic          2.175
First Order Autocorrelation       -0.106
```

Table 13: Final output of forward stepwise regression model

So the final model is Median Nitrate N = 0.024*%Dairying - 0.154*Stream Order for which the R^2 value is 0.874.

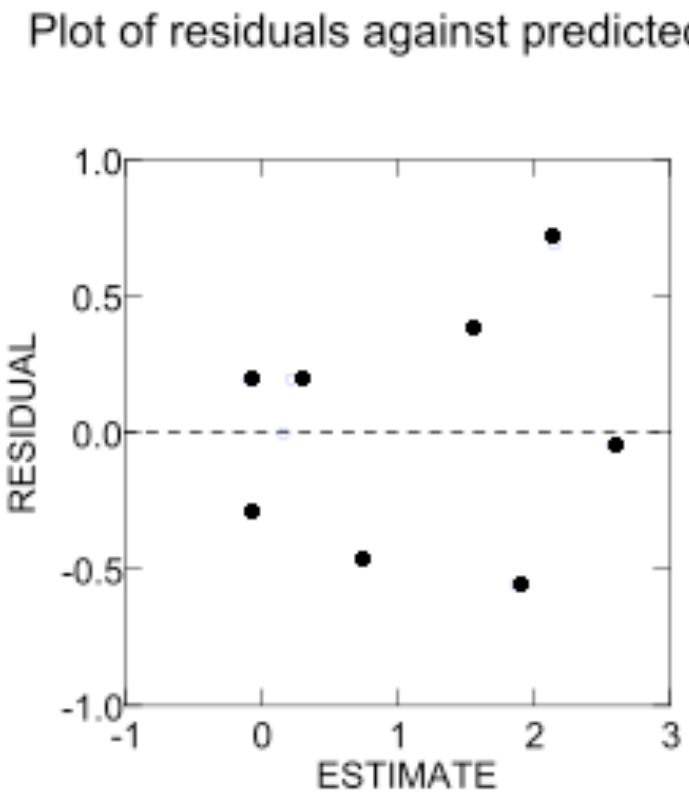


Figure 39: Plot of residuals against predicted values
There is a good spread of residuals above and below the line which is great.

Backward Elimination

With backward elimination, the model starts with all the predictor variables first and then sequentially takes out variables that are least correlated with Y.

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	%DAIRYING	0.034	0.030	1.188	0.01655	1	1.278	0.340
3	MEDIANFLOW	0.003	0.495	0.003	0.10400	1	0.000	0.995
4	SUHA	-0.324	0.723	-0.497	0.01487	1	0.201	0.685
5	STREANORDER	-0.214	0.201	-0.375	0.14755	1	1.132	0.365
6	POROSITY	0.076	0.145	0.157	0.20356	1	0.274	0.637
Out		Part. Corr.						
	none							

Dependent Variable MEDIANNITRA
Minimum tolerance for entry into model = 0.000000

Table 14: Initial model with all predictor variables

Backward stepwise with Alpha-to-Enter=0.900 and Alpha-to-Remove=0.010
Step # 1 R = 0.972 R-Square = 0.945
Term removed: MEDIANFLOW

Median Flow is removed first because this variable shows the lowest correlation with Y (as indicated by the very low F value)

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	VDAIRYING	0.034	0.026	1.188	0.01662	1	1.710	0.261
4	SUHA	-0.324	0.620	-0.498	0.01516	1	0.274	0.628
5	STREANORDER	-0.213	0.111	-0.373	0.36475	1	3.694	0.127
6	POROSITY	0.076	0.120	0.157	0.22343	1	0.404	0.560
Out		Part. Corr.						
3	MEDIANFLOW	0.004	.	.	0.10400	1	0.000	0.995

Table 15: Backward stepwise regression model with Median Flow removed
Next SU/HA is removed

Step # 2 R = 0.970 R-Square = 0.941

Term removed: SUHA

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	%DAIRYING	0.021	0.006	0.731	0.22355	1	10.177	0.024
5	STREANORDER	-0.193	0.096	-0.337	0.41515	1	4.030	0.101
6	POROSITY	0.060	0.107	0.124	0.23980	1	0.312	0.600
Out		Part. Corr.						
3	MEDIANFLOW	0.039	.	.	0.10603	1	0.006	0.942
4	SUHA	-0.253	.	.	0.01516	1	0.274	0.628

Table 16:Backward stepwise model with stocking units/ ha removed

Step # 3 R = 0.968 R-Square = 0.938

Next porosity is removed

Term removed: POROSITY

	Effect	Coefficient	Std Error	Std Coef	Tol.	df	F	'P'
In								
1	Constant							
2	%DAIRYING	0.024	0.003	0.841	0.87846	1	59.866	0.000
5	STREANORDER	-0.154	0.062	-0.269	0.87846	1	6.131	0.048
Out		Part. Corr.						
3	MEDIANFLOW	0.102	.	.	0.11453	1	0.053	0.827
4	SUHA	-0.174	.	.	0.01627	1	0.156	0.710
6	POROSITY	0.242	.	.	0.23980	1	0.312	0.600

Step # 4 R = 0.935 R-Square = 0.874

Table 17: Backward stepwise model with porosity removed

Finally it summarises the recommended model

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	5.782	1	5.782	48.587	0.000
Residual	0.833	7	0.119		

Table 18: Model summary

```
*** WARNING ***
Case            1 has large leverage    (Leverage =        0.928)
Case            8 is an outlier         (Studentized Residual =        4.151)

Durbin-Watson D Statistic          2.484
First Order Autocorrelation       -0.252
```

So the final equation is Median Nitrate N = 0.024*%Dairying - 0.154*Stream Order

Conventional Stepwise Regression

Conventional stepwise regression procedures stop and check whether variables either in the model are or not are the best combination for that step (i.e for a given number of predictors). If the contribution of any variable inside the model is found to be insignificant it is removed form the model. If the contribution of any variable outside the model is found to be significant, it is added to the model.

Analysis of variance

I would often use a simple one way analysis of variance to compare the mean value of a water quality variable at various sites. The test first answers the question of whether there are significant differences amongst the sites and then allows for pairwise comparisons of all site combinations to be made so that you can then determine which sites are statistically significantly different to each other.

Understanding the ANOVA output

First lets look at a one way ANOVA output.

```
Data for the following results were selected according to:
      season$="Summer"

Effects coding used for categorical variables in model.


Categorical values encountered during processing are:
SITE$ (6 levels)
   DS Ripia , DS Taharua , DS Waipunga, US Ripia , US Taharua , US Waipunga


Dep Var: QMCIVALUE   N: 36   Multiple R: 0.901   Squared multiple R: 0.812

Analysis of Variance
```

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
SITE$	42.120	5	8.424	25.905	0.000
Error	9.756	30	0.325		

Table 19: ANOVA output summary

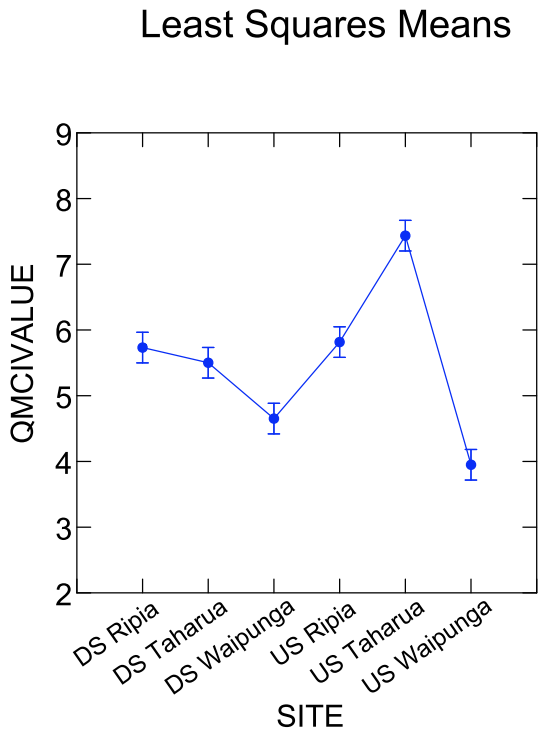
$R^2 = 42.12/(42.12+9.756) = 0.812$

$R = \sqrt{0.812} = 0.901$

The F statistic = MS Site$/MS Error

= 8.424/0.325 = 25.905 for which the p value suggests that there is a significant difference between the mean QMCI values. This is clearly evident in the plot that follows (least squares mean) which shows the mean QMCI value for each site with a 95% confidence interval for each.

The confidence interval displays the upper and lower bounds of where the true mean of the population is likely to be, given the amount of sampling that has been undertaken. The more sampling you do the smaller the confidence intervals become as you are gaining a better estimation of the population mean.




```
Durbin-Watson D Statistic          2.495
First Order Autocorrelation       -0.251
```

Figure 40: Least squares means plot

```
COL/
ROW SITE$
  1  DS Ripia
  2  DS Taharua
  3  DS Waipunga
  4  US Ripia
  5  US Taharua
  6  US Waipunga
Using least squares means.
Post Hoc test of QMCIVALUE
Using model MSE of 0.325 with 30 df.
```

Matrix of pairwise mean differences:

	1	2	3	4	5
1	0.000				
2	-0.231	0.000			
3	-1.081	-0.850	0.000		
4	0.084	0.315	1.165	0.000	
5	1.702	1.933	2.783	1.618	0.000
6	-1.783	-1.552	-0.702	-1.867	-3.485

Table 20: Table of pairwise differences

This first table simply shows the matrix of pairwise differences e.g. the difference between the mean of site 1 and the mean of site 2 is -0.231, the difference between the mean of site 3 and site 1 is -1.081 etc.

Tukey HSD Multiple Comparisons.

Matrix of pairwise comparison probabilities:

	1	2	3	4	5
1	1.000				
2	0.980	1.000			
3	**0.028**	0.133	1.000		
4	1.000	0.928	**0.015**	1.000	
5	**0.000**	**0.000**	**0.000**	**0.000**	1.000
6	**0.000**	**0.001**	0.299	**0.000**	**0.000**

	6
6	1.000

Table 21: Matrix of pairwise comparison probabilities

The second table shows which site by site comparisons are statistically significantly different. Obviously not all site by site comparisons are statistically significantly different to each other but most certainly are.

Two Way ANOVA

A Two way ANOVA might be used if I wanted to see whether the differences between site mean values were affected by the influence of another factor. In this example we examine the effects season has on differences between mean site QMCI values

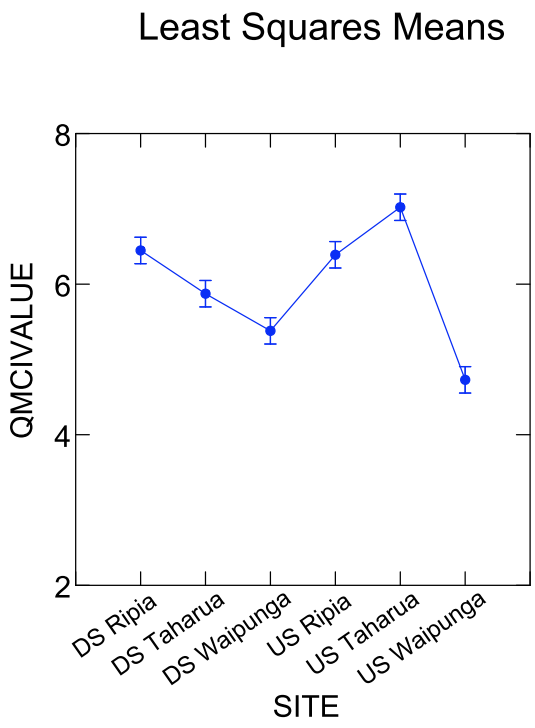
```
Effects coding used for categorical variables in model.

Categorical values encountered during processing are:
SITE$ (6 levels)
   DS Ripia , DS Taharua , DS Waipunga, US Ripia , US Taharua , US Waipunga
SEASON$ (2 levels)
   Summer, Winter

Dep Var: QMCIVALUE   N: 72   Multiple R: 0.869   Squared multiple R: 0.755

Analysis of Variance
```

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
SITE$	40.872	5	8.174	22.134	0.000
SEASON$	15.173	1	15.173	41.085	0.000
SITE$*SEASON$	12.270	5	2.454	6.645	0.000
Error	22.159	60	0.369		

Table 22: ANOVA summary

$R^2=(40.872+15.173+12.270)/(40.872+15.173+12.270+22.159) = 68.315/90.474 = 0.755$
$R = \sqrt{0.755} = 0.689$

The p values indicate that the F value for Site$, Season$ and Site$* Season$ are highly significant. So what does that mean?
It means that there are significant differences between sites!
There are also significant differences between seasons!
And lastly the interaction term (site* seasons) is significant meaning that the differences between sites depends on the season!



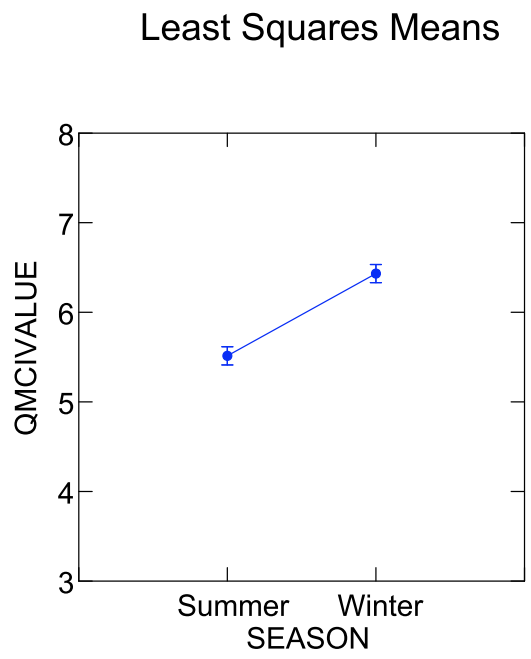
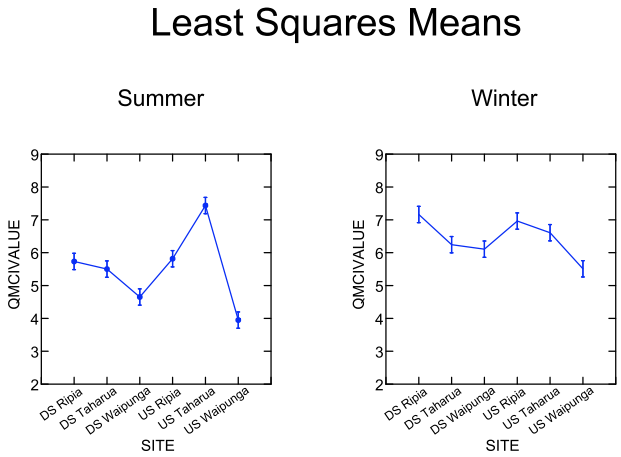


Figure 41: Least squares means plot for sites and seasons

We can see that the Winter season has higher QMCI values. This is quite likely as benthic aquatic communities are often subject to less stress during winter owing to cooler water temperatures, less periphyton growths and regular freshes keeping the stream bed clean.




```
*** WARNING ***
Case           61 is an outlier        (Studentized Residual =       -4.565)

Durbin-Watson D Statistic          2.416
First Order Autocorrelation       -0.208
```

Figure 42: Least squares means plots for seasons

Sometimes you may be faced with having to compare the mean values of two independent populations but the sample size taken from each population is different e.g. supposing you were comparing water clarity measurements taken from two monitoring sites e.g. Clive River for which the sample size was 18 (n=18) and the Karamu Stream for which the sample size was 27 (n=27). In this instance it is best to use a 2 sample t test. The t test is the preferred method if the sample sizes of the populations are different.

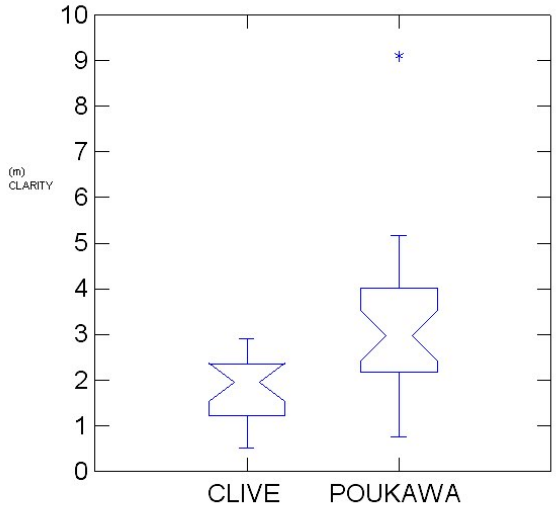


Figure 43: Water Clarity box and whisker plots for the Clive River and Poukawa Stream

Paired samples t test on CLIVE vs POUKAWA with 18 cases
Mean CLIVE = 1.804
Mean POUKAWA = 3.320
Mean Difference = -1.516 95.00% CI = -2.533 to -0.499
SD Difference = 2.045 t = -3.145
df = 17 Prob = 0.006
Bonferroni Adjusted Prob = 0.006

The null hypothesis for the t test is that the mean water clarity for the two sites is the same. The above output shows that the probability of this being so is very small (0.006) therefore we reject the null hypothesis of the sites having the same water clarity. The Bonferroni adjusted probability gives the adjusted probability based on sample size, it also confirms that the mean water clarity readings from these sites are statistically significantly different.

Calculating a percentile

Percentiles are often used in setting standards for compliance monitoring. While it is easy to ask the computer to calculate a percentile value for you BE CAUTIOUS AS TO WHAT THE SOFTWARE IS DOING. Do not use excel as it can give erroneous results. You are better off to use a reputable statistics package for this process.

Provided you know the mean and standard deviation of your data, you are in a position to calculate percentiles using the Z statistic. As you probably know the area under the normal distribution bell shape curve = 1

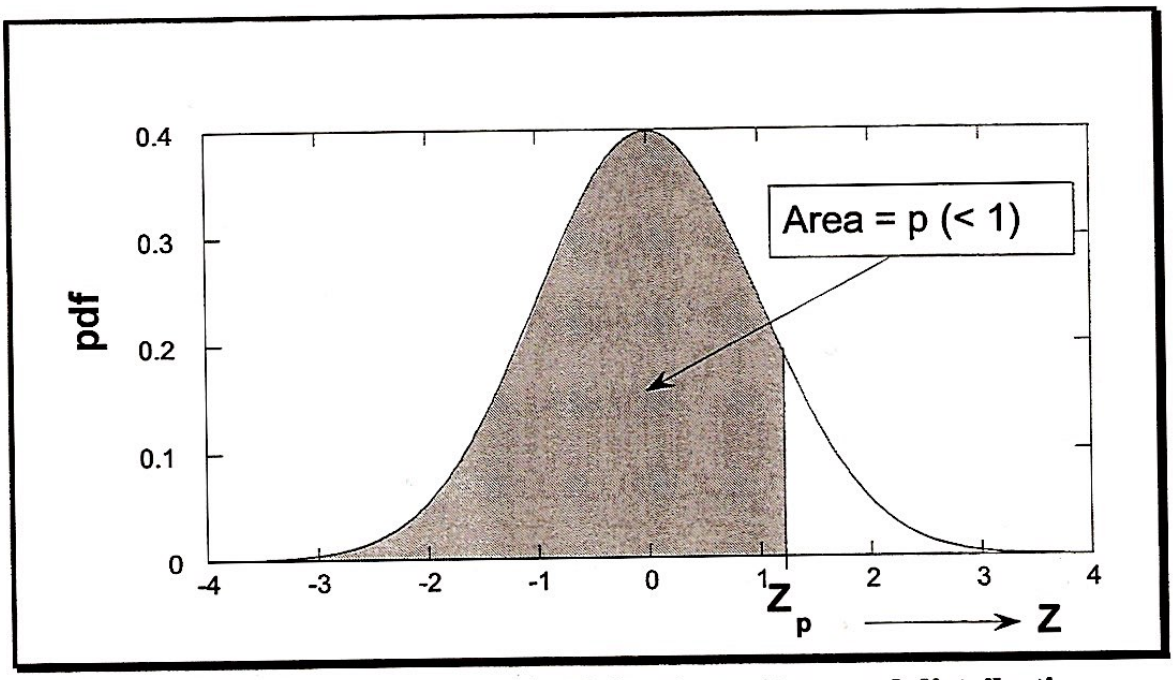


Figure 44: Probability density function (McBride 1999)

The area to the left of Zp is p (in this instance your percentile value). Using this distribution we only need to look up one standard table and do some simple calculations using the standard normal table. This table will tell you what proportion of the population are within z standard deviations of the mean. Supposing we had the following data set for nitrate nitrogen concentrations in groundwater beneath a dairy farm.

Date	Nitrate nitrogen (mg/l)	Deviation From Mean	Squared
13/01/2009	**0.87**	-0.70	0.49
17/02/2009	**1.05**	-0.52	0.27
14/03/2009	**1.59**	0.02	0.00
18/04/2009	**2.32**	0.75	0.56
19/05/2009	**0.96**	-0.61	0.38
20/06/2009	**0.88**	-0.69	0.48
15/07/2009	**1.74**	0.17	0.03
18/08/2009	**1.32**	-0.25	0.06
19/09/2009	**1.55**	-0.02	0.00
13/10/2009	**0.99**	-0.58	0.34
11/11/2009	**1.43**	-0.14	0.02
15/12/2009	**1.88**	0.31	0.09
17/01/2010	**2.05**	0.48	0.23
19/02/2010	**2.58**	1.01	1.02
21/03/2010	**2.67**	1.10	1.20
15/04/2010	**2.01**	0.44	0.19
16/05/2010	**1.77**	0.20	0.04
18/06/2010	**1.65**	0.08	0.01
14/07/2010	**1.12**	-0.45	0.20
13/08/2010	**1.02**	-0.55	0.31
Mean	**1.57**		
Sum			5.92
Variance			0.30
Std Deviation			0.54

Table 23: Some nitrate nitrogen data

In the above example, the variance is the sum of the squared deviations of each datum from the mean divided by the number of samples (in this case 5.92/20) and the standard deviation is simply the square root of the variance (in this case sqrt of 0.296 = 0.544). So we know that the mean of the data is 1.57 mg/l with a standard deviation of 0.54. Supposing we wanted to set a compliance at the 90 percentile value for this data set as the current data set represented the farm pre intensification. Therefore using the table overleaf we could calculate the 90th percentile as follows:

$X = \mu + \sigma Z$

or in English!

the 90th percentile value = mean + standard deviation * Z

so from the table below we can see that an area of 0.9 equates to a z value of 1.29 standard deviations. The table is read by using the rows to find the first digit, and the columns to find the second digit of a Z-score.

Therefore the 90th percentile = 1.57+(0.54*1.29) = 2.27 mg/l nitrate nitrogen.
Note for the above data set excel estimates the 90th percentile to be 2.346 while Statistica estimates the 90th percentile to e 2.45 because they are using different percentile methods.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999

Table 24: Z score statistics and corresponding percentile areas of the unit normal distribution.

Once you have calculated your percentile standard you need to estimate the required sample size to confidently confirm that a breach of the standard has occurred.

Estimating The Correct Sample Size For A Compliance Programme

In order to test for compliance with percentile standards, you need to have an understanding of the binomial distribution. It applies when a collection of random samples can have one of two outcomes - e.g. pass or fail, exceed or not exceed. Then the distribution of those outcomes always follows the binomial distribution, regardless of the distribution that the data happen to follow.

Say you are interested in testing whether an effluent concentration exceeds a certain limit for more than 5% of a year. Let's suppose that the effluent was borderline i.e it exceeded the limit for exactly 5% of the year. Then the probability of any individual random sample exceeding the limit is 5%. Say 20 weekly samples are taken to assess compliance and that a breach of the effluent standard will be inferred if there is more than 1 exceedance of the limit in those 20 samples. One should ask, what then is the risk of falsely inferring that a breach has occurred? Say 2 or more of the 20 samples exceeded the limit, and so a breach of standard would be inferred. What if in fact over the course of the year the effluent was borderline? (i.e. the limit was in fact exceeded for a little less than 5% of the year and so to infer a breach based on the 20 samples is an error). Tables of the binomial distribution can be used to calculate the risk. The first two columns of a cumulative binomial probability table has n (the number of samples), e, (the number of exceedances) and then the probabilities. The following table is derived from a very cool website http://stattrek.com/Tables/Binomial.aspx

n	e	p=	.05	.10	.15
20	0		.3585	.1216	.0388
	1		.7358	.3917	.1756
	2		.9245	.6769	.4049
	3		.9841	.3917	.1755

So Pr (0 or 1 exceedances of the limit) = 0.7358. The required probability is therefore (P 2 or more exceedances of the limit) = 1-0.7358 = 0.2642 or 26%! Most people are surprised to find such a high risk. Of course the risk is less if the effluent was not borderline, but that may not mollify the dischargers concerns. Obviously we want our risk of 26% to be less than 5% to be confident that > 5% exceedances have occurred so we have two ways of doing this

1. increase the number of allowable exceedances, you'll see in the above table that even if the discharger had incurred 2 exceedances, Pr = 0.9245 so 1-0.9245 = 0.075 which is still greater than 0.05. It is likely that 3 exceedances of the twenty samples would give us enough confidence that the 5% of samples threshold had been exceeded. (Pr = 0.9841, so 1-0.9841 = 0.0159 which is < 0.05
2. increase the number of samples for the compliance analysis e.g. for n = 30 and exceedances = 4

n	e	p=	.05	.10	.15
30	4		.9843	.8245	.5244

In this case if we had n = 30 and 4 exceedances Pr = 0.9843, the required probability = 1-0.9843 = 0.0157 which is < 0.05. Doing these calculations is a good way of determining the desired sample size to test compliance with percentile standards.

Determining Appropriate Sample Size For Resource Surveys

When designing an environmental monitoring programme, you need to be sure that the error in your estimation of the population does not exceed the degree of environmental change you are trying to determine.

Biggs (2000) gives a very detailed account of how to estimate the number of replicate samples needed for general resource surveys using the following equation provided by Zarr (1996):

$$n = \frac{s^2 t^2}{d^2}$$

where s is the standard deviation of the preliminary data, t =$t_{a(2)}$, n-1 the two tailed critical value of the Student's t distribution with v = n-1 degrees of freedom and α=0.05, and d is a preselected half width of the desired confidence interval of the sample mean. So let's give this a go!

Suppose we want to know the mean chlorophyll a concentration of periphyton of a stream. We might want to estimate this mean biomass with a precision that the sample mean is to lie within an interval i.e +/- 20% of the mean (i.e the sample mean is to be within +/- 20% of the population mean at p = 0.05). How many samples would be required? Preliminary survey data of chlorophyll a from one reach were tested and found to be approximately normally distributed. The mean of these data was 267.5 mg/m^2 chlorophyll a (n=10), with a standard deviation of 86.4 and the detectable difference d= 0.2*267.5 = 53.5. We would then consult a t table to find the t statistic used in the equation.

df\ p	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
1	0.32492	1	3.077684	6.313752	12.7062	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.5407	5.84091	12.924
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.94318	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.306	2.89646	3.35539	5.0413

df\ p	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.36343	1.795885	2.20099	2.71808	3.10581	4.437
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.681	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.34503	1.76131	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.75305	2.13145	2.60248	2.94671	4.0728

Table 25: The t table

We then need to start the process of iteration by guessing the number of samples that might be required and using this as a basis to select a critical value for the t distribution. We start by guessing that 15 samples would be required (for the iterations it is better to initially overestimate the number required). Therefore , the critical value for the t distribution is $t_{0.05,\ 15\text{-}1}$ = 2.145. Inserting our values for s, t and d in the above equation we get

$n = s^2t^2/d^2 = 86.4*2.145^2/53.5^2 =$ 12.00 samples

We then iterate the equation again to see if we can get close to 12 by using a smaller starting value than 15. If we use 12 samples as the starting point we insert a critical value for the t distribution for n=12-1 degrees of freedom (=2.201). This iteration then gives

$n= 86.4^2*2.201^2/53.5^2 = 12.63$ samples

This is close enough to the first estimate of 12 to conclude that we probably need 12-13 replicates to enable us to be 95% confident that our sample mean will be within +/- 20% of the population mean.

Non Parametric Methods

In this section I outline an alternative to using transformations as a method of preparing the data for hypothesis testing. This approach involves replacing the actual observed data values by their ranks i.e the smallest data value is replaced by 1, the second smallest replaced by 2 and so on up to the largest data value being replaced by n. If there are several data with the same value then they are all assigned the average rank they would have gotten if they received a tiny random increment before being placed in order e.g. if the 3rd, 4th, 5th and 6th smallest data values are all the same then the four corresponding points are all assigned rank 4.5 which is the average i.e 3+4+5+6/4. We then analyse the ranks as if they are the original data.

In principle then we could simply apply the usual data analysis techniques (t -tests, ANOVA, correlation coefficients etc) to the W=rank(Y) data and quote the p value accordingly.
We don't usually do this for small data sets however for large samples it would be quite a reasonable approach as the distribution of W values is symmetrical and therefore the usual data analysis methods - relying W being Normally distributed will work pretty well. The main difference to the standard hypothesis tests would be that they would need to be expressed in terms of medians rather than means.

A two group hypothesis test would be based on computing W (the ranks of Y) for all the n data values together and then comparing the mean of the n_1 ranks in the first group of observations to the mean for the n_2 ranks for the second group of observations. If there is no difference in population medians for the two groups then we should not be able to reject the hypothesis that the mean W values are the same: hence the hypothesis test.

In practice there are a couple of complications

1. For small samples the distribution of W=rank(Y) is not Normal because it is discrete taking only integers or averages of integers. Fortunately statisticians have long since worked out the distribution of W for many simple situations including two sample tests, one way ANOVA, two way ANOVA with balanced numbers, correlation coefficients (the correlation between rank Y and rank X) and some others. So the exact distributions can be used for hypothesis tests.
2. These exact distributions of W=rank(Y) usually depend on the assumption of no ties i.e no equal ranks. Since ties often occur in practice then we need to use methods that are modified or corrected to handle ties.

Two Dependent Samples

For the nonparametric equivalent of a one sample t-test for H_O: $u=u_{0,}$ we use the Wilcoxon Ranked Sign Test for $H_o = \eta=\eta_0$ where η (greek letter eta) is the population median. This testing is most appropriate when the two samples are dependent e.g. upstream and downstream of a discharge to water, ideal for analysis of a STP discharge
so let's give this a go with some fictitious log 10 Faecal coliform data.

[1]

Date	U/S Discharge	D/S Discharge	Difference	Rank Difference	Ri	Ri2
8/10/1997	3.24	2.53	0.71	27	27	729
4/11/1997	2.71	3.14	-0.43	4	-4	16
9/12/1997	2.32	2.23	0.09	17	17	289
6/01/1998	2.38	2.34	0.04	15	15	225
3/02/1998	2.72	2.64	0.07	16	16	256
3/03/1998	2.77	2.73	0.04	14	14	196
7/04/1998	2.59	2.15	0.44	24	24	576
5/05/1998	2.72	2.15	0.57	26	26	676
9/06/1998	2.71	2.28	0.43	23	23	529
7/07/1998	2.99	2.60	0.39	21	21	441
4/08/1998	3.20	2.79	0.42	22	22	484
8/09/1998	2.99	2.71	0.28	19	19	361
6/10/1998	3.18	2.96	0.21	18	18	324
10/11/1998	2.58	2.76	-0.18	8.5	-8.5	72.25
8/12/1998	2.45	2.58	-0.13	11	-11	121
12/01/1999	2.43	2.41	0.02	13	13	169
10/02/1999	3.03	2.53	0.50	25	25	625
10/03/1999	2.30	2.82	-0.52	3	-3	9
6/04/1999	3.27	2.91	0.36	20	20	400
4/05/1999	2.30	3.55	-1.25	1	-1	1
8/06/1999	2.66	3.25	-0.58	2	-2	4
6/07/1999	2.60	2.78	-0.18	8.5	-8.5	72.25
4/08/1999	3.87	2.43	1.44	28	28	784
8/09/1999	2.89	3.15	-0.25	5	-5	25
5/10/1999	2.64	2.85	-0.20	7	-7	49
9/11/1999	3.78	2.26	1.52	29	29	841
7/12/1999	2.58	2.79	-0.21	6	-6	36
11/01/2000	2.64	2.78	-0.13	10	-10	100
8/02/2000	2.25	2.31	-0.06	12	-12	144
7/03/2000	3.81	2.18	1.64	30	30	900

Table 26: Rank transformed data

The data is easily ranked using the function RANK in excel or in the stats programme you may be using.

First we should check whether the log 10 Faecal coliform is normally distributed. The output overleaf shows the Shapiro Wilk test for normality for which the null hypothesis is that the data is not normal.

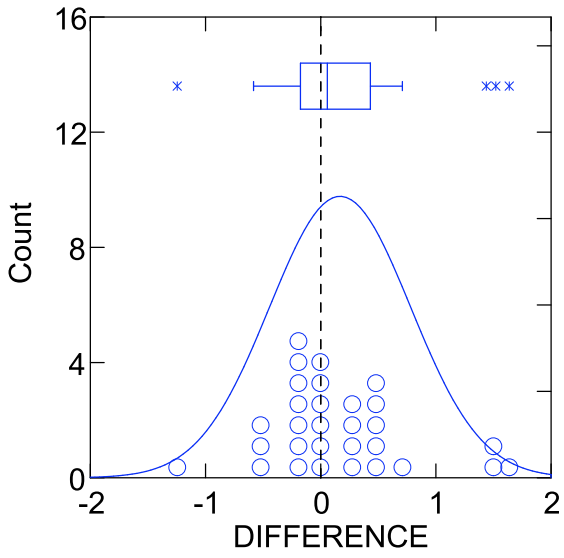
[1] Note there is an important difference between the Wilcoxon rank sum test and the Wilcoxon Signed Rank test. The former is used for comparing independent samples, the latter is used for comparing dependent samples.

	U/S Discharge	D/S Discharge	DIFFERENCE
N of cases	30	30	30
Minimum	2.246	2.146	-1.246
Maximum	3.869	3.547	1.637
Mean	2.820	2.652	0.168
Standard Dev	0.444	0.349	0.612
SW Statistic	0.897	0.959	0.930
SW P-Value	0.007	0.284	0.050

Table 27: Checking for normality of residuals[2]

The d/s of discharge shows normality yet the u/s of discharge does not, the difference column is bang on 0.05, so I will proceed with a single t test on the differences.

```
One-sample t-test of DIFFERENCE with 30 cases
Ho: Mean = 0.000 against Alternative = 'not equal'

Mean                        =        0.168
95.00% CI                   =       -0.061 to 0.397
SD                          =        0.612
t                           =        1.503
df                          =           29
p-value                     =        0.144
Bonferroni adj p-value      =        0.144
```



The output indicates that the differences between up and downstream is not significantly different to zero (p=0.144).

And now for a Wilcoxon Signed Ranks Test

Wilcoxon Signed Ranks Test Results

Counts of differences (row variable greater than column)

	U/S Discharge	D/S Discharge
U/S Discharge	0	18
D/S Discharge	12	0

Table 28: Wilcoxon Rank counts of differences

[2] The Shapiro Wilk test has a null hypothesis that the residuals follow a normal distribution

This part of the output just tells me how often is the rank of the u/s of discharge column > than the rank of the downstream column (18) and how often is the rank of the d/s of discharge column > than the rank of the u/s of discharge column (12)

Z = (Sum of signed ranks)/square root(sum of squared ranks) The following table provides the z values

	U/S Discharge	D/S Discharge
U/S Discharge	0.000	
D/S Discharge	-1.265	0.000

Two-sided probabilities using normal approximation

	U/S Discharge	D/S Discharge
U/S Discharge	1.000	
D/S Discharge	0.206	1.000

Table 29: Wilcoxon Rank output

The output indicates that there is no statistically significant difference between the ranked data of u/s and d/s of the discharge (p=0.206). To conclude the two statistical tests are in agreement that there is no statistically significant difference between the up and downstream sites.

Two Independent Samples

The non parametric equivalent of a 2 sample t test (comparing 2 independent samples) is the Mann-Whitney test. This type of testing might be applicable if you are comparing two sites for which there is an unbalanced design (i.e the number of samples for each site may be different).

so lets give this a go!

Supposing I wish to compare nitrate N concentrations between two different ground water bores. Note this example is a balanced design (number of samples is the same for each site) but it need not have been.

[3]

Date	Site 5100	Site 5135
15/09/1993	0.02	0.001
7/07/1994	0.001	0.01
27/01/1995	0.02	0.001
2/05/1995	0.001	0.001
4/08/1995	0.001	0.02
18/09/1995	0.01	0.001
26/03/2001	0.01	0.01
24/09/2002	0.01	0.01
19/03/2002	0.03	0.01
17/09/2002	0.03	0.03
26/03/2003	0.03	0.03
1/10/2003	0.023	0.03
24/03/2004	0.006	0.001
14/06/2004	0.009	0.023
30/09/2004	0.022	0.001
14/12/2004	0.023	0.001
23/03/2005	0.001	0.086
2/08/2005	0.014	0.004
6/10/2005	0.001	0.004
13/12/2005	0.001	0.004
22/03/2006	0.007	0.004
22/06/2007	0.001	0.003
17/06/2008	0.001	0.0025
2/09/2008	0.001	0.001
2/12/2008	0.0031	0.001
3/03/2009	0.0033	0.0049

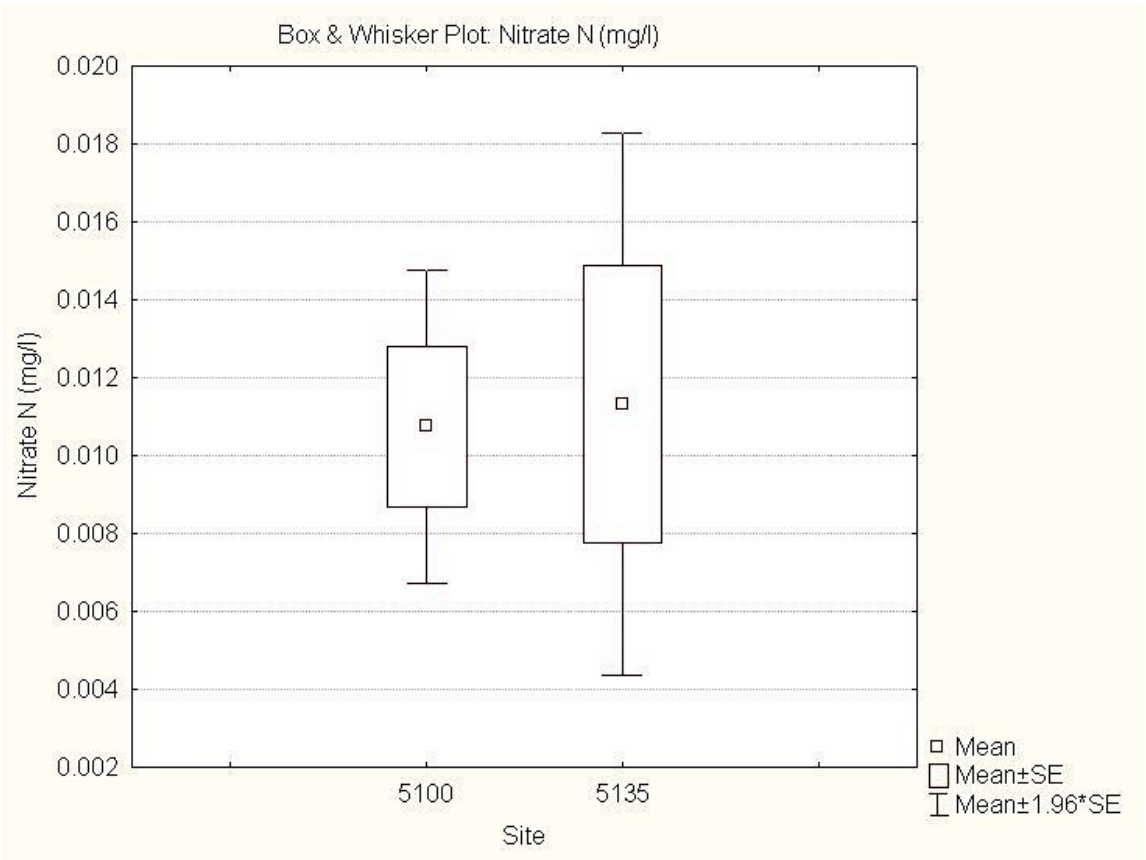
	T-tests; Grouping: Site (groundwater nitrate example) Group 1: 5100 Group 2: 5135										
Variable	Mean 5100	Mean 5135	t-value	df	p	Valid N 5100	Valid N 5135	Std.Dev. 5100	Std.Dev. 5135	F-ratio Variances	p Variances
Nitrate N (mg/l)	0.010746	0.011323	-0.140660	50	0.888704	26	26	0.010452	0.018115	3.003908	0.007843

Variable	Levene F(1,df)	df Levene	p Levene
Nitrate N (mg/l)	0.915169	50	0.343349

Table 30:Nitrate nitrogen data and t test output on site differences

The two sample t test indicates that the sites are not statistically significantly different to each other (p=0.88). The F ratio variances indicate that the samples do not have equal variance (a requirement of t tests and ANOVA's), however the more robust Levene test indicates otherwise. This is a situation where we need to look at the data more closely.

[3] The cells highlighted are those where the results were less than detection, in which case the detect value is halved.



The variance of the two samples certainly does appear to be quite different.

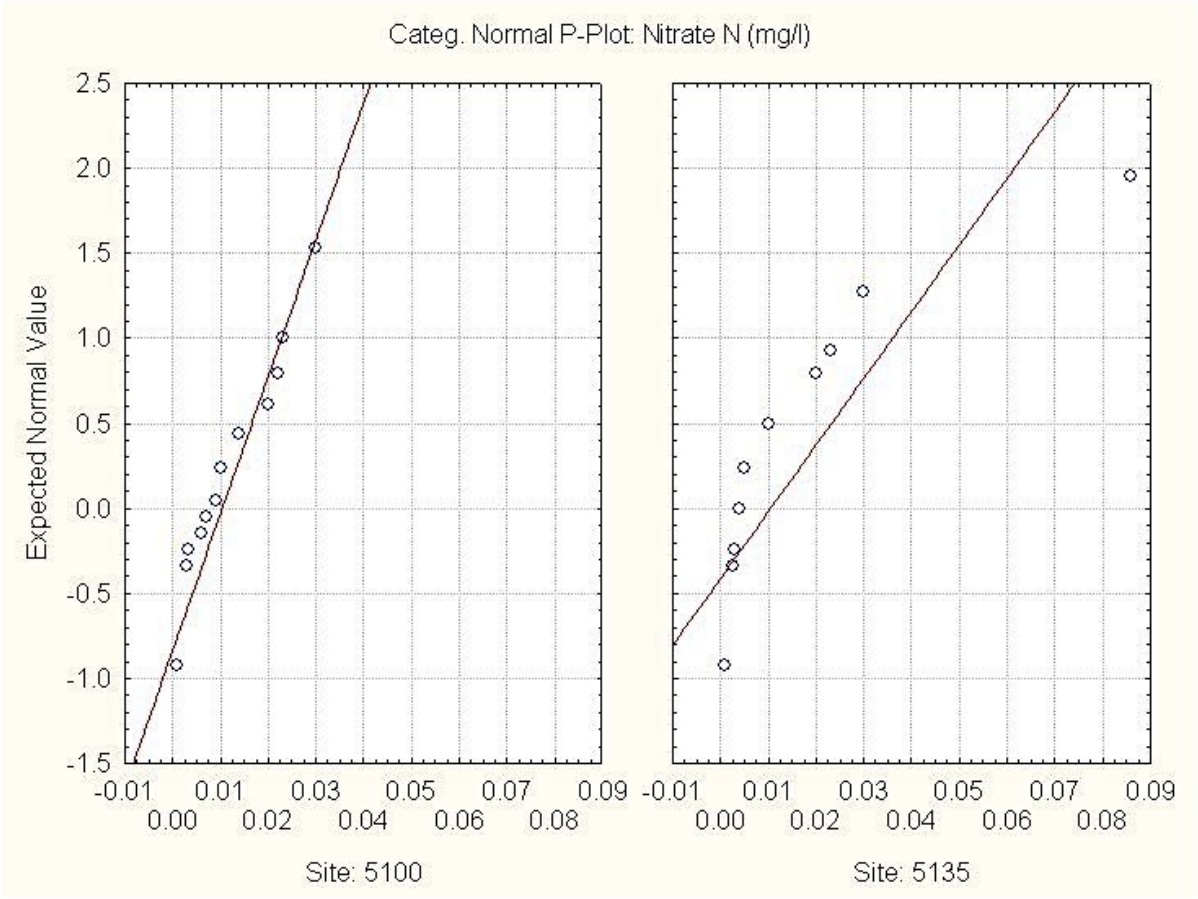


Figure 45: Box plots and normal probability plots for the two sites

The probability plot would suggest that site 5100 conforms quite closely to a normal distribution however site 5135 is questionable. On these grounds I would treat the 2 sample t test results with caution. Let's now proceed with the nonparametric alternative, the Mann-Whitney test.

	Mann-Whitney U Test (groundwater nitrate example) By variable Site Marked tests are significant at p <.05000									
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-level	Z adjusted	p-level	Valid N Group 1	Valid N Group 2	2*1sided exact p
Nitrate N (mg/l)	711.0000	667.0000	316.0000	0.402626	0.687223	0.412223	0.680177	26	26	0.696380

Table 31: Mann-Whitney test output

The Mann-Whitney test confirms that the two sites nitrate nitrogen concentrations are not statistically significantly different to each other. To conclude both tests conclude that there is no statistically significant difference between the two sites, however the t test results are questionable due to the equality of variance testing and the normal probability plot of site 5135. Therefore I would base my decisions on the Mann Whitney test.

Greater Than Two Independent Samples

I will now analyse a data set containing more than 2 samples (e.g. 4 sites) for a comparison of means (one way ANOVA) or medians (Kruskal Wallis Test)

SITE$	SITEID$	DATE$	CATCHM$	SIZE$	SITEID2
STATUS$	ORDER	NZREACH	CLIMATE$	SOF$	GEOLOGY$
LANDCOV$	VLYFRM$	CSOF$	CSOFG$	CSOFGL$	RECLAS$
BLACK_DISC	DO	COND	ECOLI	ENTERO	FAECAL
NH4_N	NO3N	PH	DRP	TN	TP
FLOW	LOG10DRP				

The following results are for:
CATCHM$ = AROPA
Data for the following results were selected according to:
siteid2<5

	DRP
N of cases	51
Minimum	0
Maximum	0.04
Mean	0.012
Standard Dev	0.011
SW Statistic	0.866
SW P-Value	0

The following results are for:
CATCHM$ = WAIKARI
Data for the following results were selected according to:
siteid2<5

	DRP
N of cases	45
Minimum	0
Maximum	0.04
Mean	0.007
Standard Dev	0.008
SW Statistic	0.766
SW P-Value	0

The following results are for:
CATCHM$ = MOHAKA
Data for the following results were selected according to:
siteid2<5

	DRP
N of cases	76
Minimum	0
Maximum	0.03
Mean	0.008
Standard Dev	0.008
SW Statistic	0.845
SW P-Value	0

The following results are for:
CATCHM$ = WAIHUA
Data for the following results were selected according to:
siteid2<5

	DRP
N of cases	55
Minimum	0
Maximum	0.053
Mean	0.01
Standard Dev	0.014
SW Statistic	0.746
SW P-Value	0

Table 32: Checking for normality of residuals for the sites.
Data is not normally distributed so I'll try a log 10 transformation.

The following results are for:
CATCHM$ = AROPA
Data for the following results were selected according to:
siteid2<5

	LOG10DRP
N of cases	49
Minimum	-3
Maximum	-1.398
Mean	-2.204
Standard Dev	0.605
SW Statistic	0.819
SW P-Value	0

The following results are for:
CATCHM$ = WAIKARI
Data for the following results were selected according to:
siteid2<5

	LOG10DRP
N of cases	43
Minimum	-3
Maximum	-1.398
Mean	-2.417
Standard Dev	0.523
SW Statistic	0.868
SW P-Value	0

The following results are for:
CATCHM$ = MOHAKA
Data for the following results were selected according to:
siteid2<5

	LOG10DRP
N of cases	68
Minimum	-3
Maximum	-1.523
Mean	-2.326
Standard Dev	0.52
SW Statistic	0.869
SW P-Value	0

The following results are for:
CATCHM$ = WAIHUA
Data for the following results were selected according to:
siteid2<5

	LOG10DRP
N of cases	47
Minimum	-3
Maximum	-1.276
Mean	-2.301
Standard Dev	0.616
SW Statistic	0.863
SW P-Value	0

Table 33: Checking for normality using log 10 data

Still not normally distributed however I'll proceed with an ANOVA even though I shouldn't.

```
Data for the following results were selected according to:
      siteid2<5

Effects coding used for categorical variables in model.

Categorical values encountered during processing are:
CATCHM$ (4 levels)
   AROPA, MOHAKA, WAIHUA, WAIKARI
32 case(s) deleted due to missing data.

Dep Var: LOG10DRP   N: 207   Multiple R: 0.127   Squared multiple R: 0.016

Analysis of Variance
```

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
CATCHM$	1.061	3	0.354	1.112	0.345
Error	64.574	203	0.318		

Table 34: ANOVA summary

This first part of the output tells me that there is no statistically significant difference in mean values of LOG 10 DRP amongst the sites(p=0.345). Some site pairwise comparisons follow.

```
Durbin-Watson D Statistic          1.119
First Order Autocorrelation        0.434
COL/
ROW CATCHM$
  1  AROPA
  2  MOHAKA
  3  WAIHUA
  4  WAIKARI
Using least squares means.
Post Hoc test of LOG10DRP
Using model MSE of 0.318 with 203 df.

Matrix of pairwise mean differences:
```

	1	2	3	4
1	0			
2	-0.122	0		
3	-0.097	0.025	0	
4	-0.213	-0.091	-0.116	0

Table 35: Matrix of pairwise mean differences

```
Tukey HSD Multiple Comparisons.

Matrix of pairwise comparison probabilities:
```

	1	2	3	4
1	1			
2	0.656	1		
3	0.835	0.995	1	
4	0.271	0.842	0.764	1

Table 36: Matrix of pairwise comparison probabilities

So no significant difference with respect to log 10 DRP

and now for a Kruskal Wallis test

Data for the following results were selected according to:
siteid2<5

Categorical values encountered during processing are:
CATCHM$ (4 levels)
AROPA, MOHAKA, WAIHUA, WAIKARI

Kruskal-Wallis One-Way Analysis of Variance for 207 cases
Dependent variable is LOG10DRP
Grouping variable is CATCHM$

Group	Count	Rank Sum
AROPA	49	5757.500
MOHAKA	68	6867.500
WAIHUA	47	4945.000

WAIKARI 43 3958.000

Kruskal-Wallis Test Statistic = 4.547
Probability is 0.208 assuming Chi-square distribution with 3 df

So no significant difference with respect to the KW Test also.
What I don't like about Systat here is that it does not provide me with pairwise comparisons of the KW Statistic where as Statistica DOES.

To conclude both the parametric test (ANOVA) and the non parametric test (KW) indicate no significant differences in the concentration of DRP amongst the sites. Because the data set is not normally distributed, I would use the results of the Kruskal Wallis test for my analysis.

Correlation Analysis

I would now like to look at some outputs comparing correlation analysis. Correlation analysis measures the strength of association between two continuous variables, where neither variable is considered dependent on the other. Three measures of correlation are in common use - Kendall's tau, Spearman's rho and Pearson's r. It is very important to plot the data first to assure yourself that the calculated correlation coefficient is consistent with the actual form of the relationship between the two variables.

The correlation coefficient formula is calculated as

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2 (y-\bar{y})^2}}$$

For Spearman's rho, the correlation is based on the strength of the linear dependence between the ranks of the data, so if you didn't have spearman's rho option on your computer you could simply undertake a Pearson's r computed on ranks of the data with ties replaced by the average rank and this would give you Spearman's rho. Sometimes Spearman's rho is called Spearman's rank correlation coefficient so try not to get confused over it, they mean the same thing.

Kendall's tau is the basis of the Mann-Kendall test for trend that I will discuss later. While requiring greater calculation effort, it has the advantage (over Spearman's rho) of approaching the unit normal distribution more quickly.

In brief tau is a measure of concordance; two pairs of data (e.g. temperature =7, PM10 = 13), (temperature =9, PM10 = 22) are concordant because PM10 increases with temperature; discordant pairs would be e.g. (temperature=7, PM10 = 52), (temperature=9, PM10= 40). Kendall's tau is simply the proportion of concordant pairs among all possible pairs of data.

OK so lets give this a go with assessing the correlation between water clarity and turbidity measurements taken from a Wellington Rivers SOE programme. Firstly I'll start with the parametric Pearson Correlation Coefficient.

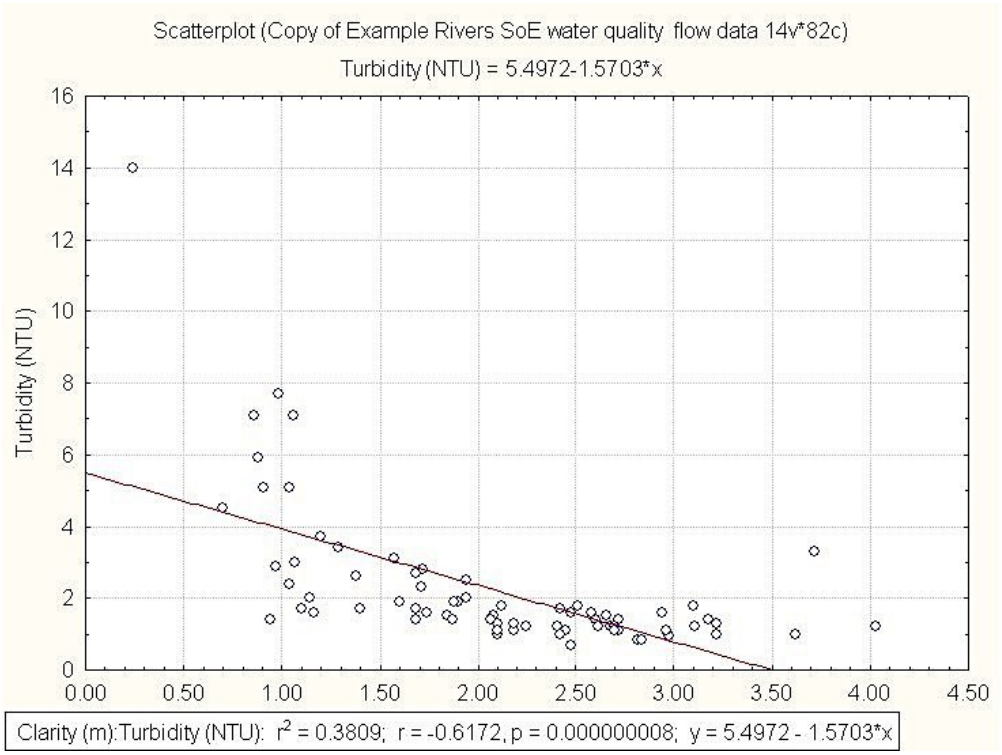


Figure 46: Scatterplot of clarity vs turbidity

Looks like a negative relationship of turbidity with water clarity i.e as turbidity concentrations decline, water clarity increases. The Pearson R is -0.6172 and the slope is highly significant. Certainly the top left point looks to be an outlier and the data is not evenly spread above and below the line, the turbidity concentrations seem to be more spread out at the higher values suggesting some skewness. I could try log transforming the data (log turbidity: log clarity) to see if this improves the strength of the relationship.

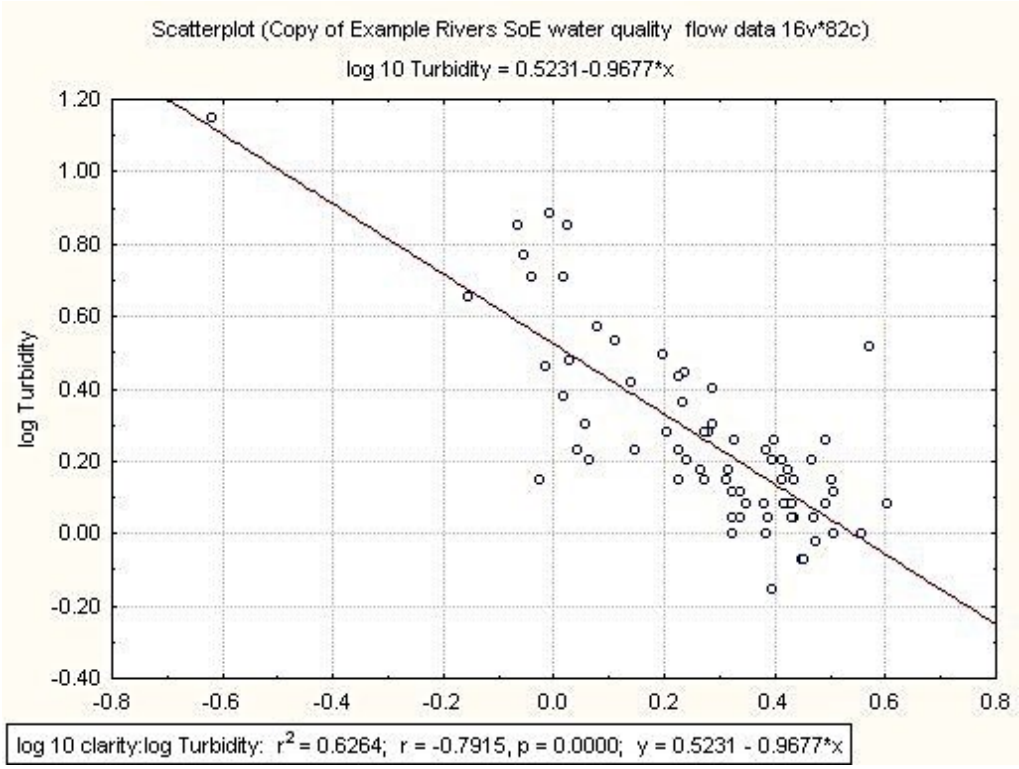


Figure 47: Scatterplot of log10 clarity vs log10 turbidity

By using the log 10 transformation I have improved the strength of the relationship between Turbidity and Water Clarity as Pearsons R now = -0.7915 and the p value is also highly significant (<0.0001). Note the outlier I had mentioned before in the top left is now far closer to the fitted line.

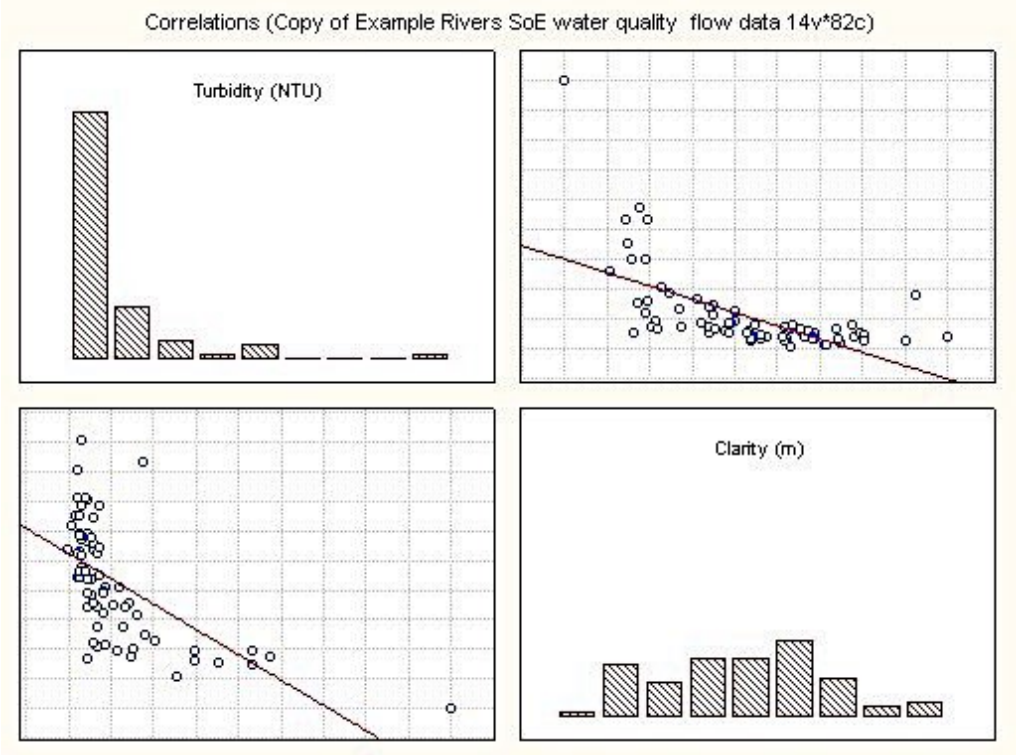


Figure 48: Spearman Rank correlation plot

	Spearman Rank Order Correlations (Copy of Example Rivers SoE water quality flow data) MD pairwise deleted Marked correlations are significant at p <.05000	
Variable	Turbidity (NTU)	Clarity (m)
Turbidity (NTU)	1.000000	-0.722935
Clarity (m)	-0.722935	1.000000

Table 37: Spearman Rank Order correlations

The Spearman's Rank Correlation indicates a strong negative relationship for which rho = -0.723. The p value indicates that the slope is statistically significant. On this output, Statistica has provided me with histograms of each variable, note how Turbidity certainly tends to show skewness, something I suspected from the initial Pearson's plot.

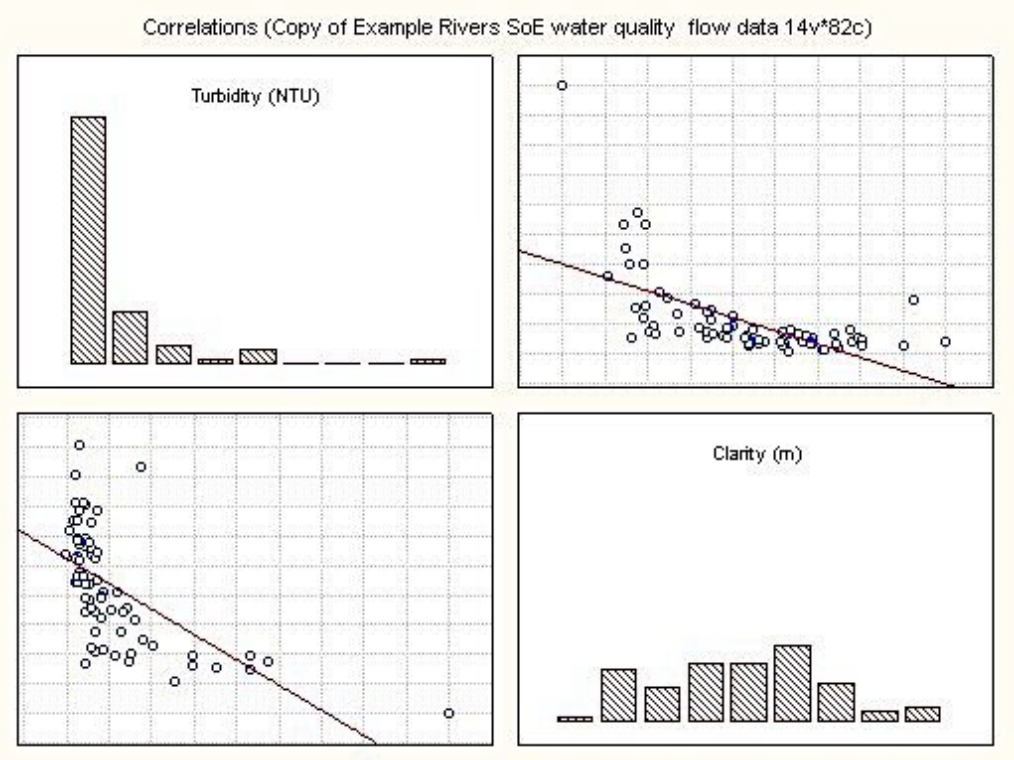


Figure 50: Kendall Tau correlation plot

	Kendall Tau Correlations (Copy of Example Rivers SoE water quality flow data) MD pairwise deleted Marked correlations are significant at p <.05000	
Variable	Turbidity (NTU)	Clarity (m)
Turbidity (NTU)	1.000000	-0.563048
Clarity (m)	-0.563048	1.000000

Table 38: Kendall Tau correlations

And finally here is the kendall Tau analysis, again we see a strong negative relationship of turbidity with clarity which is statistically significant at $p<0.05$.

To conclude, all of the above correlation tests agree that there is a strong negative relationship of clarity with turbidity (and vice versa) and the slope of the line is statistically significant (from zero). The parametric and non parametric results suggest different correlation values because they are measuring different things, the Pearson correlation is creating a slope based on least squares, the Spearman Rank Correlation is creating a slope based on ranks of the data and the Kendall tau is creating a slope based on the proportion of concordant pairs amongst the x,y data.

Time Series Analysis

The objective of time series analysis is to understand the structure that generated the time oriented data, fit a model, monitor and forecast (if necessary). The nature of structural variations in a time series are broadly classified under three major headings:

1.Trend - representing long term positive (upward) or negative (downward) movement

2.Seasonal - a periodic behaviour happening within a block (seasons) of a given time period (say a calendar year) and this periodic behaviour will repeat fairly regularly over time. Exploratory data analysis tools (EDA) are useful for judging the presence of seasonality. e.g.. a simple scatter plot of

the response variable against time which can be highlighted by grouping variables such as month and year.

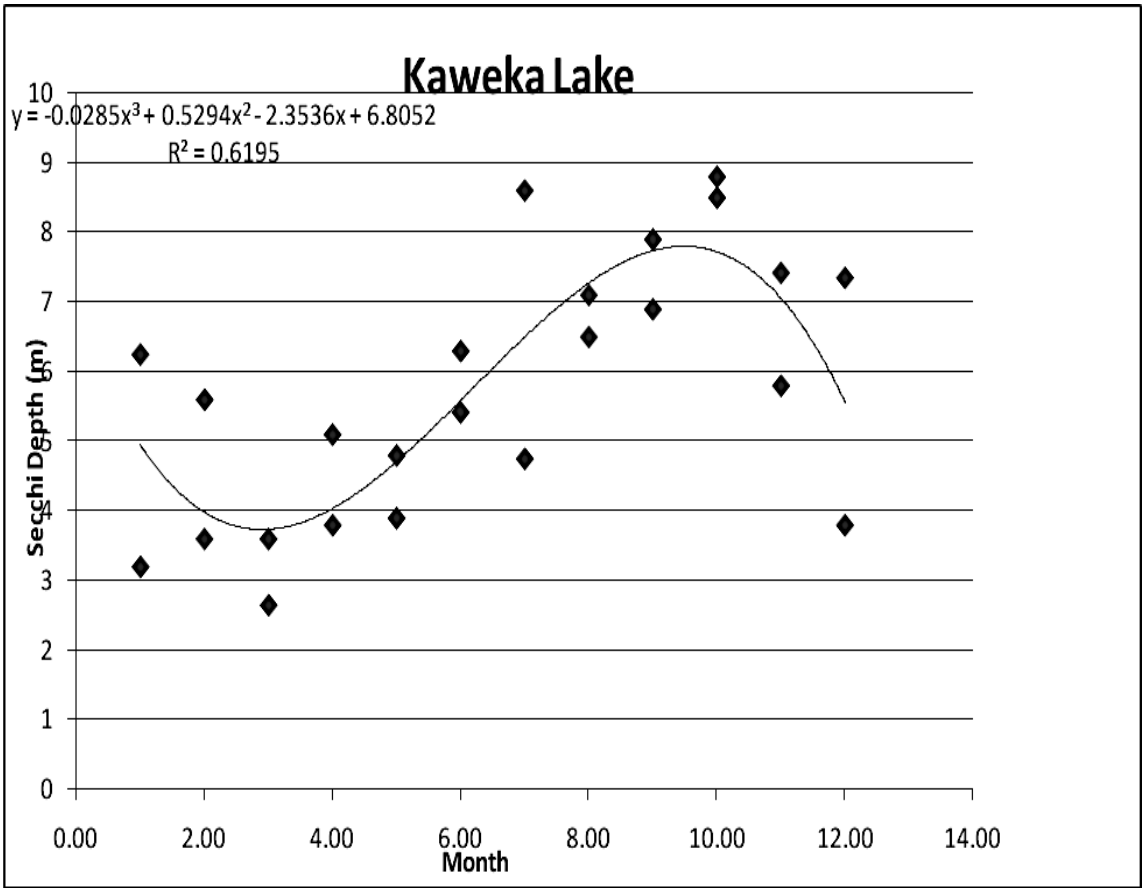


Figure 51: Plotting seasonality of secchi depth for a lake.

Or alternatively plotting seasonal box and whisker plots.

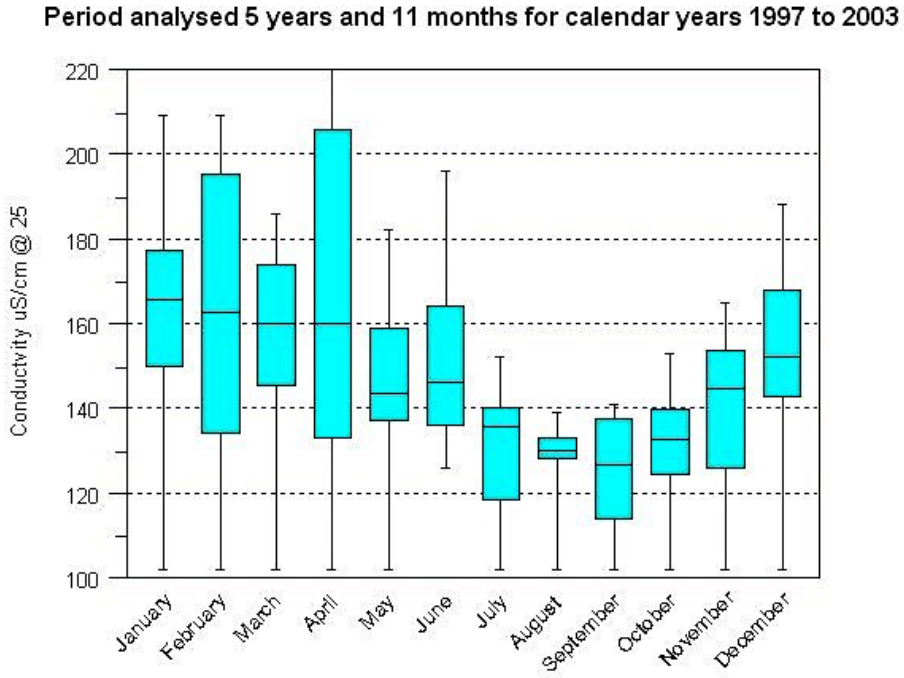


Figure 52: Box and whisker plots of conductivity by month

Clearly this seasonal box and whisker plot produced demonstrates a large degree of seasonality. If you wanted to get real keen you could follow on with a Kruskal Wallis test of the month groups to confirm.

Also useful are Autocorrelation plots - this is where the correlation within the time series is examined at certain lags . By lags I mean comparing the correlation between the observation at time t and t-1, t-2, t-3, t-4 etc, the lag value is presented on the x axis and the correlation value is presented on the y.

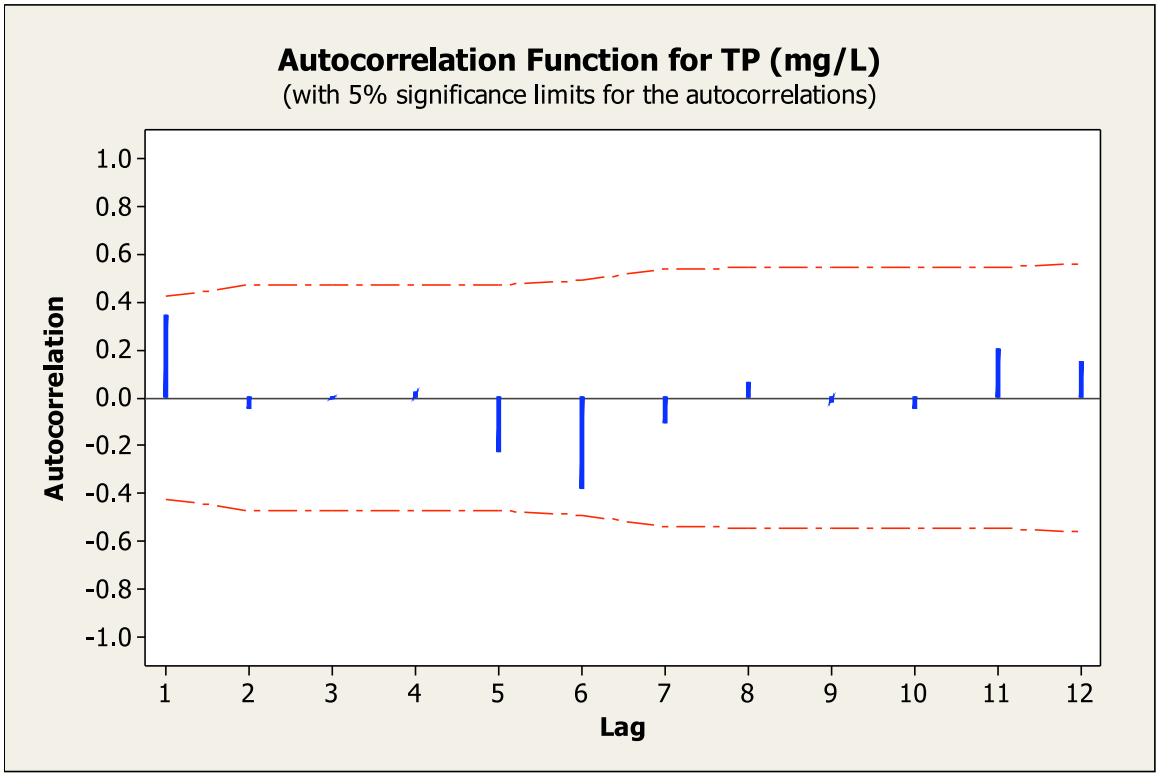


Figure 53: Autocorrelation function plot for TP (mg/l)

In the above plot all lags fall within the 5% lag indicating no significant correlation therefore the monthly monitoring programme is appropriate.

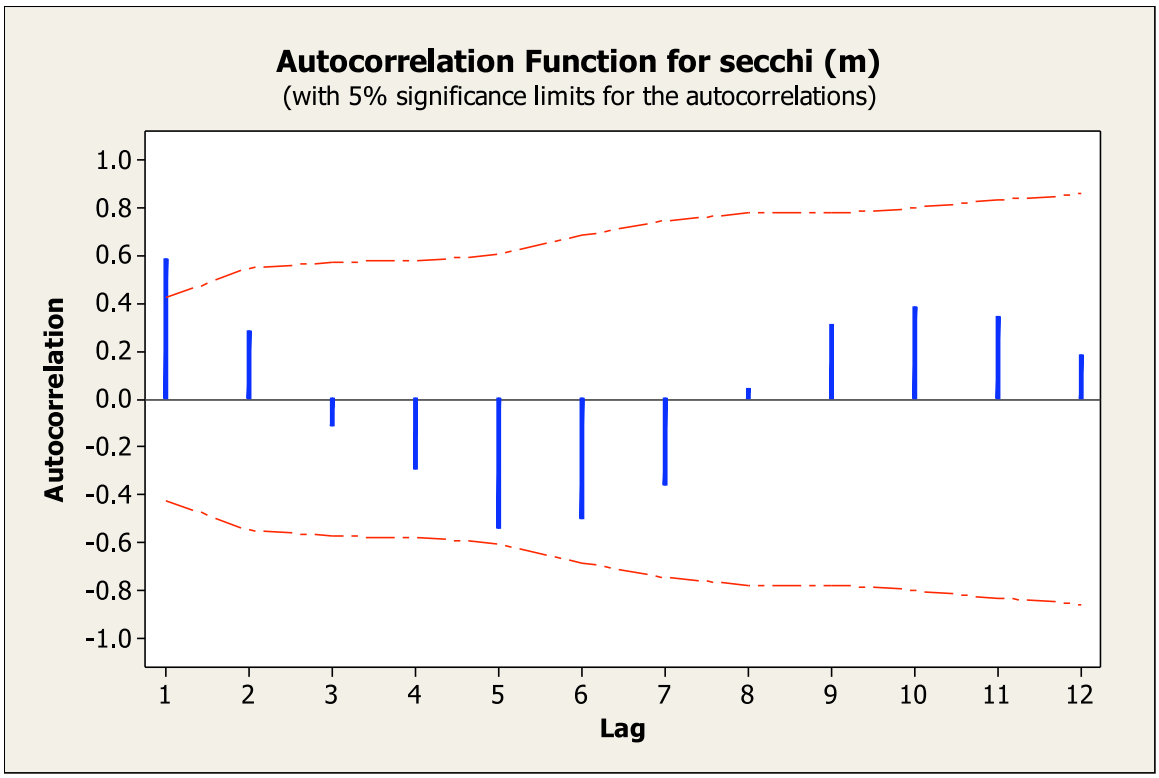


Figure 54: Autocorrelation function plot for secchi depth (m)

In this plot there is a significant amount of autocorrelation at lag =1, this would suggest that you may get away with a bi monthly sampling programme, however given the previous TP plot, a monthly monitoring programme is most appropriate. Remember, what appears to be trend at a smaller scale (e.g.. a 5 yr upward trend), may be better modeled as correlation as larger time scales (e.g.. southern oscillation index with associated climatic effects).

3.Finally as with all models there is an error component

It is good practice to assume a typical structural interrelationship between the trend and season or cyclical components of a time series. This may be done using an additive model for the observation at time t say x_t as

x_t = Trend + Seasonal + Error or

using the multiplicative model

x_t = Trend *Seasonal + Error

The multiplicative model takes the interaction between the components into account whereas the additive model assumes independence of the components. Before I proceed with the basics of the non parametric Mann Kendall Trend test, I'd like to touch on a couple of other techniques primarily because unlike the Mann Kendall trend test , they can be used for forecasting trends, which I think is important for regional councils.

The Moving Average Model

Here we compute the mean of successive smaller sets of numbers of immediate past data. The period of length (also called span) of the moving average is the number of observations (including the present one) used for averaging. The general expression of the moving average is $M_t = [x_t+x_{t-1} +...x_{t-N+1}] / N$ where x_t is the observation at time t and N the moving average length.

I'll demonstrate further with the table below, supposing we had what appeared to be increasing MCI values in the Waikanae River

Year	MCI	Moving Average of length N=2	Explanation	Moving Average of length N=3	Explanation
1978	80	*	1	*	1
1979	78	79	(80+78)/2	*	2
1980	81	79.5	(78+81)/2	79.66	(80+78+81)/3
1981	83	82	(81+83)/2	80.66	(78+81+83)/3
1982	89				
1983	82				
1984	95				
1985	97				
1986	99				
1987	106				

1 Note for the first observation there is no preceding observation and hence the moving average is not computed.

2 Note for N=3, there are no two preceding observations and hence the moving average is not computed.

Table 39: Moving average model

When placing the moving averages against time, placing them in the middle time period is more appropriate. Strictly speaking the moving average must fall at t =1.5, 2.5, 3.5 etc when the period of the moving average is an even number. Hence we smooth again the moving average smoothed values to place the moving averages at t = 2,3,4 etc (called centered moving average).

The major disadvantages of the moving average smoothing methods is that if there is any strong trend in the original data, the moving average smoothing process will be very slow to pick it up.

Exponential (Average) Smoothing

In moving average smoothing all past observations are given equal weight. in exponential average smoothing past observations (i.e as the observations become older) are given exponentially decreasing weights. That is recent observations are given relatively more weight than the older

observations. The average computed using exponentially decreasing weights is known as Exponentially Weighted Moving Average.

EWMA starts by setting S_O to x_1 , where S stands for the smoothed observation and x for the observation. The subscript x refers to time periods t=1,2,....n,. For the second period, $S2 = \alpha x_2 + (1-\alpha)S_1$ and so on. Here the parameter α is called the smoothing constant, the weight given to the current observation. Let's try EWMA on the MCI data in the previous example using α = 0.6.

t	Year	MCI at xt	St EWMA, $\alpha = 0.6$	Explanation $S_t = \alpha x_1+(1-\alpha)S_{t-1}$
1	1978	80	80	0.6*80+0.4*80
2	1979	78	78.8	0.6*78+0.4*80
3	1980	81	80.12	0.6*81+0.4*78.8
4	1981	83	81.8	0.6*83+0.4*80.12
5	1982	89	86.12	0.6*89+0.4*81.8
6	1983	82	83.65	0.6*82+0.4*86.12
7	1984	95	90.46	0.6*95+0.4*83.65
8	1985	97	94.38	0.6*97+0.4*90.46

Table 40: Exponential average smoothing model

Lowess

Locally weighted scatterplot smoothing is a popular smoothing technique sometimes applied to water quality data to observe general trends. A logical regression model is fit to each point and the points closest to it (the number determined by the tension).

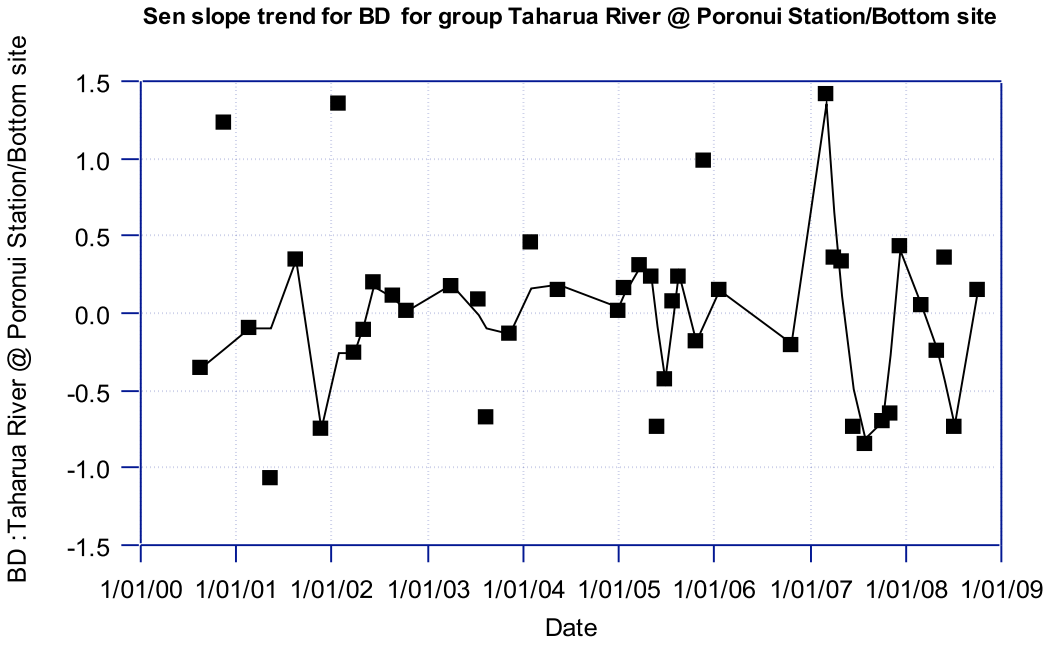


Figure 55: Lowess plot at tension 10%
The above plot shows 10% Lowess Span – the result = quite a squiggly line
Note the tails of the time series use a smaller span than data in the middle of the series i.e in the above case a 10% tension is set. This means that for data in the middle of the series the regression line is calculated using 5% of the data either side of each point x y that is being smoothed, however at the beginning of the series there is no data before it, therefore the first x, y data point uses 5% of the data to the right of it. Alternatively the last data point can only use the previous 5% worth of data before it as there are no data points occurring after it.
The same data with a 30% span

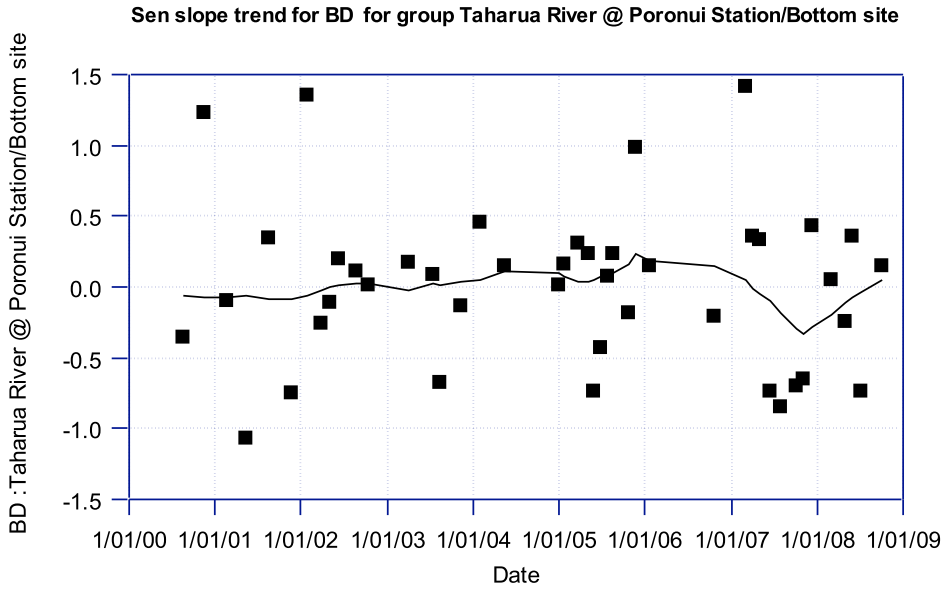


Figure 56: Lowess plot at 30% tension

And finally the same data with a 50% span

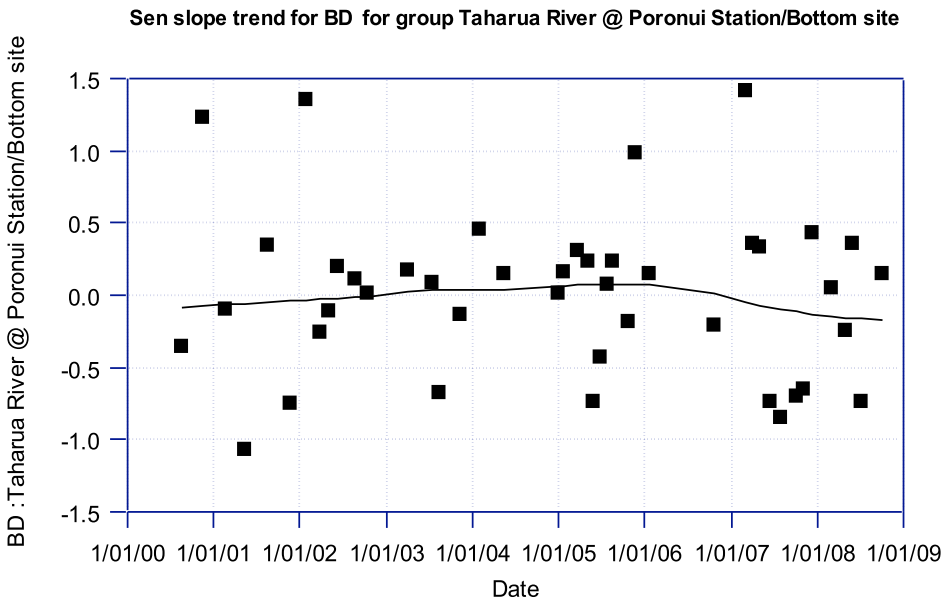


Figure 57: Lowess plot at 50% tension

The Mann-Kendall Trend Test

The Mann Kendall trend test is a non parametric method used to determine time series trends. Supposing we wish to analyse the following time series of n= 10 nitrate nitrogen concentrations . Is there a trend at the 90% level?

0.98	0.697	0.74	0.687	0.604	0.567	0.708	0.865	0.806	1.113
1	2	3	4	5	6	7	8	9	10
	-	-	-	-	-	-	-	-	+
		+	-	-	-	+	+	+	+
			-	-	-	-	+	+	+
				-	-	+	+	+	+
					-	+	+	+	+
						+	+	+	+
							+	+	+
								_	+
									+

Table 41: Record of concordance and discordance of the data

The above table simply records whether the difference between each pair is positive or negative i.e top left, the change from 0.98 at t = 1 to 0.697 at t= 2 is a negative change, the change from 0.98 at t=1 to 0.74 at t=3 is a negative change etc. The test statistic S = (number of positive differences) - (number of negative differences), so S = 26-19 = 7 and the variance of S

$$var(S) = \frac{n_i(n_i-1)(2n_i+5) - \sum[t_{ip}(t_{ip}-1)(2t_{ip}+5)}{18}$$

where ni = the number of data and t_{ip} = the number of tied data.

Now fortunately there are no tied data set (yes I deliberately did this to make it a bit more understandable), so var S = 10*9*25-0/18 = 125 and so the normalised test statistic is Z= (7-1)/ √125 =0.54

Since $Z_{0.9}$ = 1.282 > 0.54, we would not reject the null hypothesis and conclude that there is not a statistically significant monotonic trend at the 90% confidence level.

The Sen Slope Estimator

This is the easy bit, it's simply the median slope of all possible pairwise comparisons (i.e the first point with the second point, the first point with the third point, the first point with the fourth point etc through to the last point of the series, the test then goes back and computes slopes for the second point with the third point, the second point with the fourth point, etc until all possible sequential slopes have been calculated and then calculates the median of all these slopes).

Seasonal Kendall Trend Test

This is almost the same as the Mann-Kendall trend test except that the test accommodates seasonality using the following procedure. It is the preferred trend test to use if your data displays seasonality.

1. Partition the data set according to season (e.g.. for k=12 monthly data put the Januarys together, the February's together etc).
2. Compute statistics S_i and var(S_i) for each season
3. Compute sums over all seasons :$S' = \sum_{i=1}^{k} S_i$ and $var(S') = \sum_{i=1}^{k} (S_i)$
4. Then compute the test statistic from
5. $Z = \frac{S'-1}{\sqrt{var(S')}}$ if $S' > 0$

$Z = 0$ if $S' = 0$

$Z = \frac{S'+1}{\sqrt{var(S')}}$ if $S'<0$

Seasonal Kendall Sen Slope Estimator

The seasonal kendall sen slope estimator should be used if your time series displays seasonality. In this case it compute slopes between pairs of points within each season only and select the median as the slope estimate. The seasonal Kendall slope estimator is faster to calculate than the regular Sen slope estimator since the number of points considered is much smaller.

Note if there are a large number of tied data (e.g many non detects), you can find anomalous results using the Seasonal Kendall slope estimator and the Seasonal Kendall Trend test. e.g.. slope = 0.0000 that is statistically significant. This is because the two procedures are completely separate calculations. This is where you need to bring your professional judgement in. I have adopted Environment Waikato's approach which is to say if the slope of the trend is < 1% of the mean of the data set, then it is not environmentally significant, this was also an approach I used when I worked for Greater Wellington many years ago.

Determining Environmental Significance

As we all know statistical significance does not necessarily equate to environmental significance when analysing water quality data. Fortunately statistics can provide a support tool that enables the scientist to test an interval beyond which they consider the change (it may be a time trend, it may be differences between sites etc) is environmentally significant.

The neat thing about interval hypothesis testing can be summarised as follows:

- Interval hypothesis testing allows the user to specify bounds that describe an "environmentally " significant change
- You can accept or reject an interval!
- You can reject a point null hypothesis if $p < \alpha$ but you cannot accept a point null hypothesis. You can only say you have failed to reject it. This is because for example if you were comparing two sample mean values from two populations they may look to be equivalent but they may not be particularly if taken to the zillionth decimal place.

I'll now show you an output for some analysis I did this year to give you an idea of the strengths of this technique, but first some background. We undertook some pre consent monitoring for a development in the Manawatu and found that the macroinvertebrate community index varied seasonally throughout the year. This supported research that suggested that the MCI value could vary by as much as 10% during the course of a year.

So rather than asking the question "is the mean MCI at a site after windfarm construction statistically significantly different when compared to pre construction monitoring", we ask the question " is the change in the mean MCI value greater than 10% when comparing post construction data with pre construction data?" The statistical output follows:

Equivalence test of means

Reference site is "Trib T1 Jul"

Equivalence limits are -10.0% to 10.0% of mean value of data for reference site

No practically important difference (difference may be trivial compared to the limits)

Variable: "MCI Value"		
Grouping variable	**"Trib T1 Jul"**	**"Trib T1 June"**
N	4	4
Means	124.505	130.155
SD	2.93959	3.3762
H_0: no difference	**Reject, P = 0.04503**	
H_i: difference lies beyond limits (inequivalence)	Reject, P = 0.01143	
H_e: difference lies within limits (equivalence)	**Accept, P = 0.98857**	
Bayesian posterior probability (%) that difference is within limits	98.8477	

Table 42: Interval hypothesis testing output

This output shows that while the traditional point null hypothesis testing shows that the mean values are statistically significantly different (p=0.045), the difference lies within the limits (in this case 10%) of the interval. Therefore this site shows compliance.

Interval hypothesis testing is a very useful tool to use when wanting some decision support on determining environmental significance and it should be used a lot more!

I have to congratulate you for hanging in there with the data analysis section of this course, you have done well to have understood the basic concepts of the tests I have demonstrated. To be fair we have only scratched the surface of the statistical analysis methods to apply to water quality data however the methods I have discussed in this course cover many applications in day to day water management. I thank you for attending and participating in this course and I appreciate your feedback of the days events.

GLOSSARY OF TERMS

Hydrosphere - the hydrosphere describes all the places on earth where water is stored including the atmosphere, oceans, wetlands, rivers, glaciers, snowfields and groundwater. Water moves from one storage are to another via evaporation, condensation, precipitation, deposition, runoff, infiltration, sublimation, transpiration and groundwater flow.

Sublimation - the process where ice changes into vapour without first becoming liquid. Think of it as the vapor you see that emanates from ice.

Specific Heat- the heat capacity of a unit mass of a substance or heat needed to raise the temperature of 1 gram (g) of a substance 1 degree Celsius.

Runoff -The topographic flow of water from precipitation to stream channels located at lower elevations. Occurs when the infiltration capacity of an area's soil has been exceeded. It also refers to the water leaving an area of drainage. Also called overland flow.

Surface Tension - Tension of a liquid's surface. Due to the forces of attraction between molecules.

LITERATURE CITED

Argue, J. R.; Pezzaniti, D. 1997: Are constructed wetlands the total solution for treating stormwater in an urban environment? Proceedings of the 24th hydrology and water resources symposium. November 24-28, Sheraton Auckland Hotel and Towers, Auckland. PP 13-24

Biggs, B.J.F & Kilroy, C. (2000) Stream Periphyton Manual. Prepared for the Ministry for the Environment. ISBN 0-478-09099-4

Burns, N., Bryers, G. & Bowman, E. (2000) Protocol for Monitoring Trophic levels of New Zealand Lakes and Reservoirs. Prepared for the Ministry for the Environment December 2000. ISBN 0 478 09096X

Clayton, J. and Edwards, T.,(2006) Aquatic plants as environmental indicators of ecological condition in New Zealand lakes. Hydrobiologia 570: 147 - 151.

Donnison, A. (1998) The potential for ultraviolet disinfection of meat processing wastewater. Water and Wastes in New Zealand. March 1988

Graham, A. A. (1990) Siltation of stone-surface periphyton in rivers by clay –sized particles from low concentrations in suspension. Hydrobiologia 199: 107-115

Hickey, C.W. & Martin, M.L (2009) A Review of nitrate toxicity to freshwater aquatic species. Technical Report Prepared For Environment Canterbury. Report No. R09/57 ISBN 978-1-86937-997-1

Joy, M.K. & Death, R.G. (2004) Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. Environmental Management, 34, 415-428.

Lenat, D. R. (1988) Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. Journal of the North American Benthological Society 7 (3): 222-233.

Makepeace, DK; Smith, DW; Stanley, SJ. 1995. “Urban storm-water quality: summary of contaminant data.” Critical Reviews of Environmental Science and Technology. 25:93-139.

McBride, G. M. (1999) Statistical Methods For Water Quality Studies.NIWA Short Course, Quality Hotel, Hamilton, 8-10 March 1999.

Nagels, J.W, Davies-Colley, R.J & Smith, D.G *(2001)A water quality index for contact recreation in New Zealand. Water Science and Technology Vol 43 (5):282-295

Ott, W.R. (1978). Environmental Indices. Theory and Practice. Ann Arbor Science, Ann Arbor, Michigan.

Stark, J. R. 1985. A macroinvertebrate community index of water quality for stoney streams. Water & Soil Miscellaneous Publication 87: 53p. (National Water and Soil Conservation Authority: Wellington, New Zealand).

Stark, J. R. & Maxted, J. R. (2007b) A user guide for the Macroinvertebrate Community Index. Cawthron Report No. 1166

Marshall, J.W. 1974; A biological investigation of the Leeston Drain, Canterbury, new Zealand. MSc thesis, Universtiy of Canterbury Christchurch

Rogers, D.C. () Aquatic macroinvertebrate occurrences and population trends in constructed and natural vernal pools in Folsom, California. PAGES 224-235 in: C.W. Witham, E.T. Bauder, D. Belk, W.R. Ferren Jr., and R. Ornduff (Editors). Ecology, Conservation, and Management of Vernal Pool Ecosystems – Proceedings from a 1996 Conference. California Native Plant
Society, Sacramento, CA. 1998.

Ryan, P.A. (1991) Environmental effects of sediment on New Zealand streams: a review. New Zealand Journal of Marine and Freshwater Research 25:207-221

Ryder, G. I.(1989) Experimental studies on the effects of fine sediment on lotic invertebrates. PhD. thesis University of Otago, Dunedin

Sharpin, M. & Breen, P. (1997) An ecosystem health approach to managing urban stormwater. Proceedings of the 24th hydrology and water resources symposium. November 24-28, Sheraton, Auckland hotel and Towers, Auckland. Pp 223-228

Snelder , T. & Trueman, S. 1995: The environmental impacts of urban stormwater runoff. Technical publication 53 Auckland Regional Council, Auckland.

Snelder, T., Biggs, B. & Weatherhead, M. (2003) New Zealand River Environment Classification A Guide To Concepts And Use. Prepared for the Ministry for the Environment. NIWA Project MFE 02505. NIWA Christchurch PO Box 8602 Christchurch.

Stark, J.D. 1998: SQMCI: a biotic index for freshwater macroinvertebrate coded abundance data. New Zealand journal of Marine and Freshwater Research 32:55-66.

Williamson, R. B. (2003) Urban runoff data book. Water Quality Centre publication 20. NIWA, Hamilton.

Zarr, J.H. (1996) Biostatistical Analysis. Prentice-Hall International, Inc.

APPENDIX 1: Nutrients Essential For Life

Typical concentrations sufficient for plant growth. After E. Epstein. 1965. "Mineral metabolism" pp. 438-466. in: Plant Biochemistry (J.Bonner and J.E. Varner, eds.) Academic Press, London.

Element	Symbol	mg/kg	percent	Relative number of atoms
Nitrogen	N	15,000	1.5	1,000,000
Potassium	K	10,000	1	250,000
Calcium	Ca	5,000	0.5	125,000
Magnesium	Mg	2,000	0.2	80,000
Phosphorus	P	2,000	0.2	60,000
Sulfur	S	1,000	0.1	30,000
Chlorine	Cl	100	--	3,000
Iron	Fe	100	--	2,000
Boron	B	20	--	2,000
Manganese	Mn	50	--	1,000
Zinc	Zn	20	--	300
Copper	Cu	6	--	100
Molybdenum	Mo	0.1	--	1
Nickel	Ni	0.1	--	1